Information Heterogeneity and the Demand for Healthcare: Evidence from GPs in England*

Zach Y. Brown
University of Michigan & NBER

Christopher Hansman
Emory University

Jordan Keener
University of Michigan

Andre Veiga
Imperial College London

June 2026

Abstract

Do different patient choices reflect what they value or what they know? We study this question using variation created by rounded star ratings for GP practices, which changes the information about providers that patients can publicly observe. Differences in the demand response to these ratings indicate that some groups—notably high-income patients—are better informed about provider quality. We develop an empirical demand model that allows both information and preferences to vary across patients and find that heterogeneity in choice overwhelmingly reflects differences in information. Standard full-information demand models underestimate the gains from improving providers by more than 50%, and fail to capture the incidence of those gains. Low-income and other low-information groups benefit substantially from such improvements even when their observed demand for highly-rated GPs is weak.

JEL Classification: L15, I11, I14, D83

*We are grateful to Martin Gaynor, Matt Grennan, Ben Handel, Maria Polyakova, Carol Propper, and seminar participants at APPAM, Arizona, ASHEcon, Berkeley, Caltech, the CEPR virtual IO seminar, Columbia, Duke, Groningen, GSU, HEC Paris, Helsinki GSE, Indiana, Johns Hopkins, Maryland, Michigan, NBER Summer Institute, NYU Stern, Penn State, SITE, Stanford, SUNY Buffalo, Toulouse, Tsinghua, UC Irvine, UCSB, UGA, UIUC, UT Austin, WBEC, Yale, and the ZEW Conference on Information and Communication Technologies for valuable comments. We thank Radoslaw Kowalski for sharing data. A previous version of this article was titled, "Information and Disparities in Health Care Quality: Evidence from GP Choice in England."

1 Introduction

Observed demand reflects both what consumers value and what they know. If some consumers are less informed, differences in demand need not reflect differences in preferences. Identifying the role of information matters both for interpreting observed choices and for evaluating policies that disclose information or change the options consumers face. The distinction is especially important in health care, where patients often choose among providers on dimensions of care that are difficult to observe. The challenge is that choice data alone generally do not reveal whether demand heterogeneity reflects preferences or information.

We address this challenge in the context of the English National Health Service (NHS), where a natural experiment allows us to study the role of information heterogeneity in provider choice. We focus on demand for general practitioners as a function of patient ratings. These ratings are a review-based measure of patients' experiences with providers. Demand for highly rated providers is strong on average but varies substantially across neighborhoods and demographic groups. For example, individuals in high-income neighborhoods are about twice as likely to enroll with a local practice in the top decile of ratings; similar gradients appear by age and other characteristics.

Our empirical strategy exploits variation in public information generated by a widely used NHS website that displayed coarse star ratings for each GP practice. The website constructed these star ratings by rounding recent patient reviews, so practices just above and below each rounding threshold received different public ratings despite having similar underlying reviews. We measure both the discrete demand response at rounding thresholds and the slope of demand with respect to the underlying review measure, and we show that both margins vary across observable patient groups. We then develop and estimate a structural demand model that leverages these reduced-form moments to separate heterogeneity in preferences from heterogeneity in information.

Our central finding is that heterogeneity in information, rather than preferences, explains most of the variation in demand for highly rated providers. Certain groups—such as low-income patients—react sharply to the public signal provided by the website, but show little sensitivity to underlying ratings otherwise. Others, including high-income patients, sort systematically toward practices with better underlying reviews, but show little response to star ratings. The structural model attributes this pattern to differences in the precision of patients' information. Both groups value higher ratings, but low-income patients rely heavily on the public signal, whereas high-income patients have better information about providers even absent the star ratings on the website.

Accounting for heterogeneous information changes the evaluation of both disclosure and supply-side policies: public ratings improve sorting and reduce disparities in GP choice, and models that abstract from information heterogeneity understate the aggregate and distributional gains from provider improvements.

English primary care is a useful setting in which to study provider choice for several reasons. GPs are the first point of contact with the health care system, registration is free, and patients choose between nearby practices that differ substantially along observable measures of quality. The setting also offers unusually rich data: individual and neighborhood-level enrollment records can be linked to a range of provider characteristics, including patient ratings and clinical quality metrics.

We begin by using these data to document substantial heterogeneity across demographic groups in the types of GPs patients choose. In particular, patients differ sharply in their tendency to enroll with highly rated practices in their choice set. The same holds true with respect to other measures of quality.

To guide the analysis, we introduce a simple theoretical model of patient learning in which individuals combine a private signal about provider quality with a public rounded star rating. The model implies that coarse public ratings should affect beliefs most for patients with imprecise private signals, whereas patients with precise private signals will update little. The model also shows that two reduced-form objects are jointly informative: the jump in demand at rounding thresholds and the slope of demand with respect to the underlying review measure. Patients with precise private signals tend to choose providers with high underlying quality but respond little when the public star rating changes. Patients with imprecise private signals should exhibit flatter within-bin sorting but larger discrete responses at thresholds. By contrast, weaker preferences for quality attenuate both margins. This distinction guides both the reduced-form analysis and the structural model.

Motivated by this framework, we first use a regression discontinuity design that compares GP demand just above and just below the rounding thresholds for star ratings. We find sharp discontinuities in enrollment when stars are visible, and show that these discontinuities disappear when ratings are not displayed on the website. We do not find any corresponding discontinuities in wait times or other observable provider characteristics and do not find bunching in the distribution of GP ratings. We find similar patterns among individuals who must actively choose a new GP after changing address and in a panel-regression strategy focusing on GPs that experience star-rating changes over time. These results support the interpretation that the design isolates the effect of public information rather than supply-side differences or manipulation of the rating system.

We then show that the average effect masks substantial heterogeneity across demographic groups. Patients in low-income neighborhoods respond sharply at rounding thresholds but are not sensitive to differences between GPs with the same displayed star rating. Patients in high-income neighborhoods, by contrast, respond little to star thresholds but sort strongly toward practices with better underlying average reviews within a star bin. Similar patterns emerge when we split patients by other characteristics. Given our model, these cross-group differences are consistent with meaningful heterogeneity in information precision. We fielded a simple online survey of patients which points to a similar conclusion and suggests that social networks play a role in determining who is informed.

To quantify the role of information relative to preferences, we embed the learning framework in a structural model of GP choice. Patients choose among practices based on posterior expected quality, combining a private signal with the coarsened public signal from the displayed half-star bin. When forming expectations, they incorporate the fact that star ratings are less informative when they are based on fewer reviews. Patients with imprecise private signals place substantial weight on the rounded public signal, generating large discontinuities in choice probabilities at star-rating thresholds and weak sorting on the underlying review measure within a star bin. Patients with precise private signals update little at the threshold but continue to sort within bins. The demand system incorporates standard choice attributes, including distance and congestion, as well as practice fixed effects and an outside option, and allows both preferences and the precision of private information to vary across patients. To abstract from inertia and related considerations, we estimate the model using individual-level data on geographic movers in Greater London who must make an active choice about their GP practice.

We consider a sequence of specifications that allow increasingly rich heterogeneity in information precision and preferences: by neighborhood income; by income, age, and disability status; and by flexible principal components that summarize a large set of neighborhood demographics. Our estimation approach uses a penalized likelihood that combines individual-level choice data with auxiliary reduced-form moments. In this way, the structural model is disciplined by the same group-specific threshold discontinuities and within-bin slopes that we recover in our RD design.

The estimates imply substantial heterogeneity in information but comparatively little heterogeneity in preferences for higher-rated providers. In our preferred specification, the mean weight that patients place on their private signal is roughly 0.6, with a standard deviation of 0.3. This implies that some patients rely heavily on the public star signal while others are close to fully informed even without it. By contrast, the preference parameter—while large on average—exhibits limited dispersion across groups. Counterfactuals that equalize information effectively eliminate heterogeneity in responsiveness to patient ratings. Equalizing preferences leaves this heterogeneity largely unchanged. Thus, differences in information rather than tastes account for most of the observed variation in responsiveness to quality.

These differences matter for policy. We show that providing full information improves sorting to higher-rated GPs, with substantial effects on consumer surplus for low-information patients. Even coarse public stars can achieve much of this benefit. Making stars visible delivers more than half of the ratings gain from full information and reduces the correlation between chosen ratings and income by about one-quarter. Better information also amplifies the gains from supply-side improvements: raising the quality of a random quarter of practices by one star increases aggregate consumer surplus by 15% more under full information than in the no-stars environment. The additional gains are especially large for low-income and other low-information patients.

Distinguishing information from preferences is also central for welfare analysis. A standard full-information demand model understates the welfare gains from this supply-side improvement by roughly 55% because it treats weak baseline responsiveness as evidence that patients place little value on ratings. It therefore also fails to capture the incidence of the gains: for low-income patients, the welfare gain in our model is roughly twice the gain implied by the full-information benchmark.

These findings contribute to three literatures. First, the paper contributes to work on consumer responsiveness to quality information in health care and the associated equilibrium effects (e.g., Dranove et al. 2003; Cutler et al. 2004; Pope 2009; Kolstad 2013). Kolstad and Chernew (2009) provides a review. Related work documents consumer uncertainty about quality in health care settings (e.g., Jin and Sorensen 2006; Dafny and Dranove 2008; Werner et al. 2012; Darden and McCarthy 2015; Grennan and Town 2020) and shows that physicians and their families make different health care decisions, consistent with greater expertise (Bronnenberg et al. 2015; Chen et al. 2019; Artmann et al. 2022). Recent work also emphasizes supply-side responses to public quality disclosure for health insurance (Vatter 2025). Relative to this literature, our contribution is to distinguish heterogeneity in preferences from heterogeneity in private information and to show how that distinction changes welfare and incidence.

Second, the paper relates to work on online rating systems and learning from reviews (e.g., Chevalier and Mayzlin 2006; Anderson and Magruder 2012; Lewis and Zervas 2016; Luca 2016; Luca and Zervas 2016; Reimers and Waldfogel 2021; Xiao 2010; Newberry and Zhou 2019). Work has also studied review systems in health care markets, including effects on pricing, doctor congestion, and interruptions in care (Dor et al. 2020; Kummer et al. 2021; Chartock 2025). Our structural approach also complements work estimating patient preferences in primary care markets, such as Huitfeldt et al. (2025). The primary contribution is to use responses to public ratings to identify heterogeneity in private information.

Finally, the paper contributes to broader work on disparities in health care choices (e.g., van Doorslaer et al. 2006; Cookson et al. 2016; Balarajan et al. 2011; Devaux 2015; Handel et al. 2020). We show that information frictions can generate differences in realized quality even in a setting with free registration and near-universal coverage. The mechanism echoes evidence from other choice settings, perhaps most notably education, where information interventions can affect both sorting and inequality (Hastings and Weinstein 2008; Hastings et al. 2015; Kapor et al. 2020).

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

2 Background and Motivation

This section describes GP enrollment in England, the NHS ratings website, and the data used in our analysis.

2.1 GPs and the NHS Choices Website

Choosing a GP practice. In the English National Health Service (NHS), GPs are the first point of contact for patients. GPs provide a wide range of services, including checkups, screenings, vaccinations, simple surgeries, and referrals to secondary care. They are organized into practices comprising several doctors, which contract directly with the NHS.[1]

All residents of England can register with a GP practice free of charge. Individuals register with a practice, not an individual doctor. For this reason, we conduct our analysis at the practice level (for simplicity, we refer to GP practices, potentially comprising multiple doctors, as "GPs"). Absent capacity constraints or a set of specialized circumstances, practices must accept all patients who wish to register.[2] Registration is straightforward and can often be completed online. Patients are free to switch GPs at any time, and medical records are automatically transferred. GPs do not advertise. The practice of selectively enrolling patients is likely minimal since funding is weighted by patient complexity and GPs face strong legal disincentives to cream-skim patients.

Capacity constraints and waiting times may nonetheless shape demand, so it is useful to distinguish between registration and realized access. Our administrative outcome is practice enrollment, not whether or when a patient is subsequently seen. This distinction matters because a patient may register with a practice even if there are delays in getting an ap-

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

[2] The right to choose a GP is directly outlined in the constitution of the NHS. The constitution states "You have the right to choose your GP practice, and to be accepted by that practice unless there are reasonable grounds to refuse." Patients generally register at a GP within assigned catchment areas, although since 2015 they have been allowed to register with a GP outside their designated geographic area. The practice is allowed to refuse a patient if there is concern that the patient lives too far away and traveling will be inconvenient or dangerous given the patient's health status. See `https://www.gov.uk/government/publications/the-nhs-choice-framework`.

pointment. At the same time, enrollment is an important margin of access: routine primary care in England is organized around the practice with which the patient is registered, and switching registration changes the patient's default provider for primary care. Formal practice closures to new patients are rare, but anticipated crowding could still affect choice even when practices remain nominally open.[3] We therefore treat congestion as a potentially important non-quality attribute throughout the paper and return below to direct evidence on wait-time proxies and other capacity-related robustness checks.

Online Reviews. The NHS maintains a website (www.nhs.uk) that provides information to help patients choose a GP practice, hospital, dentist, or other health care provider.[4] NHS Digital reported that the website received more than 1 billion visits per year near the end of our sample.[5] We focus on the GP rating system that, during our main sample, was part of *NHS Choices*. The system allowed patients to leave a written review of a GP practice and to provide a rating from one to five stars, similar to online review platforms such as Yelp or TripAdvisor. Between 2007 and the end of 2019, these reviews were used to construct summary star ratings that were displayed prominently at the top of each GP practice's page and in search results. These summary ratings were calculated as the average rating over the previous two years and rounded to the nearest half-star. For example, average ratings of 3.26 and 3.74 were both displayed as 3.5 stars. Panel A of Appendix Figure A-1 shows an example of a GP page with a star rating. Given that the website was widely known and heavily trafficked, it is reasonable to assume that most patients searching for a GP saw these star ratings.

At the end of 2019, the summary star ratings were removed from the provider pages but the individual ratings and user comments were kept (see Panel B of Appendix Figure A-1). In theory, individuals could still manually calculate each GP's star rating, but the information was significantly more difficult to access. The removal of the star ratings from the website coincided with the start of the COVID pandemic which heavily affected GP

[3] According to Freedom of Information requests, less than 2% of GPs were officially closed to new enrollment in 2016–2018. See FOI-056173. We are not aware of data allowing us to identify the specific GPs that do not accept new patients.

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

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

enrollment. For this reason, our primary analysis considers only the period prior to January 2020 in which the star ratings were directly displayed. We use data from the later period without visible star ratings only in falsification exercises.

2.2 Data

Reviews and Star Ratings. We construct a monthly panel of individual reviews from May 2015 to June 2022.[6] For practice j in month t, let r_{jt} denote the mean rating over the preceding two years and n_{jt} the number of reviews in that window. The displayed star rating s_{jt} equals r_{jt} rounded to the nearest half-star. Both s_{jt} and the review count n_{jt} were displayed to patients on the website prior to 2020.

We interpret the review index r_{jt} as a patient-valued provider attribute. The index may capture communication, access, convenience, practice organization, and aspects of clinical care, but it is neither an exhaustive nor a direct measure of clinical quality. Appendix Table A-2 documents its relationship with external measures. The review index has a correlation of 0.52 with overall experience in the representative GP Patient Survey. Its correlation with the QOF clinical score is 0.20 in the full sample and 0.63 among practices with more than 100 reviews. These associations establish that the index is an economically meaningful provider attribute and align with evidence linking UK patient surveys and NHS Choices hospital ratings to mortality and readmission rates (Greaves et al. 2012b,a). Appendix Section A.1 describes the survey and QOF measures. For expositional convenience, throughout the paper we refer to this review-based, patient-valued attribute as quality. When we mean clinical quality specifically, we say so and use measures such as QOF.

The first panel of Appendix Table A-1 reports summary statistics for the rating system. Columns 1–2 cover the visible-star period while columns 3–4 cover the later period without visible star ratings. Our sample includes 7,635 unique GP practices, over 18 million GP$\times$LSOA$\times$quarter observations and over 350,000 individual reviews. During this period, the mean GP has an index (r_{jt}) of 3.2 stars. There is substantial dispersion in ratings across practices: the standard deviation across practices is 1.0 stars and over 10% of GPs have 2

[6]We collected individual reviews and ratings for each GP practice from the NHS Choices website for the period April 2016 to June 2022. We combine this with previously collected data for the period May 2013 to April 2016. See Kowalski (2017).

stars or less. Appendix Figure A-2 shows that the distribution of average reviews r_{jt} across GP practices is smooth, including near the rounding thresholds.

Because much of our analysis examines demographic heterogeneity, one concern is that lower-income patients may rate the same practice differently. Using restricted GP Patient Survey data with respondent income terciles, Appendix Table C-1 shows that raw income differences in reported experience largely disappear after conditioning on practice fixed effects. For nearly all survey questions, within-practice differences between the highest- and lowest-income terciles are about one percentage point or less. This suggests that raw income gradients mainly reflect sorting across practices rather than systematic differences in how patients evaluate a given practice. A text-based check yields the same conclusion: sentiment maps into star ratings similarly for practices serving high- and low-income groups.[7]

A further concern is manipulation or fake reviews (as documented in other review systems, e.g. Mayzlin et al. 2014; Luca and Zervas 2016). NHS reviews are government-sanctioned and the NHS collects reviewer information, including email and IP addresses, and can withhold or remove suspicious comments.[8] We treat manipulation as an empirical question and test for it directly in Section 4.3.

Neighborhood-Level GP Enrollment. We match the review and star rating data with enrollment data for the universe of GP practices in England (NHS 2022). For each GP, we observe quarterly enrollment by Lower Super Output Area (LSOA). LSOAs are detailed geographic areas with an average of about 700 households. Because enrollment stocks are persistent, our primary outcome is net new enrollment, defined as the quarterly change in GP-LSOA enrollment.

We merge on LSOA characteristics from the 2019 English Indices of Deprivation and the Office for National Statistics Census 2021, including income, disability share, health, education, and employment.[9] The GP enrollment files also report registered patients by age

[7]As shown in Appendix Figure C-1, the relationship between positive and negative sentiment in the review text and the chosen star rating is nearly identical for GPs serving high and low-income groups.

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

[9]Our measure of neighborhood income is defined as the negative of the income deprivation score. For details on how income deprivation is computed, see `https://www.gov.uk/government/statistics/english-indices-of-deprivation-2019`. Because the main income measure is an income-deprivation index rather than earnings, we also benchmark it against available LSOA mean-income estimates in Greater London. In that sample, mean LSOA income ranges from roughly £20,000 at the bottom of the distribution to more than

bin, which we use to construct average age for each LSOA-quarter. Unless otherwise noted, demographic variables in the descriptive and reduced-form analyses are measured at the LSOA level.

Appendix Table A-1 shows summary statistics on enrollment and patient demographics. The average practice has about 8,000 patients. Mean enrollment at a practice from a given LSOA is 58 patients, and grows by 0.17 patients per quarter on average. The characteristics of registered patients reflect those of the English population, consistent with virtually all individuals in England being registered with a GP practice. The median individual lives only 1.4 km from their chosen GP, consistent with a preference for nearby practices.[10]

Individual-level GP Enrollment for Movers. As an additional data source on enrollment, we obtained restricted individual-level enrollment records from the NHS's Data Access Request Service (DARS). The data include all individuals who changed their GP and changed their address over the period May 2015 to December 2019. The records include an individual identifier, the mover's age in the month of the switch (in 5 year bins), gender, the month of the switch, the chosen GP, and the patient's destination LSOA. These individual-level data allow us to identify those individuals that had to make an active choice because they changed address and registered with a new practice. To emphasize this distinction, we refer to individuals in these data as *movers*. We present summary statistics for this sample of movers in Appendix Table D-1.

GP characteristics. We also combine the enrollment data with quarterly data on GP characteristics, namely the number of practitioners, mean experience of practitioners, and the age of the practice. These data were obtained from NHS Digital. We describe GP characteristics and other supplemental data sources in Appendix Section A.1. We geocode addresses for all GPs, allowing us to calculate the distance from each GP to the centroid of each LSOA.

Estimation Sample and Choice Sets. For our regression discontinuity approach, we use the neighborhood-level GP enrollment data and show robustness results using movers. For the

£140,000 at the top.

[10] Appendix Figure A-3 shows a histogram of patient distance to GPs and the location of GP practices in England.

structural model, we leverage the individual-level data which consists of movers making an active choice. We focus on those age 25 and older and drop person-month observations with multiple address changes or multiple GP changes so that each observation corresponds to a well-defined relocation and registration decision. Because the model requires constructing an individualized choice set for each mover, for computational tractability we focus the estimation on the Greater London area, which still yields more than 1 million observations.

For each mover, the inside options are the GP practices whose catchment areas include the destination LSOA. Since patients in England may register outside their local catchment area during our sample period, our outside option includes registration at any practice outside the catchment area. Appendix Table D-1 reports summary statistics for the resulting estimation sample and choice sets. In the mover sample, income, disability, and other neighborhood characteristics are assigned from the destination LSOA, while age comes from the individual enrollment record.[11]

2.3 Heterogeneity in Choice

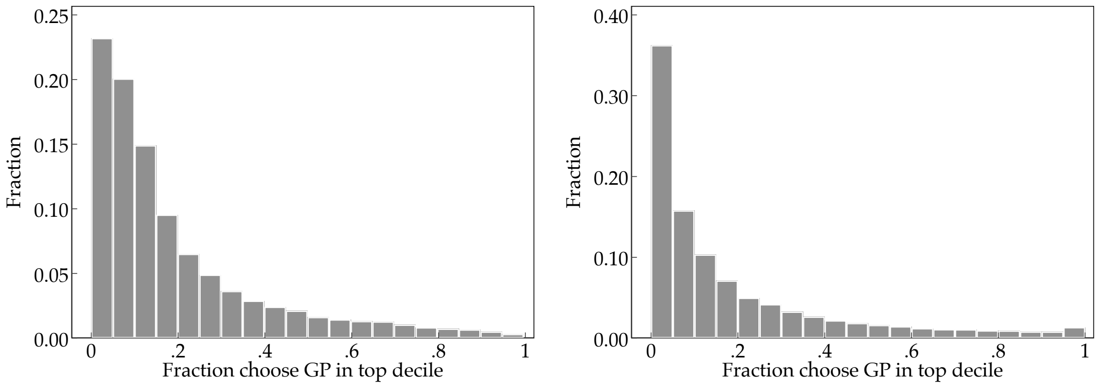
We document pronounced heterogeneity in sorting toward high-quality GPs—as measured by both patient ratings and a proxy for clinical quality—across places and demographic groups. We define a GP as high quality for a patient if it is in the top decile within that patient's local choice set. We construct this measure using two indices: the review index that determines the website star rating and the clinical quality score from the Quality and Outcomes Framework (QOF). Because high quality is defined relative to each patient's local choice set, the measure captures sorting behavior rather than differences in access to better or worse local options.

Figure 1 shows large variation in sorting intensity across neighborhoods (LSOAs). In more than 20% of LSOAs, less than 5% of people enroll at a GP in the top decile of quality, as measured by reviews. The percentage is even higher when quality is measured using the QOF clinical score. Conversely, there is a significant fraction of neighborhoods in which more than half of individuals enroll at a top-decile practice by either measure.

Panel (a) of Figure 2 shows a steep gradient in sorting intensity by income and age: the

[11] Disability share is defined as the fraction of LSOA residents whose day-to-day activities are limited a little or a lot by long-term physical or mental health conditions or illnesses.

Figure 1
Histogram of Probability of Choosing a Top-Decile GP
Within the Local Choice Set across Neighborhoods



a. Star Rating b. Clinical Quality Score (QOF)

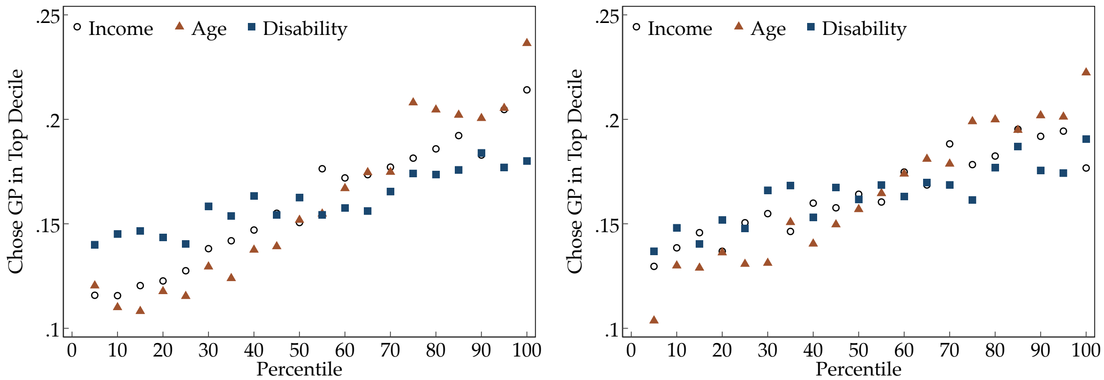
Notes: Figure shows a histogram of the probability that individuals are enrolled at a GP in the top decile of quality in their catchment area across LSOAs. Panel (a) measures quality using subjective patient reviews which determine the star rating on the NHS website. Panel (b) measures quality using objective clinical quality measures from the Quality and Outcomes Framework (QOF).

probability of enrolling with a top-decile GP rises sharply with neighborhood income and with neighborhood age. Individuals in higher-income neighborhoods are about twice as likely to enroll at a high-quality GP. We also observe a positive, though noisier, relationship by disability status.

When we replace average reviews with the QOF clinical score, the same gradients appear: patients in higher-income and older LSOAs are more likely to register with GPs in the top decile of QOF within their local choice sets. This can be seen in Panel (b) of Figure 2.

Several mechanisms could generate these facts. One is preference heterogeneity over other GP attributes such as convenience, language, or wait times for appointments. Another is information: some patients may hold imprecise beliefs about GP quality and rely differentially on public signals. The next section develops a simple framework showing how these two forces have distinct empirical implications in the presence of rounded star ratings, which we explore in the rest of our analysis.

Figure 2
Probability of Choosing a Top-Decile GP Within the Local Choice Set
by Demographic Group



a. Star Rating b. Clinical Quality Score (QOF)

Notes: Figure shows the probability that individuals are enrolled at a GP in the top decile of quality in their catchment area for each quantile of LSOA income, age, and disability share. Panel (a) measures quality using subjective patient reviews which determine the star rating on the NHS website. Panel (b) measures quality using objective clinical quality measures from the Quality and Outcomes Framework (QOF).

3 A Learning Model of GP Demand

Our central question is whether differences in information are an important driver of patient choice. The challenge is that heterogeneity in observed choices can reflect either heterogeneity in preferences for quality or heterogeneity in beliefs about quality (or both). In this section, we present a simple learning model that clarifies how a coarse public rating—such as the half-star categories displayed on the NHS website—interacts with private signals to shape demand. The model makes clear that heterogeneity in information and preferences have distinct empirical implications. The structural model in Section 5 builds on the simple model presented in this section by incorporating a more detailed model of belief updating.

3.1 Beliefs and Expected Quality

Let i index individuals and j index GP practices. Let $\tilde{r}_j$ denote latent quality of practice j. The public signal is the displayed star rating s_j. Practices that display the same star rating are said to be within the same *"star-bin"*. To fix ideas, we begin with the case in which latent quality conditional on the displayed star rating is normally distributed:

$$\tilde{r}_j \mid s_j \;\sim\; \mathcal{N}\big(\mu_{s_j},\, \eta^2\big), \qquad \mu_s \equiv \mathbb{E}_{\tilde{r}_j}\big[\tilde{r}_j \mid s_j = s\big]\,,$$

where μ_s is the within-star-bin mean and η^2 captures within-bin dispersion in quality. For simplicity, we assume η^2 is constant across star bins.

Each individual receives a private signal about $\tilde{r}_j$:

$$\tilde{z}_{ij} = \tilde{r}_j + \sigma_i \varepsilon_{ij}, \qquad \varepsilon_{ij} \sim \mathcal{N}(0,1),$$

where σ_i^2 is an individual-specific parameter capturing the information variance for that individual (e.g. noisier internet research, weaker social networks, or other frictions).

Bayes' rule implies a posterior mean that is a convex combination of the private signal and the bin mean:

$$b_{ij} \equiv \mathbb{E}_{\tilde{r}_j}\big[\tilde{r}_j \mid \tilde{z}_{ij}, s_j\big] \;=\; \omega_i\, \tilde{z}_{ij} + \big(1-\omega_i\big)\mu_{s_j}. \tag{1}$$

where ω_i is the weight on the private signal, given by

$$\omega_i \;\equiv\; \frac{\eta^2}{\sigma_i^2 + \eta^2}. \tag{2}$$

Individuals with more precise private signals (smaller σ_i^2) place more weight on their private signal $\tilde{z}_{ij}$ and less on the coarse public star category s_j.

3.2 Utility and Choice

Patients are risk-neutral and value expected quality. Let β_i^q denote individual i's preference for quality and let ν_{ij} be an idiosyncratic Gaussian taste shock. Abstracting from other observed practice attributes, expected utility from practice j is

$$\mathbb{E}[u_{ij}] \;=\; \beta_i^q\, b_{ij} + \nu_{ij} \;=\; \beta_i^q\, \omega_i\, \tilde{z}_{ij} \;+\; \beta_i^q\, (1-\omega_i)\, \mu_{s_j} \;+\; \nu_{ij}. \tag{3}$$

Writing $\tilde{z}_{ij} = \tilde{r}_j + \sigma_i \varepsilon_{ij}$ yields the decomposition

$$\mathbb{E}[u_{ij}] \;=\; \underbrace{\beta_i^q \big(\omega_i \tilde{r}_j + (1-\omega_i)\mu_{s_j}\big)}_{\delta_{ij}} \;+\; \underbrace{\beta_i^q \sigma_i \omega_i}_{\kappa_{ij}} \,\varepsilon_{ij} \;+\; \nu_{ij}.$$

Thus $\kappa_{ij}\varepsilon_{ij}$ reflects individual i's residual belief error for practice j. We assume ν_{ij} is normally distributed, so choice probabilities take the multinomial probit form.

3.3 Implications for Ratings and Patient Choice

The model has sharp predictions for how demand varies with underlying quality $\tilde{r}_j$ and with the displayed star category s_j. Two reduced-form objects are particularly informative: (i) the *within-star-bin slope* of demand with respect to $\tilde{r}_j$ holding s_j fixed, and (ii) the *jump* in demand at rounding thresholds where s_j changes discretely.

Holding preferences fixed, higher information variance σ_i^2 pulls the mean of posterior beliefs more strongly toward the bin mean μ_s. As a result, demand varies less with $\tilde{r}_j$ within a star-bin, but reacts more sharply when the star category changes. In contrast, holding information fixed, a weaker preference for quality β_i^q uniformly attenuates responses to quality: both within-bin slopes and threshold jumps become smaller.

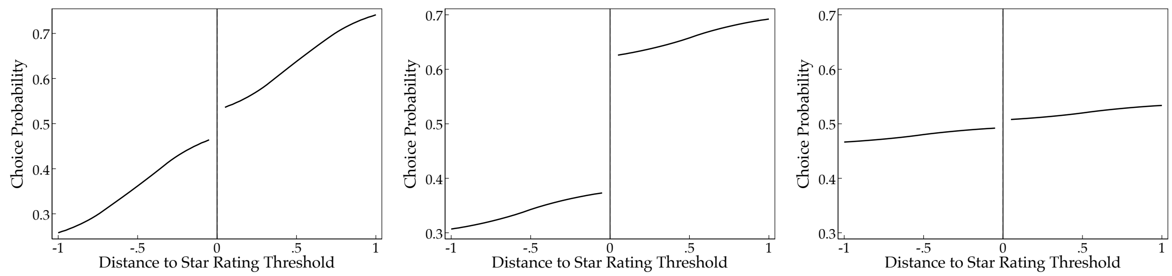
Figure 3 illustrates these mechanisms by plotting average choice probabilities against underlying quality $\tilde{r}_j$ in a two-star version of the model.[12] We simulate 1 million individuals, each with a different choice set, and average over the simulations.

Patients with more precise private signals about quality (low σ_i) exhibit a strong relationship between underlying practice quality $\tilde{r}_j$ and the choice probability, which is visible in the steep slope of the demand curve in panel (a) of Figure 3. However, because these individuals already have precise beliefs, the coarse star category provides little incremental information, so demand changes only modestly at the threshold.

Individuals with less precise private signals about quality (high σ_i) show the opposite pattern: demand varies relatively little with $\tilde{r}_j$ within a star-bin, but responds sharply when the star category changes, as in panel (b). This pattern is distinct from panel (c), which holds

[12]For expositional clarity, the simulation uses a binary public signal $s_j \in \{0,1\}$ with $s_j = \mathbb{1}\{\tilde{r}_j > 0\}$ and summarizes the within-bin distribution of $\tilde{r}_j$ with a Gaussian approximation. Section 5 replaces this approximation with the exact truncated-normal updating implied by half-star-bins and allows the informativeness of the public signal to vary with the number of reviews.

Figure 3
Simulated Demand Response to Star Rating Threshold



a. High signal precision
$\sigma_i = 0.001, \beta_i^q = 1$

b. Low signal precision
$\sigma_i = 0.5, \beta_i^q = 1$

c. Low preference for quality
$\sigma_i = 0.001, \beta_i^q = 0.1$

Notes: Simulated demand as a function of underlying quality when individuals observe a coarse public star signal. We assume individuals choose between two GPs with $\tilde{r}_j \sim \mathcal{N}(0, 1)$ and a binary star signal $s_j \in \{0, 1\}$, where $s_j = 1$ if $\tilde{r}_j > 0$ and 0 otherwise. Panel (a) shows the case with precise private signals, panel (b) shows the case with noisy private signals, and panel (c) shows the case with a low preference for quality.

information precision fixed at a high level but reduces the preference for quality (low β_i^q). In that case, both the within-bin slope and the threshold jump are small.

Our data contain the continuous review index r_{jt}, which determines s_{jt}. In the model, the public signal we use for identification is the rounded star rating, together with the displayed review count; other information is captured by private signals. Therefore, the combination of within-bin slopes and threshold jumps allows us to disentangle heterogeneity in preference for quality from heterogeneity in information precision. This insight guides both the reduced-form RD estimation in Section 4 and the structural estimation in Section 5.

4 RD Evidence on Patient Responses to Star Ratings

Motivated by the simple model above, we examine how different demographic groups respond to the star ratings displayed on the NHS website. We use a regression discontinuity approach that examines jumps in demand at half-star rounding thresholds.

4.1 RD Methodology

Our strategy exploits the fact that the website rounds each practice's average review r_{jt} to the nearest half star to obtain s_{jt}. We use r_{jt} as the running variable and compare changes in enrollment just above and below the thresholds between 1.25 and 4.75.

Our main specification collapses the data and jointly estimates the average effect of crossing any threshold. Let y_{jlt} represent an enrollment outcome for practice j, LSOA ℓ, and quarter t, and let c_s be the rounding threshold just above star rating s. We stack our data by defining our running variable as $r^0_{jt} = r_{jt} - c_s$ (the distance between the average review and the relevant threshold). We start by estimating a parametric RD specification given by

$$y_{j\ell t} = \alpha + \tau \mathbb{1}\{r^0_{jt} > 0\} + \beta r^0_{jt} + X_{jt}\gamma + \varepsilon_{j\ell t}, \tag{4}$$

where $\mathbb{1}\{r^0_{jt} > 0\}$ is an indicator for being above the threshold and X_{jt} is a vector of covariates, namely practice age, the number of reviews, and cutoff fixed effects. The parameters of interest are τ (the effect of crossing the threshold) and β (the slope away from the threshold). We also consider a non-parametric local linear approach following Cattaneo et al. (2019). [13]

Our key identification assumption for τ is that the relevant average potential outcomes functions are continuous at each threshold.[14] This may fail if there is endogenous sorting at rounding thresholds, or if characteristics of GPs change sharply at the thresholds. We examine this assumption in Section 4.3.

For our RD analysis, our primary outcome variable is the change in enrollment at the GP-LSOA-quarter level. This measure of demand allows us to capture enrollment flows rather than the level, which is heavily influenced by GP choices in the past. 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%. We also exclude practices with fewer than 5 reviews in our primary RD analysis since the evidence is consistent with patients placing less weight on star ratings when the website shows they are based on very few reviews.[15] In Section 5, by contrast, we retain these practices and let the displayed review count govern how informative the public signal is.

We proceed in three steps. First, we show that rounded star ratings impact enrollment

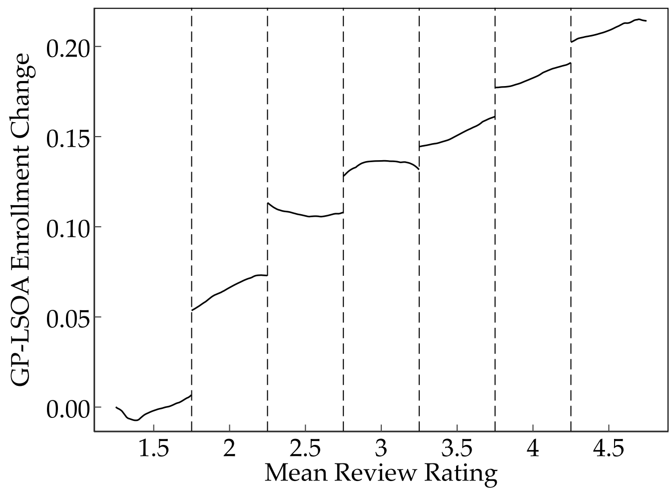
[13]We use a triangular kernel, select symmetric MSE-optimal bandwidths and cluster our standard errors at the GP level using a plug-in residual approach. Unless otherwise noted, we follow Calonico et al. (2019) and the earlier Calonico et al. (2014) when including covariates, selecting bandwidths, and computing standard errors. We refer to these bandwidths as CCT bandwidths.

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

[15]Average ratings are less correlated with other measures of quality when the number of reviews is small (see Appendix Table A-2) and we find little evidence of a discontinuity in demand for GPs with fewer than 5 reviews.

on average. Second, we show that there is important heterogeneity by demographic characteristics, and that this is consistent with higher-income groups having more precise private information about quality. Finally, we discuss potential alternative mechanisms and robustness.

Figure 4
GP Demand and Review Thresholds



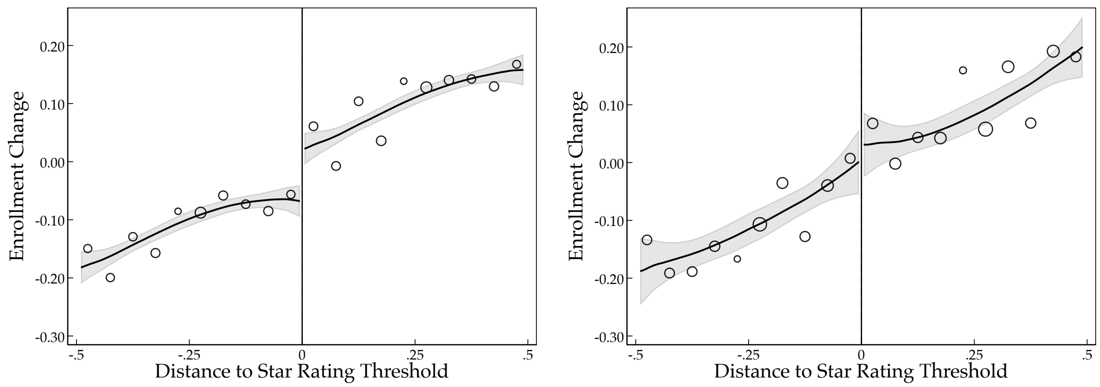
Notes: Relationship between average reviews and GP enrollment change by GP-LSOA-quarter for the period when the NHS website displayed star ratings. Lines are smoothed using a local linear regression within each star rating bin. Vertical lines show thresholds for rounding star ratings.

4.2 RD Results

Average Impact of Star Ratings. Figure 4 plots GP–LSOA enrollment changes against the RD running variable (the two-year average review index), with separate local linear fits within each star-bin. Enrollment is discontinuously higher at each half-star threshold, consistent with patients responding to the rounded star ratings displayed on the site. Within each star-bin, demand rises with the mean review, which suggests that patients also value underlying quality even when the displayed star is unchanged.

Figure 5 panel (a) zooms in on the threshold design by stacking all thresholds. Enrollment change is approximately linear in the distance to the cutoff on both sides, and there is a clear jump at the threshold. Parametric estimates in Table 1 imply that crossing a half-star threshold increases quarterly enrollment change from an LSOA by about 0.08 patients. The mean LSOA-to-GP quarterly change is only 0.17 since LSOAs are quite small, so the effect is

Figure 5
Effect of Star Rating Threshold on GP Enrollment
Before and After Website Change



a. Visible star ratings on website

b. No star ratings on website

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

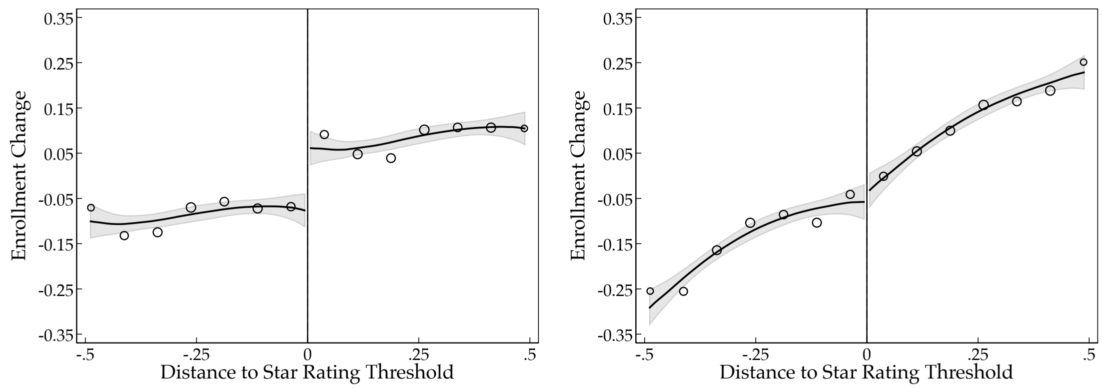
about 47% of the mean. Nonparametric specifications yield similar results, and the effect is statistically significant.

A falsification test supports the interpretation that the discontinuity is driven by the star display rather than unmodeled features of reviews. In the period after the NHS website stopped showing rounded star ratings (beginning January 2020), we find no detectable discontinuity even though slopes on either side of the threshold remain similar.[16] Table 1 shows no detectable jump during the no-stars period, in line with panel (b) of Figure 5.

Differences by Demographics. We next study how responses vary across groups. Separately analyzing LSOAs with above and below median income yields a clear pattern. Low-income neighborhoods exhibit a pronounced discontinuity at each half-star threshold, while high-income neighborhoods show little or no jump. Column (4) of Table 1 implies that crossing a threshold raises quarterly enrollment changes by about 0.15 patients in below-median income LSOAs and by about 0.02 in above-median income LSOAs; the difference is statistically significant. Figure 6 shows the corresponding binned RD plots. Appendix Figure B-4 shows that this pattern is broadly monotone across income quartiles.

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

Figure 6
Effect of Star Rating Threshold by Income



a. Low-Income b. High-Income

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

The within-bin slopes are also informative. In high-income LSOAs, demand varies steeply with the underlying review index on both sides of the cutoff, even though the threshold jump is small. In low-income LSOAs, within-bin slopes are flatter and the threshold jump is larger. In the language of the model, this pattern is consistent with higher-income groups having more precise private information and therefore relying less on the public star signal.

We see similar patterns for other local demographics. Table 1 and Appendix Figure B-1 indicate larger discontinuities for LSOAs with younger patients, and for LSOAs with a higher disability share. We examine heterogeneity for a wide array of additional observable characteristics in Appendix Table B-1. In general, characteristics indicating that individuals are of higher socioeconomic status are associated with weaker responses at the thresholds, consistent with more precise private signals.

These results use the neighborhood-level data for all patients in England registered with a GP, but they are robust to using the individual-level mover data. Movers are useful because they must actively choose a new GP. We assign each mover the characteristics of the destination LSOA. In this sample, low-income movers display a sizable jump at thresholds, whereas high-income movers place more weight on within-bin variation in reviews. The jump and slope estimates by income, age, and disability mirror the baseline. Appendix Table B-6 reports the corresponding estimates. These results imply that endogenous switching

Table 1
Effect of Star Ratings on GP Demand
Regression Discontinuity Estimates

	CCT RD	Parametric RD				
				RD Effect Heterogeneity Visible Star Ratings		
	Visible Star Ratings	Visible Star Ratings	No Star Ratings	Income	Age	Disability
Above Threshold	0.130**	0.081**	0.022			
	(0.058)	(0.034)	(0.071)			
Above × Low				0.145***	0.134***	0.064
				(0.042)	(0.045)	(0.043)
Above × High				0.019	0.021	0.088**
				(0.041)	(0.039)	(0.041)
P-Value	0.026	0.016	0.754			
CCT Robust P-Value	0.038					
High/Low Diff P-Value				0.010	0.030	0.640
Bandwidth	0.13	0.25	0.25	0.25	0.25	0.25
N	916,688	1,621,667	405,476	1,621,667	1,573,646	1,573,646

Notes: Dependent variable is the change in enrollment at the GP-LSOA-quarter level. Visible star ratings indicate the period May 2015 to December 2019. No star ratings indicate the period January 2020 to June 2022. Low (high) is defined as LSOAs below (above) the median with respect to the demographic characteristic. Column 1 uses local linear regressions with triangular kernels and follows Calonico et al. (2019) and Calonico et al. (2014) to select bandwidths and calculate robust standard errors. Columns 2 to 6 use a parametric linear regression specification. Specifications control for practice age, number of reviews, and cutoff fixed effects. Standard errors are clustered at the GP-practice level and shown in parentheses.

is not driving the main findings.

Taken together, the heterogeneity points to information as the key channel: the star display shifts choices most for groups with less precise private signals, while groups with more precise private signals respond mainly to the underlying continuous measure of quality.

4.3 Robustness and Potential Alternative Mechanisms

Endogenous sorting around the threshold. Our key identification assumption is that the relevant average potential outcomes functions are continuous at each threshold. This may fail if there is endogenous sorting of GPs at rounding thresholds, or if the observable or unobservable characteristics of GPs change sharply for other reasons. For instance, a number of GPs might exert effort to stay just above a threshold.

We perform manipulation tests in the spirit of McCrary (2008) to help rule out jumps in the distribution of the running variable r_{jt} across the threshold. We implement the tests outlined in Cattaneo et al. (2018) based on local polynomial density estimators and find no evidence of a discontinuity in the density of average reviews at the threshold. We present our results in Appendix Figure B-2. This suggests that GP practices in our sample did not differentially manipulate reviews (either falsely or through the provision of effort) to gain a higher star rating.

We test the smoothness of covariates across the thresholds to help rule out discrete jumps in the observable features of GPs above versus below each threshold. Appendix Figure B-3 shows that observable characteristics of GPs are continuous at rounding thresholds.[17] 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. All five covariates are smooth through the threshold, with small t-statistics, implying no discrete change in observable practice characteristics.

Capacity constraints and RD estimates. A concern is that practices just above a threshold may register additional patients, leading to increases in waiting times that offset informational gains, especially in areas where demand is already high (e.g., Chartock 2025). We do not find evidence of such sharp local congestion responses. Appendix Figure B-3 shows the share of survey respondents reporting that they waited more than one week for an appointment is smooth through the thresholds, even though enrollment jumps discretely.

A related concern is that differential capacity constraints explain the results. To explain the lack of a discontinuity for high-income groups, it would have to be the case that they face more capacity constraints. In Appendix Table B-5 we find that the enrollment RD is similar when we compare high- and low-income groups living in close proximity, who therefore face similar local menus of practices and capacity constraints. Appendix Table B-5 also shows similar results when we exclude practices that appear plausibly capacity constrained according to our proxy based on unusually static enrollment.

[17] For our tests, we implement the non-parametric RD methodology described above at the GP level with five different GP characteristics on the left-hand side: (i) the fraction of patients that report having to wait one week or more for an appointment (in survey data) (ii) the number of months the GP has been active, (iii) a survey-based outcome of patient trust, (iv) the QOF clinical score, and (v) the payments each GP receives per-patient from the NHS.

Taken together, these exercises make it less likely that the threshold estimates are driven by discontinuous changes in crowding or access. We emphasize, however, that this evidence should be interpreted carefully. Our data do not allow us to observe realized appointments at the patient level, so we cannot rule out congestion responses away from the threshold or longer-run supply-side adjustment.

Robustness to London sample. We examine the results separately for Greater London and the rest of England. Appendix Table B-2 shows similar results by income and age. The magnitudes of the estimates differ when the sample is split by disability, but the standard errors are large. Given these similarities, for computational feasibility, our demand model below focuses on the Greater London area.

Alternative identification strategies. We also find similar results using a complementary panel fixed-effects strategy that exploits within-GP changes in the rounded star rating s_{jt} over time. These changes occur as new reviews are added, or as older reviews are excluded from the two-year moving average. We implement this approach by regressing enrollment changes on the star rating and the star rating interacted with LSOA income, while controlling for GP fixed effects, quarter×year fixed effects, and time-varying GP controls. While this strategy relies on stronger identification assumptions than our RD approach, we still find that the low-income group is more responsive to *changes* in star ratings. We present the estimation details and results in Appendix Section B.2.

4.4 Mechanisms and Survey Evidence

The results raise the question of *why* some individuals have more precise private signals than others. To shed light on this issue, we fielded a short online survey of adults registered with a GP in England and asked how they gathered information when choosing their current practice. We consider three broad channels: access to social information, differences in search effort, and differences in preferences.[18] We present the results in Appendix Table C-2.

Two patterns emerge. First, social information sources differ by income. Respondents in

[18]The survey was conducted through Amazon Mechanical Turk from May to November 2024 with standard quality screens. We provide a description of the survey methodology and questions in Appendix Section C.3.

high-income households report consulting more people before choosing a GP (3.53 versus 2.30 among low-income households; $p = 0.035$). They also report a larger share of friends or family who work in health care (2.63 versus 1.76; $p = 0.026$). These gaps are consistent with richer networks that can transmit information about relative practice quality.

Second, search intensity looks similar across income groups. The share listing any website as a source is nearly identical, as is the number of distinct sources used. Reported hours researching and the number of GP websites visited do not differ in a statistically meaningful way. Taken together, the survey suggests that income gradients in information are better explained by social networks than by the use of online resources.[19] While our survey is a small convenience sample rather than a representative survey, we note that the results are broadly consistent with earlier survey evidence on GP ratings that finds low-income patients are more likely to use doctor-rating websites (Galizzi et al. 2012).

Finally, we ask whether preferences differ across groups. Low- and high-income respondents assign similar importance to clinical quality, distance, waiting times, and friendliness while differences by health are modest. These results largely agree with Appendix Table C-1 showing that low- and high-income groups rate a given GP in a similar way, suggesting similar preferences.

5 Empirical Model of Demand

While the reduced-form evidence highlights differential information, it does not allow us to quantify its relative importance compared to differences in preferences. To address this, this section develops and estimates a model of GP demand that accounts for heterogeneity in preferences, information and access.

Individuals observe a public star signal and a private signal of quality, where we once again consider the dimension of quality defined by reviews. Importantly, the precision of the private signal is allowed to vary. The model also accommodates heterogeneity in preferences over GP attributes and incorporates GP congestion. Our estimation approach leverages the

[19] We also compare responses by self-reported health. Those in better health consult more individuals, call or visit more practices, engage more with online forums, and consider more options, although differences in website use per se are modest. These patterns again point to variation in the reach of informal information rather than systematic differences in reliance on the NHS website.

same variation in the public signal caused by the rounded star ratings as the regression discontinuity approach in Section 4. We then use the structural parameter estimates to quantify the roles information and preferences play in shaping demand heterogeneity, and to evaluate counterfactuals, including policies related to the provision of information.

5.1 Empirical Model Setup

This section embeds the learning framework from Section 3 into a richer demand system. The core mechanism is unchanged: patients choose a GP based on posterior expected quality, which combines a public signal with a private signal. Relative to the stylized model, the empirical specification adds three conceptual features. First, the informativeness of the public signal varies with the number of reviews available for each GP. Second, beliefs use the exact posterior implied by observing the rounded half-star bin rather than a normal approximation. Third, utility depends on distance, congestion, and practice fixed effects.

Sample and Choice Sets. We estimate the structural model on the mover sample described in Section 2 which consists of patients age 25 and over who move residence and register with a new GP in the same month between May 2015 and December 2019. For computational tractability, we restrict the structural estimation to movers in Greater London, while retaining a large sample of more than one million observations. As shown in Section 4, the reduced-form patterns are similar in Greater London and in England as a whole.

For a mover i who changes LSOA in quarter t, the choice set J_{it} consists of inside options J_{it}^I and a composite outside option $j = 0$. The inside options are the GP practices whose catchment areas include the mover's destination LSOA. The outside option captures registration at a GP practice outside this catchment-based set. Appendix D.1 describes the exact choice-set construction, including the treatment of LSOAs with more than 30 catchment practices, and Appendix Table D-1 reports summary statistics for the resulting estimation sample and choice sets.

Latent quality and public information. Latent practice quality $\tilde{r}_j$ is assumed to be fixed over the estimation period, so time variation in the website's signal reflects finite-sample review noise rather than changes in underlying quality. We use the long-run average of observed individual reviews as a proxy for latent quality. A practice fixed effect absorbs persistent demand shifters not captured by observed covariates or by this review-based quality proxy.

Patients hold a Gaussian prior, $\tilde{r}_j \sim \mathcal{N}(m_0, s_0)$, where m_0, s_0 denote the prior mean and variance. In each quarter t, the NHS website calculates a continuous review-based rating r_{jt} based on the previous two years of n_{jt} individual reviews. We model this index as a noisy signal of $\tilde{r}_j$:

$$r_{jt} \mid \tilde{r}_j, n_{jt} \sim \mathcal{N}(\tilde{r}_j, \varsigma_{jt}^2), \qquad \varsigma_{jt}^2 \equiv \frac{\sigma_r^2}{n_{jt}}. \tag{5}$$

The same displayed star rating is more informative (ς_{jt} smaller) when it is based on many reviews than when it is based on few. For details, see Appendix D.2.

For instance, if the individual observes a GP with 3.5 stars, then they know that the continuous review-based rating lies in the interval $\mathcal{B}(3.5) = [3.25, 3.75)$.

Private signals. Each mover i also receives a private signal about j,

$$\tilde{z}_{ij} = \tilde{r}_j + \sigma_i \varepsilon_{ij}, \tag{6}$$

where $\varepsilon_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1)$. We assume $\{\varepsilon_{ij}\}_{j \in J_{it}^I}$ are independent across practices and independent of the public-review noise in Equation (5) conditional on the prior $\tilde{r}_j$. As in Section 3, smaller σ_i^2 means more precise private signals about quality and therefore less reliance on the displayed stars.

Belief updating. Since both are Gaussian, combining the prior $\tilde{r}_j$ with the private signal $\tilde{z}_{ij}$ yields the normal posterior $\tilde{r}_j \mid \tilde{z}_{ij} \sim \mathcal{N}(m_{ij}, v_{ij})$, with mean and variance

$$m_{ij} = v_{ij} \left(\frac{m_0}{s_0} + \frac{\tilde{z}_{ij}}{\sigma_i^2} \right), \qquad v_{ij} = \left(\frac{1}{s_0} + \frac{1}{\sigma_i^2} \right)^{-1}. \tag{7}$$

However, there is an additional round of belief updating, because individuals also observe the public star rating s. Intuitively, the individual observes a coarsened Gaussian public signal whose variance decreases with the number of reviews n_{jt}. The coarsening arises because individuals only see which interval this signal falls into, namely $\mathcal{B}(s)$. The star rating nudges beliefs up or down depending on which half-star bin the (hidden) review average is in, relative to the individual's beliefs $\tilde{r}_j \mid \tilde{z}_{ij}$. The size of that nudge depends on n_{jt}.

The individual has prior $\tilde{r}_j \mid \tilde{z}_{ij}$ and updates based on this coarsened Gaussian signal. The posterior distribution is not Gaussian in general, but its mean,

$$b_{ijt} \equiv \mathbb{E}_{\tilde{r}_j}\left[\tilde{r}_j \mid \tilde{z}_{ij}, s_{jt}, n_{jt}\right]$$

has a closed-form expression, derived in Appendix D.2, equation (D-1). This is the empirical analogue of Equation (1) in the simple model in Section 3. In the simple model, beliefs are simply a convex combination of the private signal and the star-bin mean. Here beliefs take a similar form, but the informativeness of the public signal varies with the number of reviews n_{jt}. When n_{jt} is large, the star-bin acts like a tight restriction on beliefs; when n_{jt} is small, posterior beliefs remain close to m_{ij}. Equation (D-1) shows that the posterior mean equals m_{ij} plus a correction that accounts for these factors.

Local approximation. For computational tractability, when computing the model's likelihood, we use a first-order Taylor approximation to posterior expected quality as a function of the private signal. The posterior mean can be written $b_{ijt}(\tilde{z})$ to emphasize its dependence on the private signal. Define the local pass-through of the private signal by

$$\omega_{ijt}(\tilde{z}) \equiv \frac{\partial b_{ijt}(\tilde{z})}{\partial \tilde{z}}.$$

For the likelihood approximation, we evaluate this derivative at the expansion point $\tilde{z} = \tilde{r}_j$ and write $\omega_{ijt} \equiv \omega_{ijt}(\tilde{r}_j)$. Using Equation (6), the first-order expansion is

$$b_{ijt}(\tilde{z}_{ij}) \approx b_{ijt}(\tilde{r}_j) + \omega_{ijt}\left(\tilde{z}_{ij} - \tilde{r}_j\right) = b_{ijt}(\tilde{r}_j) + \omega_{ijt}\sigma_i\varepsilon_{ij}.$$

Appendix D.2 shows that $\omega_{ijt}(\tilde{z})$ has a closed form. In the simple model, the pass-through of the private signal is constant within a star bin.[20] Here, it varies with review counts and

[20] This definition of ω_{ijt} is the analogue, in this more complex model, of Equation (2), where $\omega_i \equiv \frac{\partial b_{ij}}{\partial \tilde{z}_{ij}} = \frac{\eta^2}{\sigma_i^2 + \eta^2}$.

with the mover's beliefs relative to the rounding thresholds. The term ω_{ijt} can be interpreted as measuring the importance of patient i's private information about j in period t.

The approximation separates a deterministic component of beliefs—given the mover, the practice, and the public signal—from residual belief heterogeneity generated by idiosyncratic private information. Appendix D.2 gives the exact posterior, the closed-form expression for $\omega_{ijt}(\tilde{z})$, and the expansion point used in estimation.

Utility. Let Z_i denote a vector of observable mover and destination-neighborhood characteristics, including a constant, and let X_{ijt} include observed non-quality attributes—distance, congestion or capacity, and the outside-option indicator—interacted with Z_i. Let ξ_j denote a GP practice fixed effect for inside options, with $\xi_0 = 0$. For an inside option $j \in J^I_{it}$, mover i's full-information utility in quarter t is

$$u^{\mathrm{FI}}_{ijt} = \beta^q_i \tilde{r}_j + X'_{ijt}\beta + \xi_j + \nu_{ij}.$$

Given posterior beliefs, expected utility when making a choice is given by

$$u_{ijt} = \mathbb{E}_{\tilde{r}_j}\left[u^{\mathrm{FI}}_{ijt} \mid \tilde{z}_{ij}, s_{jt}, n_{jt}\right] = \beta^q_i\, b_{ijt}(\tilde{z}_{ij}) + X'_{ijt}\beta + \xi_j + \nu_{ij}.$$

The deterministic component of the posterior expected quality at the expansion point is $b_{ijt}(\tilde{r}_j)$. Substituting the local approximation into expected utility gives

$$u_{ijt} \approx \beta^q_i \left[b_{ijt}(\tilde{r}_j) + \omega_{ijt}\sigma_i\varepsilon_{ij}\right] + X'_{ijt}\beta + \xi_j + \nu_{ij} = \delta_{ijt} + \kappa_{ijt}\varepsilon_{ij} + \nu_{ij}, \tag{8}$$

where $\nu_{ij} \sim \mathcal{N}(0, \sigma^2_\nu)$. We define

$$\delta_{ijt} \equiv \beta^q_i\, b_{ijt}(\tilde{r}_j) + X'_{ijt}\beta + \xi_j, \qquad \kappa_{ijt} \equiv \beta^q_i \omega_{ijt}\sigma_i.$$

We assume the taste shocks ν_{ij} are independent across options and independent of the belief shocks ε_{ij}. For the outside option, we set $b_{i0t} = 0$, $\omega_{i0t} = 0$, $\xi_0 = 0$, and $\varepsilon_{i0} = 0$, so its utility is captured entirely by the outside-option terms in X_{i0t} and by ν_{i0}. Relative to the simple model, both the deterministic posterior mean entering δ_{ijt} and the residual belief dispersion κ_{ijt} vary with the displayed review count n_{jt} and with the practice's position relative to the rounding thresholds. Appendix D.1 describes the construction of the demographic variables and choice sets.

Heterogeneity. Observable heterogeneity enters in three places. First, the marginal utility of perceived quality is $\beta_i^q = \exp(Z_i'\theta_{\text{qual}})$, which constrains the quality coefficient to be positive. Second, private-signal noise is parameterized as $\sigma_i^2 = \exp(Z_i'\theta_{\text{prec}})$, so a larger value of $Z_i'\theta_{\text{prec}}$ corresponds to a noisier, less precise private signal. Third, the coefficients on non-quality attributes vary with observables through the interactions in X_{ijt}. The exact contents of Z_i vary across specifications and are described below when we present the estimates. Appendix D.2 provides details on the parameterization of heterogeneity.

5.2 Estimation

We now describe the choice probabilities implied by the model and then our penalized likelihood estimator.

Choice probabilities In Equation (8), the composite shock is the weighted sum of Gaussians $\eta_{ijt} \equiv \kappa_{ijt}\varepsilon_{ij} + \nu_{ij}$, which is independent across j and normally distributed with variance $\sigma_{u,ijt}^2 \equiv \kappa_{ijt}^2 + \sigma_\nu^2$. Hence the latent expected utilities are independently normally distributed across options, with mean δ_{ijt} and variance $\sigma_{u,ijt}^2$. The probability that option j delivers the highest expected utility is

$$P_{ij}(\theta) \;=\; \int_{-\infty}^{\infty} \frac{1}{\sigma_{u,ijt}}\phi\left(\frac{x-\delta_{ijt}}{\sigma_{u,ijt}}\right) \prod_{k\in J_{it}\setminus\{j\}} \Phi\left(\frac{x-\delta_{ikt}}{\sigma_{u,ikt}}\right)\, dx. \tag{9}$$

The expression above reflects standard choice probabilities in a multinomial probit with independent but heteroskedastic shocks. Intuitively, δ_{ijt} shifts the location of option j's latent utility distribution, while κ_{ijt} shifts the option-specific dispersion $\sigma_{u,ijt}$. Choice probabilities can be evaluated using standard 1-dimensional quadrature.

Penalized likelihood estimator that matches auxiliary RD moments. Let $y_i \in J_{it}$ denote mover i's realized choice. The average sample log-likelihood is

$$\overline{\mathcal{L}}_N(\theta) \equiv \frac{1}{N}\sum_{i=1}^{N} \log P_{i\,y_i}(\theta), \tag{10}$$

where $P_{i\,y_i}(\theta)$ is given by equation (9).[21]

Motivated by the comparative statics in Section 3.3, we augment the likelihood with the RD variation used in Section 4. Let $\hat{\tau}^{data}$ collect the reduced-form threshold jumps and within-bin slopes, including the subgroup interactions used in the empirical analysis. For any candidate parameter vector θ, we compute model choice probabilities, aggregate them to the same level as in the reduced-form sample, and apply the same RD specifications to obtain $\hat{\tau}^{model}(\theta)$. We also apply the threshold-jump specifications to the practice fixed effects and denote the resulting vector by $\Delta\xi(\theta)$. The penalized likelihood estimator is

$$\hat{\theta} = \arg\min_{\theta} \left\{ -\overline{\mathcal{L}}_N(\theta) + \sum_{k=1}^{K_{\text{mom}}} \lambda_k \left(\hat{\tau}_k^{model}(\theta) - \hat{\tau}_k^{data}\right)^2 + \lambda_{FE}\,\|\Delta\xi(\theta)\|^2 \right\}. \tag{11}$$

Appendix Section D.3 describes the auxiliary moments, the normalization of the penalty weights, and the inference procedure.

The parameters to be estimated are the preferences for quality θ_{qual}, other preferences β, GP fixed effects ξ_j, and the determinants of precision θ_{prec}. We normalize the variance of taste shocks $\sigma_v^2 = \pi^2/6$ to match the variance of the error in a standard logit. Assuming individuals have rational expectations over the distribution of GP quality, we set parameters (m_0, s_0, σ_r^2) to their empirical counterparts computed from the review microdata. For instance, m_0 is the mean individual review across all practices, s_0 is the cross-practice variance of long-run average reviews, and σ_r^2 is the within-practice variance of individual review residuals around long-run practice means.[22]

Identification. Identification follows directly from the logic in Section 3.3. Variation in threshold jumps identifies the precision of private signals: when σ_i^2 is large, posterior beliefs

[21]The normalization by N places the likelihood on a per-mover scale before it is combined with the auxiliary moments. We report the conventional summed log-likelihood separately as a model-fit statistic.

[22]In particular, we set $m_0 = 3.297$ and the standard deviation $\sqrt{s_0} = .902$. Given these parameters, only about 3% of the mass of this Gaussian distribution is outside of the interval $[1, 5]$.

are affected more strongly by the rounded public signal, which produces larger jumps at half-star thresholds and flatter within-bin slopes. This is especially true when the displayed stars are based on a large number of reviews.

Conditional on precision, the within-bin relationship between demand and the continuous rating r_{jt} identifies preference for quality, θ_{qual}. Once beliefs respond smoothly within a star-bin, steeper demand gradients with respect to that continuous rating conditional on observables reflect a stronger preference for quality rather than better information.

The remaining coefficients β are identified from substitution patterns over distance, congestion, and the outside option, as in standard demand models. Practice fixed effects ξ_j absorb persistent non-quality demand shifters and are distinct from latent quality $\tilde{r}_j$, which enters beliefs through the learning model. Time variation in the public signal coming from review noise and changes in review counts also help separately identify the precision parameters. The smoothness penalty on ξ_j prevents the fixed effects from mechanically generating the threshold discontinuities that identify information frictions.

5.3 Results and Model Fit

We report results from our empirical model in Table 2. We consider several specifications that differ in how observable heterogeneity is parameterized. The first specification allows information precision and preferences—over quality, distance and congestion—to vary only by income. The second specification allows them to vary by income, mover age, and disability. We focus on these characteristics because they summarize economically relevant dimensions of mover heterogeneity in a parsimonious way. The third specification replaces income, age, and disability with the first four principal components extracted from a wide set of demographic variables (see Appendix Table D-3). These four principal components capture 80% of the variance in the included covariates, providing a flexible summary of observable heterogeneity without imposing a specific structure on which characteristics matter.[23] This progression allows us to assess the robustness of our results to alternative ways of capturing heterogeneity while avoiding multicollinearity. Across specifications, we focus primarily

[23] Appendix Table D-3 shows the first principal component primarily loads on socioeconomic status. In addition, Appendix Table D-2 reports a summary of the principal components, including the proportion of total variance explained.

on two sets of parameters: those governing the precision of the private signal, and those governing preferences for GP quality.

Our estimates of the first key set of parameters consistently show that the variance of the private signal (σ_i^2) declines with income. In both the income-only model and the model that includes age and disability, the coefficient governing the impact of income on the private signal variance is large in magnitude, negative, and precisely estimated (the point estimates are -0.835 and -1.640, respectively). This indicates that higher income groups are better informed about quality in the absence of displayed stars. These findings align with the reduced-form RD evidence shown in Section 4: high-income neighborhoods react less at star thresholds but still respond to within-bin variation in the underlying rating. Conditional on income, precision also varies with mover age, with older movers having noisier signals on average. However, once we control for income and age, differences by disability are small and not statistically significant.

To provide a more economically interpretable characterization of these estimates, the second panel of Table 2 gives a summary of the mean and dispersion of the weight movers place on their private signal ($\omega_{ijt} \in (0,1)$). The mean weight ranges from 0.61 to 0.81 across specifications and the standard deviation of the weight is consistently large. For example, when allowing for heterogeneity by income, age, and disability, the standard deviation of the private signal weight is 0.33. This implies a nontrivial share of individuals place very little weight on the private signal and instead rely almost entirely on the displayed stars when choosing a GP, while another significant share have precise private signals and are nearly fully informed even absent the star ratings.

The second set of estimates in Table 2 indicates a strong preference for high-quality GP practices on average, with only minor differences by observable characteristics. On average, the quality coefficients are large and statistically significant: the mean quality coefficient ranges from roughly 0.4 to 0.5. Depending on the specification, our estimates imply that a patient with average demographics is willing to travel 21-26% of the mean patient-to-GP distance ($\approx 1.5\,\text{km} \times 0.25 = 0.375\,\text{km}$) for a "one-star" increase in posterior expected quality. There is modest observable heterogeneity. In the income-only model, the coefficient on income is small and is not statistically significant. In the model that includes age and disability, the coefficient on income is slightly larger in magnitude, but negative, implying

Table 2
Empirical Model Estimates

	Heterogeneity by Income		Heterogeneity by Income/Age/Disability		Heterogeneity by Principal Components	
	Mean	SE	Mean	SE	Mean	SE
Private Signal Variance:						
Constant	−2.114	(0.090)	−0.679	(0.308)	−2.008	(0.227)
× Income Score	−0.835	(0.168)	−1.640	(0.518)		
× Age			2.347	(0.625)		
× Disability			−0.097	(0.483)		
× Principal Component 1					1.466	(0.344)
× Principal Component 2					0.039	(0.370)
× Principal Component 3					−0.358	(0.312)
× Principal Component 4					−0.365	(0.295)
Preference for Quality:						
Constant	−0.702	(0.016)	−0.953	(0.038)	−0.867	(0.026)
× Income Score	0.038	(0.028)	−0.111	(0.041)		
× Age			0.011	(0.047)		
× Disability			−0.062	(0.042)		
× Principal Component 1					0.062	(0.031)
× Principal Component 2					0.060	(0.038)
× Principal Component 3					−0.017	(0.034)
× Principal Component 4					−0.033	(0.030)
Preference for Distance:						
Constant	−1.282	(0.003)	−1.278	(0.005)	−1.304	(0.005)
× Income Score	−0.027	(0.004)	−0.035	(0.005)		
× Age			0.033	(0.004)		
× Disability			−0.021	(0.007)		
× Principal Component 1					−0.030	(0.004)
× Principal Component 2					−0.036	(0.006)
× Principal Component 3					0.021	(0.005)
× Principal Component 4					0.145	(0.004)
Congestion:						
Constant	0.072	(0.008)	0.084	(0.012)	0.034	(0.013)
× Income Score	−0.056	(0.009)	−0.084	(0.011)		
× Age			0.065	(0.007)		
× Disability			−0.066	(0.011)		
× Principal Component 1					−0.025	(0.012)
× Principal Component 2					−0.093	(0.010)
× Principal Component 3					−0.025	(0.010)
× Principal Component 4					0.003	(0.009)
Private signal weight (ω_{ijt}), mean	0.809		0.612		0.775	
Private signal weight (ω_{ijt}), std	0.110		0.322		0.162	
Quality preference (β_i^q), mean	0.497		0.395		0.429	
Quality preference (β_i^q), std	0.015		0.030		0.025	
Log-Likelihood	-823,871		-831,029		-830,616	
Pseudo-R^2 (likelihood ratio)	0.423		0.418		0.418	

Notes: Table shows estimates from three versions of the model with different specifications of observable heterogeneity in information precision, preference for quality, disutility of distance, and congestion. Columns 1–2 allow heterogeneity by income; Columns 3–4 allow heterogeneity by income, age, and disability; Columns 5–6 allow heterogeneity by the first four principal components of demographic characteristics. Bootstrapped standard errors, holding nuisance parameters fixed, are shown in parentheses (Appendix Section D.3 provides details).

that lower income movers have a stronger preference for quality. Across specifications, the standard deviation of the quality coefficient is consistently less than one-tenth of the mean. This provides some initial evidence that heterogeneity in preferences may be less important than heterogeneity in information.

The remaining parameter estimates have expected signs and plausible magnitudes. For example, patients dislike practices that are farther from their LSOA, and the distance disutility is somewhat larger for high-income groups. As expected, individuals prefer GP practices with a higher number of practitioners per patient, suggesting a dislike for congestion.

The results from our principal-component specification align with the first two specifications. Even after flexibly accounting for a broad set of observable covariates, the model continues to imply substantial heterogeneity in information precision and relatively modest heterogeneity in preferences for quality. The coefficient on the first component—which loads primarily on income and socioeconomic status—is large and precisely estimated in the private-signal variance equation. The remaining components have small and statistically insignificant coefficients. Finally, the specifications have similar overall fit, as measured by the pseudo-R^2.[24]

In the analysis and examination of counterfactuals below, we focus on the specification that allows heterogeneity by income, age, and disability as our baseline. A comparison of model predictions and data is presented in Appendix Table E-2. The specification strikes a good balance between flexibility and parsimony and delivers a close match to the data. Both the RD effect at the threshold and the slope away from the threshold at the estimated parameters match the corresponding estimates from the data well. We also verify that fixed effects are smooth across the thresholds.

In the counterfactuals, our baseline scenario is an environment with no displayed summary star ratings, corresponding to the period after 2019.

[24] We report McFadden's likelihood-ratio index, computed from the raw log-likelihood at the penalized likelihood estimates. Because the estimator balances likelihood fit with RD-moment targets and a smoothness penalty on practice fixed effects, the index need not be increasing as we enrich heterogeneity.

5.4 The Relative Role of Information and Preferences as Drivers of Patient Choice

The parameter estimates provide initial evidence that information heterogeneity is more important than preference heterogeneity in shaping differences in chosen GP quality. To examine the role of information heterogeneity more formally, we use the estimated demand model to decompose patient responsiveness to GP quality into two channels: one driven by information and another driven by preferences. We summarize quality responsiveness using the within-bin semi-elasticity with respect to posterior expected quality

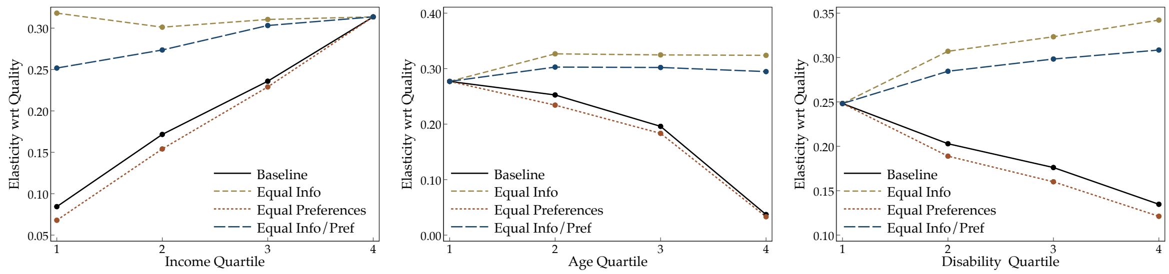
$$H_{ijt}(\theta) \equiv \beta_i^q \frac{\partial \log P_{ij}(\theta)}{\partial \delta_{ijt}}. \tag{12}$$

This varies with the value patients place on quality and the precision with which they infer it. We report the corresponding elasticity with respect to latent quality $\tilde{r}_j$, obtained by multiplying $H_{ijt}(\theta)$ by the pass-through from latent quality into posterior expected quality and then aggregating with choice-probability weights. Appendix D.4 gives the exact formula.

We use this object to decompose responsiveness across income, age, and disability. For each characteristic, we divide movers into quartiles and compute each quartile's average model-implied elasticity with respect to latent quality in four environments. First, the baseline is the no-stars environment, where patients rely on their private signals. Second, the equal-information environment, which sets each mover's information precision equal to that of movers in the top quartile of the relevant characteristic (income, age or disability), holding preferences fixed. Third, the equal-preference environment, which sets each mover's quality-preference parameters equal to those of movers in the top quartile, holding information fixed. Fourth, a final environment, which equalizes both information and quality preferences. The results are shown in Figure 7.

We find that, in the baseline environment (black line), higher income patients are more responsive to GP quality. This mirrors the observed income-quality gradient in the data as well as the reduced form evidence in Section 4. The first panel of Figure 7 reports the semi-elasticity by quartile of income. The top income quartile is quite elastic (a semi-elasticity of roughly 0.3), whereas the bottom income quartile is largely unresponsive to quality (a semi-elasticity below 0.1). The remaining panels show corresponding negative relationships by age and disability.

Figure 7
Decomposing Role of Heterogeneous Information and
Heterogeneous Preferences



a. Income quartile

b. Age quartile

c. Disability quartile

Notes: Figure shows how sensitivity to quality (as measured by the elasticity with respect to quality) changes with no information heterogeneity and no preference heterogeneity. Counterfactual simulations use the demand specification allowing for heterogeneity by income, age, and disability. The black line shows the elasticity with respect to latent quality for each quartile of income, age, and disability in the baseline scenario with no displayed summary star ratings. The equal-information counterfactual sets each individual's information precision to that of the top quartile in the corresponding panel. The equal-preference counterfactual sets each individual's preference for quality to that of the top quartile. The final counterfactual does both.

Our model implies that these differences in responsiveness cannot be explained by differences in preferences. The dashed red line presents the *equal preference* counterfactual, in which each individual's preference for quality is set equal to the average of the individuals in the top-quartile of the variable considered in the corresponding panel. Equalizing preferences actually slightly lowers the responsiveness of low-income patients, accentuating differences by income. This reflects the negative coefficient on income in the quality preference equation in Table 2. The pattern is similar for both age and disability.

Conversely, the model can explain effectively all of the observable differences in responsiveness via information heterogeneity. The dashed green line presents an *equal information* counterfactual, in which information precision is set equal to the average of the individuals in the top-quartile of the variable considered in the corresponding panel. Doing so flattens the relationship between income and responsiveness, with semi-elasticities close to 0.3 for all income quartiles. With equal information, high and low-income patients are similarly sensitive to quality, despite slightly different preferences. This finding is reinforced by an *equal-information and equal-preference* counterfactual, shown as a dashed blue line, as well as Appendix Figure E-1 which shows that the gains from equalizing information are concentrated among individuals who were least responsive to quality ex ante.

5.5 Counterfactuals: Information and Quality Improvements

We use our estimated demand model to evaluate policy counterfactuals in an environment where demand reflects heterogeneity in information alongside heterogeneity in preferences. This distinction matters for every intervention we consider. We focus first on a set of counterfactuals that involve information provision. Improving information directly impacts chosen quality and welfare, particularly for low-income and low-information individuals. As a consequence, informational interventions compress differences in observed quality choice and reduce inequality in outcomes across groups.

We then consider simple supply-side interventions that increase the latent quality of GPs. Since low demand need not reflect low valuations, these interventions have larger welfare gains than what would be implied by a model that does not take information frictions into account, especially for patients with less private information.

Consumer surplus. Because expected quality need not coincide with realized quality, welfare must account for the gap between posterior beliefs and the true quality patients ultimately experience. In environment E, individuals choose based on posterior expected quality b^E_{ijt} but ultimately experience true quality $\tilde{r}^E_j$. Following Train (2015), consumer surplus combines the expected maximum of anticipated utility with a correction for this expectation error:

$$CS^E_i = \mathbb{E}_{\varepsilon_i}\left[IV^E_i(\varepsilon_i) + \sum_{j \in J_{it}} p^E_{ij}(\varepsilon_i)\, \beta^q_i \big(\tilde{r}^E_j - b^E_{ijt}(\varepsilon_{ij})\big) \right],$$

where

$$IV^E_i(\varepsilon_i) \equiv \mathbb{E}_\nu \left[\max_{j \in J_{it}} \left\{ \beta^q_i\, b^E_{ijt}(\varepsilon_{ij}) + X'_{ijt}\beta + \xi_j + \nu_{ij} \right\} \right].$$

We examine distance-equivalent welfare, $CS^E_{i,\mathrm{km}} = -CS^E_i / \beta^d_i$, where $\beta^d_i \equiv Z'_i \beta^d$ is mover i's coefficient on distance. We report $\Delta CS = CS^{\mathrm{policy}} - CS^{\mathrm{baseline}}$ where aggregate welfare is $CS^E = \sum_i CS^E_{i,\mathrm{km}}$. Appendix D.5 provides additional details.

Information provision. We compare two information-based counterfactuals to our baseline environment with no visible star ratings. First, we consider a scenario in which patients observe star ratings.[25] The second is a full-information scenario that removes all private signal noise. We consider the effect of these interventions on the ratings of chosen GPs, consumer surplus and the elasticity with respect to quality. We present these, alongside our baseline, in the first three columns of Table 3.

Both scenarios improve sorting of patients to higher-rated GPs, with the largest gains for patients who are least informed in the baseline. Relative to the baseline (no-stars) environment, making star ratings available raises the average rating of chosen GPs from 3.27 to 3.30. This is roughly 60% of the improvement that arises in the full information counterfactual (3.32). The effects are concentrated among patients with weaker private signals. The low-information group, defined as those with below median private signal precision ($1/\sigma_i^2$), has roughly four to five times the gains (e.g., a shift from 3.21 to 3.26) of the high-information group (from 3.33 to 3.34). We see a similar pattern when comparing low versus high income groups (again defined as below versus above median).

The consequence of improved sorting is a reduction in the dispersion of realized quality. The gap in chosen ratings between low- and high-information patients falls by one-third with posted stars (from 0.12 to 0.08) and by one-half under full information (to 0.06). The analogous low- versus high-income gap falls by more than 25% with posted stars (from 0.11 to 0.08) and more than 45% with full information (to 0.06). Consistent with these changes, the correlation between the rating of the chosen GP and our inferred information index declines from 0.16 to 0.11 when stars are posted, while the correlation with income falls from 0.15 to 0.11. Under full information, these correlations fall further, to 0.08 and 0.09, respectively.

Welfare gains, measured in total distance-equivalent kilometers, mirror the sorting results. Across all individuals, providing public information in the form of rounded star ratings increases total CS by about 2,446 km, and providing full information increases total CS by about 7,880 km. Displaying the stars captures about 31% of the welfare gain obtained under full information. The gains to both interventions are disproportionately concentrated among patients with less baseline information, including low-income patients. High income and high information patients are only modestly impacted by the availability of star ratings, and experience muted gains even in the full information counterfactual.

[25] Because the model abstracts from any residual information patients may extract from raw comments in the post-removal period, this counterfactual should be interpreted as capturing the effects of the displayed stars, rather than the value of all information on the website.

Table 3
Summary of Counterfactual Simulations

	Full Model					Naive Model Assuming Full Info	
	Baseline (No Stars)	Stars on Website	Full Info	Quality Improvement (No Stars)	Quality Improvement (Full Info)	Baseline	Quality Improvement
Avg Rating	3.27	3.30	3.32	3.57	3.64	3.27	3.57
Avg Rating (Low Info)	3.21	3.26	3.29	3.50	3.62		
Avg Rating (High Info)	3.33	3.34	3.35	3.64	3.66		
Avg Rating (Low Income)	3.22	3.26	3.29	3.51	3.62	3.22	3.52
Avg Rating (High Income)	3.33	3.34	3.35	3.63	3.66	3.33	3.62
Elasticity wrt quality (ϵ_q)	0.188	0.093	0.352	0.209	0.407	0.707	0.771
Elasticity wrt quality (Low Info)	0.086	0.027	0.345	0.095	0.397		
Elasticity wrt quality (High Info)	0.314	0.216	0.359	0.357	0.419		
Elasticity wrt quality (Low Income)	0.120	0.043	0.336	0.136	0.384	0.623	0.683
Elasticity wrt quality (High Income)	0.272	0.184	0.372	0.303	0.431	0.838	0.902
Δ CS		2,446	7,880	69,991	80,201		51,876
Δ CS (Low Info)		2,327	6,663	35,710	44,617		
Δ CS (High Info)		119	1,217	34,281	35,584		
Δ CS (Low Income)		1,954	5,948	36,404	44,231		22,803
Δ CS (High Income)		493	1,934	33,590	35,976		29,068
Rating-Info Correlation	0.16	0.11	0.08	0.16	0.05		
Rating-Income Correlation	0.15	0.11	0.09	0.13	0.05	0.14	0.11
ϵ_q-Info Correlation	0.671	0.749	0.032	0.672	0.016		
ϵ_q-Income Correlation	0.431	0.542	0.071	0.426	0.046	0.293	0.282

Notes: Full model refers to the demand specification allowing for heterogeneity in information and preferences by income, age, and disability in Table 2. Naive model assuming full information refers to the demand specification in Appendix Table E-1, which allows heterogeneity in preferences by income, age, and disability. Quality improvement refers to a counterfactual in which the latent quality of a random 25% of GP practices is increased by 1 star.

The effect on the semi-elasticity with respect to quality is more nuanced. Because $H_{ijt}(\theta)$ is defined with respect to the latent continuous quality index, posting rounded stars can reduce the elasticity even as average chosen ratings rise. The reason is that rounded stars shift demand across half-star bins but make practices appear more similar (and hence reduce the elasticity) within bins. Under full information, by contrast, patients observe the latent continuous quality index, so the elasticity rises from 0.188 to 0.352 overall and from 0.086 to 0.345 for low-information patients.

Taken together, these results show that information provision raises average chosen quality, compresses disparities in who receives high-quality care, and delivers the largest benefits to individuals who are least informed ex ante. Furthermore, they suggest that even coarse informational interventions—like the star rating system—can have substantial impacts on patient choice.

Quality improvements. Policy makers have emphasized improving GP practice quality, and it is important to understand which patients benefit. We next examine counterfactuals that capture the impact of supply-side improvements in practice quality. In column 4 of Table 3, we consider a scenario where we increase the latent quality of a randomly chosen 25% of practices by one star given baseline information. In column 5 of Table 3, we consider the impact of this same increase in quality in a scenario where patients have full information. Notice that, if individuals do not re-optimize following this improvement, we expect a mechanical increase in the rating of chosen GPs of 0.25 stars.

The results indicate that quality improvements are valuable, but even more so when patients can identify which practices improved. The average rating of the chosen GP increases by 0.30 stars in the no-stars environment and by 0.32 stars under full information. For low-information and low-income patients, the increase in average chosen ratings rises from 0.29 stars to 0.33 stars, whereas for high-information and high-income patients it changes little, from about 0.30 to 0.31 stars.

The welfare gains are in line with the changes in chosen quality. In the no-stars environment, the quality improvement generates 69,991 km in consumer surplus. Jointly improving quality and providing full information generates 80,201 km of consumer surplus, an additional increase of about 15%. This suggests complementarity between information and quality. In particular, the joint gain from quality improvement and full information exceeds the sum of the standalone gain from full information and the gain from quality improvement under no stars. The additional gains are disproportionately concentrated among patients with weaker baseline information: the low-information group gains 44,617 km rather than 35,710 km (25% more), and the low-income group gains 44,231 km rather than 36,404 km (22% more). By contrast, the corresponding increases are only 4% for high-information patients and 7% for high-income patients.

We benchmark these results against a "naive" model estimated under the assumption that patients are fully informed. This model rationalizes heterogeneity in choice strictly through differences in preferences and access. Column 6 of Table 3 reports the baseline for the naive model, and column 7 reports the same quality-improvement counterfactual. This comparison allows us to quantify how a standard approach would assess the gains from quality improvements relative to our more flexible model. Details regarding this "naive" model and its estimation are in Appendix E.1.

Crucially, the overall welfare gains implied by our model (80,201 km) are roughly 55%

larger than the gains implied by the naive model (51,876 km). This is despite the fact that the average changes in provider quality are virtually identical across the two models. The reason is that the naive model attributes cross-group differences in sorting to preferences for quality rather than to differences in information.[26] As a consequence, the naive model significantly understates the gains for low-information patients. For example, the welfare gain for low-income patients after accounting for information heterogeneity (44,231 km) is roughly double what the naive model suggests.

A regulator using the naive model—and therefore ignoring information heterogeneity—would infer that quality improvements generate the largest gains in areas where revealed responsiveness to quality is strongest, such as high-income areas. Accounting for information heterogeneity changes this conclusion. The welfare gains from improving quality are slightly larger for low-income patients, even in the no-star benchmark in Column 4, because low baseline responsiveness partly reflects limited information rather than weak preferences for quality.

Equal access. Even our full information counterfactuals do not remove all cross-group differences in quality. The fundamental reason for this is differences in access to high quality providers. In Appendix Table E-3, we repeat our counterfactual experiments in an equal-access environment in which we randomly reassign patients to choice sets. In the equal access environment, baseline disparities in the ratings of chosen providers shrink, but a meaningful gap remains. For instance, under equal access, low- and high-information patients choose GPs with average ratings of 3.25 and 3.30, respectively.

Without differences in access, providing full information completely eliminates the quality gap between low- and high-income patients. In fact, with full information and equal access, low-income and low-information patients actually choose slightly higher-quality GPs, reflecting a slightly stronger preference for quality. The welfare gains are of similar magnitude to those in Table 3 and remain concentrated among patients with lower information at baseline, implying that the value of public information does not arise solely because disadvantaged patients happen to face worse choice sets. Instead, the results imply that access

[26]In particular, Appendix Table E-1 shows that the naive model implies lower estimated preferences for quality than the full model, especially for low-income groups.

and information are complementary: equalizing access narrows level differences in the quality patients receive, while better information improves how effectively patients sort within the choice sets they face.

5.6 Discussion and Robustness

Capacity constraints. Capacity constraints are a natural concern in our setting, since information-induced reallocation toward higher-rated practices could increase congestion and waiting times. In the baseline counterfactuals, we treat our congestion proxy (practitioners per 1,000 patients) as fixed at its observed level. As a robustness check, we allow capacity to adjust endogenously in the counterfactual by explicitly accounting for the feedback from simulated enrollment to congestion. Starting from observed staffing, we update each practice's practitioners-per-patient measure using the model-implied patient loads, recompute choice probabilities under the resulting capacity levels, and iterate until the implied enrollment distribution and capacity levels converge to a fixed point. This exercise captures the idea that patients may internalize anticipated crowding when deciding where to register, so demand responses to information could be attenuated by endogenous congestion.

The equilibrium outcomes under this endogenous-capacity adjustment are similar to our main results (Appendix Figure E-2 and Appendix Table E-4), indicating that the central findings on information heterogeneity and the distributional effects of public signals about quality are not an artifact of holding capacity fixed. Congestion feedback has a modest effect on sorting levels but does not alter the qualitative patterns or the main quantitative conclusions.

Supply-side changes. The counterfactual exercises above should be interpreted as addressing short-run congestion feedback rather than long-run supply adjustment. In particular, we do not model practice entry, exit, or quality investment. Allowing these margins to adjust could potentially amplify the demand-side effects we estimate. Related work finds the supply-side margin is important in other health care settings involving reviews (Vatter 2025). The stability of the counterfactuals under endogenous congestion suggests that our main conclusions are not an artifact of holding short-run access fixed, while still leaving scope for richer long-run supply responses outside the model.

Unobserved Preference Heterogeneity. A remaining concern is that unobserved heterogeneity in preferences, rather than information, may explain the variation in quality responsiveness. Our estimates are robust to allowing preferences and information to vary flexibly with a rich set of neighborhood characteristics, summarized by principal components. The resulting preference heterogeneity is similar to that in our baseline specification, while information heterogeneity remains the main source of variation in responsiveness. Thus, an omitted-preferences explanation would have to rely on preference components that are largely orthogonal to the observable demographics we use, yet sufficiently correlated with the RD responses to rationalize the differential threshold effects.

A related limitation is that most demographic characteristics are observed at the neighborhood (LSOA) level rather than the individual level. Although the registration data contain individual-level age and gender, they cannot be linked to income and other demographic measures. We therefore interpret the heterogeneity results as differences across local populations, rather than as a decomposition of within-neighborhood heterogeneity across individuals.

6 Conclusion

We study whether differences in information or preferences explain why patients enroll with lower-rated GP practices. Exploiting the NHS website's rounded star ratings, we show that demand jumps discretely when practices cross a rating threshold, and that these jumps disappear once summary stars are removed. The response is heterogeneous: low-income groups react strongly at thresholds, whereas higher-income groups display little discrete response but continue to sort toward practices with higher underlying review scores within a star-bin. Evidence on mechanisms points to differences in information, potentially driven by social networks. More broadly, the results imply that part of the persistent gradient in measures of health care quality across demographic groups reflects differences in what patients know, rather than differences in preferences or available options. That these gradients arise in a system with free registration and near-universal coverage underscores the fact that information frictions can generate disparities even when there are no financial barriers.

We embed these reduced-form facts in a structural demand model that allows information precision, preferences, distance costs, and congestion sensitivity to vary across patients. The model matches observed choice patterns and RD moments, and implies substantial heterogeneity in the precision of private signals about quality but comparatively limited heterogeneity in preferences for quality. Counterfactuals that equalize information sharply compress heterogeneity in responsiveness, whereas equalizing preferences does relatively little. They also show that access and information are complements: widening choice sets matters more when patients can identify which options are better.

These distinctions matter for evaluating disclosure and supply-side improvement policies. Relative to the current no-stars environment, posting coarse star ratings improves sorting toward higher-rated GPs, closes more than half of the gap to full information in average chosen quality, and reduces the relationship between chosen quality and income by about one-quarter. Full information yields additional gains by sharpening sorting within star-bins as well as across them. Better information also changes the incidence of supply-side improvements: the same increase in provider quality generates larger reallocation and surplus gains when patients are better informed, especially for low-income patients. A naive model that rules out information frictions misses this mechanism, attributing weak observed responsiveness to low tastes for quality and understating both the aggregate and distributional gains from quality improvement. Information heterogeneity is therefore first-order for welfare analysis, not merely for matching observed demand elasticities.

A final implication is that choice-based reforms and information policy have important interactions. Related evidence from UK hospital choice suggests that expanding provider choice can strengthen demand-side incentives and improve provider performance (Gaynor et al. 2016). Our results indicate that the distributional consequences depend on whether all patients can observe and interpret quality differences. Policies that expand choice or improve provider quality without addressing information gaps may leave substantial differences in realized care intact. Conversely, simple, salient, and credible public signals can meaningfully narrow those differences even when the signal is coarse.

Overall, in markets where quality is difficult to observe, gaps in realized quality arise not only from differences in access, but also from differences in information. Policies that pair expanded choice or quality improvement with better information are therefore likely to be more effective than policies that target access alone.

References

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

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

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

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.

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," *Review of Economics and Statistics*, 2025.

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.

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

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

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

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

Dor, Avi, William Encinosa, and Kathleen Carey, "Hospital performance standards and medical pricing: the impact of information disclosure in cardiac care," *Journal of Economics & Management Strategy*, 2020, *29* (3), 492–515.

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.

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

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

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

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

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

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

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.

Huitfeldt, Ingrid, Victoria Marone, and Daniel C Waldinger, "Designing dynamic reassignment mechanisms: Evidence from GP allocation," Technical Report, National Bureau of Economic Research 2025.

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

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

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

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

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

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

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

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

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

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

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

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

NHS, "Number of patients registered at a GP, 2017-2022. Retrieved from UK National Health Service (accessed January 2022), `https://digital.nhs.uk/data-and-information/publications/statistical/patients-registered-at-a-gp-practice`," 2022.

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

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.

Train, Kenneth, "Welfare calculations in discrete choice models when anticipated and experienced attributes differ: A guide with examples," *Journal of Choice Modelling*, 2015, *16* (C), 15–22.

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.

Vatter, Benjamin, "Quality Disclosure and Regulation: Scoring Design in Medicare Advantage," *Econometrica*, 2025, *93* (3), 959–1001.

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

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

Appendix for Online Publication

Appendix Contents

A. Data and Setting 49
A.1. Additional Data Sources 49
A.2. Descriptive Validation and Background 49
B. Reduced-Form Evidence 55
B.1. RD Heterogeneity, Validity, and Robustness 55
B.2. Panel Fixed-Effects Regressions 63
C. Mechanisms and Survey 65
C.1. Reporting and Review Content by Income 65
C.2. Survey Evidence on Information Sources 67
C.3. GP Choice Survey Methodology and Questions 69
D. Empirical Model and Estimation 71
D.1. Estimation Sample, Choice Sets, and Demographic Principal Components 71
D.2. Additional Details for the Empirical Model Setup 74
D.3. Estimation and Inference Details 78
D.4. Elasticity with Respect to Quality 80
D.5. Welfare and Counterfactual Implementation 82
E. Additional Structural Results 84
E.1. Naive Model Assuming Full Information 84
E.2. Model Fit and Decomposition 84
E.3. Additional Counterfactuals and Structural Robustness 88
Appendix References 91

A Data and Setting

A.1 Additional Data Sources

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

GP Patient Survey The GP Patient Survey is a large, independently administered survey conducted on behalf of NHS England.[27] The survey was conducted twice a year from July 2011 to March 2016, and after that point was conducted annually. We match this to quarterly data using the closest available survey date. We also use a restricted version of these data obtained from the NHS that includes respondent income tercile.

Quality and Outcomes Framework (QOF) The Quality and Outcomes Framework (QOF) is a system used for performance pay of GP practices. The QOF clinical score aggregates a number of clinical indicators, such as whether recommended vaccinations and diagnostic tests were provided for patients with specific diagnoses. We also examine the overall score, which includes indicators related to GP staff training. While these scores are available online, it is relatively difficult for patients to compare scores across GP practices.

GP Characteristics GP practice characteristics were obtained from NHS Digital Organisation Data Service. The location of each practice was obtained by geocoding the addresses in the "GP Practices" file. The opening date was also obtained from this file. The number of practitioners in a practice and each practitioner's experience were obtained from the "GP Practitioners" file.[28]

A.2 Descriptive Validation and Background

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

[28] These data are available from `https://digital.nhs.uk/services/organisation-data-service/data-search-and-export/csv-downloads/gp-and-gp-practice-related-data`.

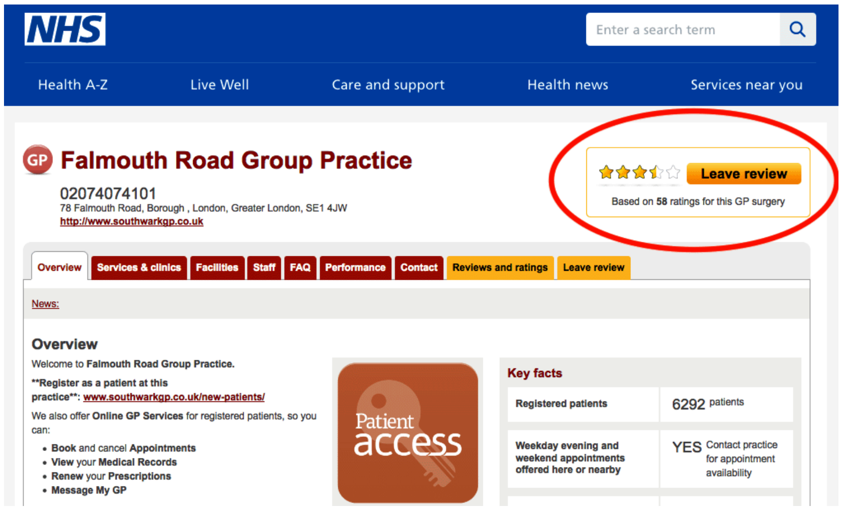
Table A-1
Summary Statistics for Reviews, Enrollment, and Demographics

	Period with Visible Star Ratings		Period without Visible Star Ratings	
	Mean	SD	Mean	SD
GP Reviews:				
Individual review	3.17	1.84	3.39	1.70
GP stars (2-yr moving avg)	3.20	1.02	.	.
GP num reviews (2-yr moving avg)	87.1	94.1	131.5	167.0
GP Enrollment:				
Total Enrollment (100s)	80.73	50.92	92.07	61.26
LSOA Enrollment (100s)	0.58	1.61	0.54	1.61
Quarterly LSOA Enrollment Change	0.17	2.08	0.09	1.80
Patient Demographics:				
Female	0.50	0.02	0.50	0.09
Age	39.93	4.55	40.32	4.55
LSOA Demographics:				
Income score	−0.13	0.10	−0.13	0.10
Fraction disabled	0.18	0.05	0.18	0.05
Unique GPs	7,640			
Total GP Observations	19,998,172			
Individual Reviews	368,644			

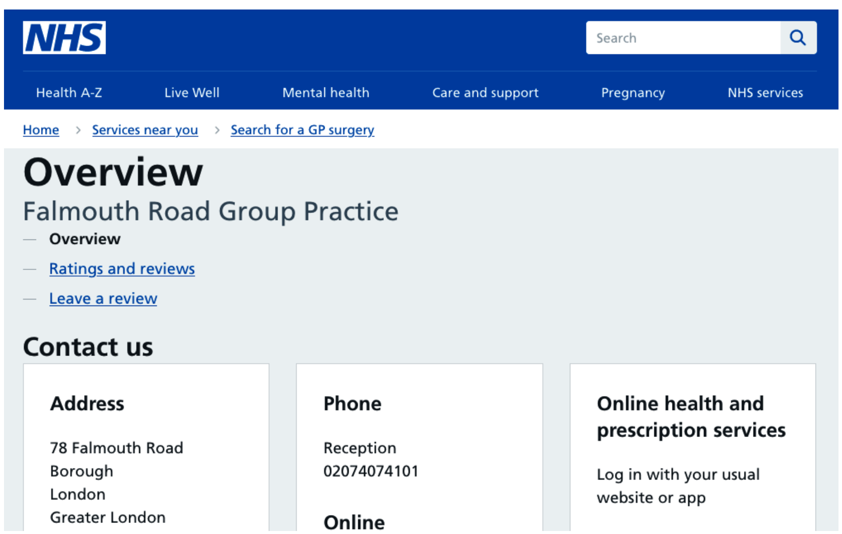
Notes: Summary statistics for reviews, enrollment, and demographics for GP practices in England. Columns 1–2 cover May 2015 to December 2019, during the period with visible star ratings. Columns 3–4 cover January 2020 to June 2022, during the period without visible star ratings. Total enrollment is measured at the GP-practice-quarter level. LSOA-level enrollment is measured at the GP-practice-LSOA-quarter level. An LSOA is a small geographic area containing roughly 700 households (about 2,000 individuals). The sample excludes GP-practice-LSOA-quarter observations in the top and bottom 2% of quarterly enrollment changes.

Figure A-1
Examples of the NHS Website

Panel A: With Visible Summary Star Ratings (Prior to 2020)

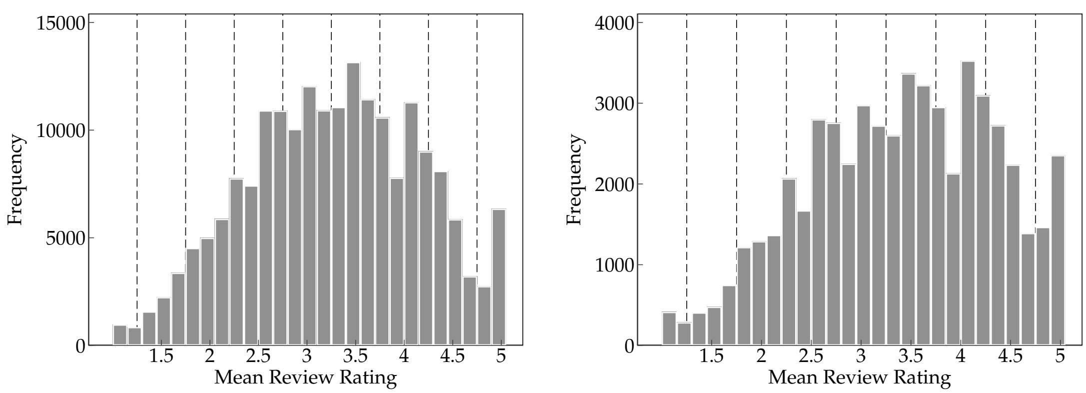


Panel B: Without Visible Summary Star Ratings (2020 and After)



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

Figure A-2
Histogram of Average GP Reviews



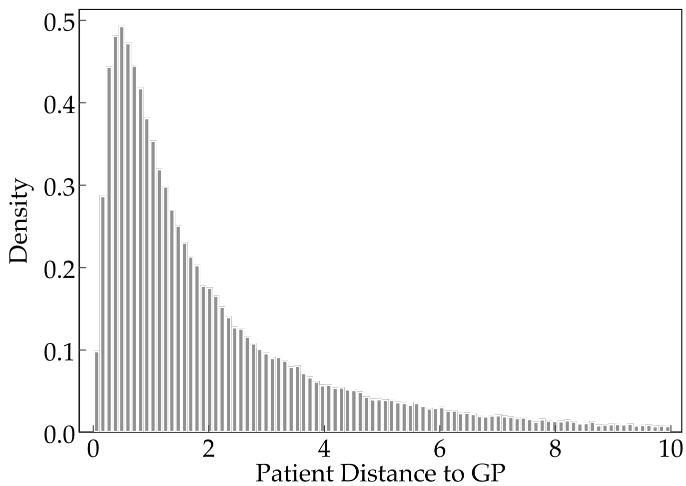
a. With visible star ratings

b. Without visible star ratings

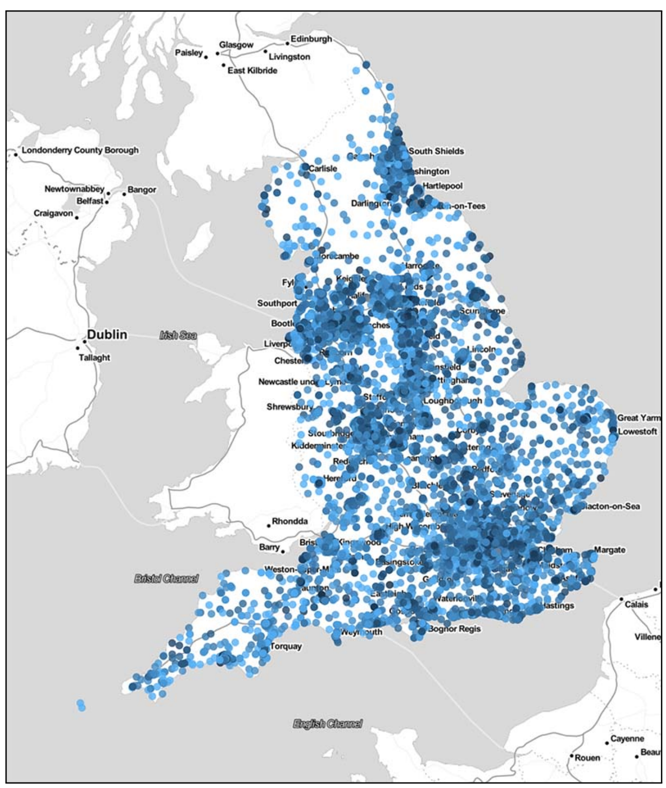
Notes: Figure shows the distribution of the review index r_{jt} during the period with visible star ratings and the later period without visible star ratings. The NHS calculated r_{jt} as the average of individual ratings over the previous two years. Vertical lines mark the half-star rounding thresholds.

Figure A-3
Travel Distance to GPs

Panel A: Histogram of Distance to Chosen GP



Panel B: GP Location and Enrollment



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

Table A-2
Correlation of Subjective Reviews with Other Quality Measures

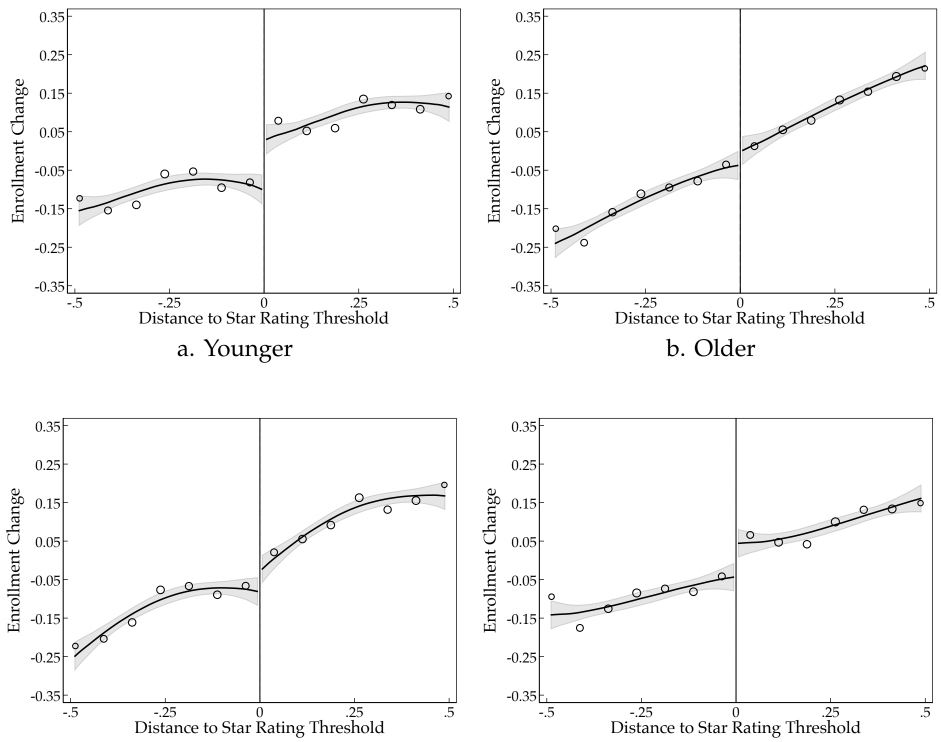
			> 100 Reviews	
	Corr	p-value	Corr	p-value
Patient Surveys:				
Easy getting through to GP	0.48	0.000	0.57	0.000
Receptionist was helpful	0.46	0.000	0.52	0.000
Able to get appointment	0.47	0.000	0.57	0.000
GP involved you	0.41	0.000	0.48	0.000
GP treated with care & concern	0.43	0.000	0.52	0.000
Confidence and trust in GP	0.38	0.000	0.46	0.000
Overall experience good	0.55	0.000	0.61	0.000
Quality & Outcomes Framework:				
Clinical (z-score)	0.20	0.000	0.63	0.000
Overall (z-score)	0.19	0.000	0.59	0.000

Notes: The table reports correlations between various measures of clinical and non-clinical quality and the review index r_{jt}, together with the corresponding p-values. Columns 1–2 report the main sample of GP practices with more than 5 reviews. Columns 3–4 report the subsample with more than 100 reviews.

B Reduced-Form Evidence

B.1 RD Heterogeneity, Validity, and Robustness

Figure B-1
Effect of Star Rating Threshold on GP Demand
Heterogeneity by Demographics



a. Younger

b. Older

c. Low Disability

d. High Disability

Notes: Mean enrollment change around the threshold for rounding star ratings. The sample used covers only the period with visible star ratings (May 2015 to December 2019). The size of the circles corresponds to the number of observations in each bin. The fitted line is from a local linear regression using a triangular kernel. Shaded area shows 95% confidence interval.

Table B-1
Effect of Star Ratings on GP Demand
Heterogeneity by Additional Characteristics

	Education	English	LFP	Foreign	Married	Migrant	Mover	Nonwhite	Environment	Housing
Above × Low	0.076*	0.067	0.063	0.048	0.123***	0.024	0.035	0.046	0.115**	0.095**
	(0.040)	(0.047)	(0.043)	(0.041)	(0.046)	(0.038)	(0.036)	(0.045)	(0.046)	(0.044)
Above × High	0.066	0.081*	0.084**	0.102**	0.023	0.124***	0.107**	0.107**	0.043	0.058
	(0.042)	(0.042)	(0.041)	(0.047)	(0.039)	(0.047)	(0.046)	(0.046)	(0.040)	(0.041)
High/Low Diff P-Value	0.836	0.808	0.690	0.357	0.055	0.062	0.144	0.315	0.189	0.499
Bandwidth	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25	0.25
N	1,573,713	1,573,713	1,573,713	1,573,713	1,573,713	1,573,713	1,573,713	1,573,713	1,621,745	1,621,745

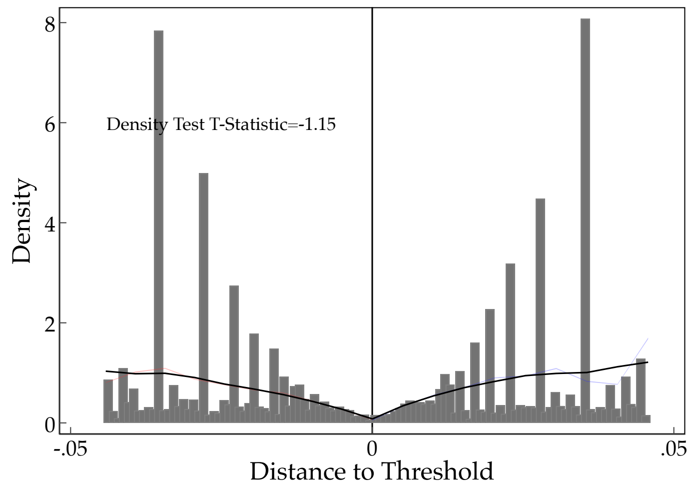
Notes: Dependent variable is the change in enrollment at the GP-LSOA-quarter level. LSOA-level demographic variables (Education, English, labor force participation (LFP), Foreign, Married, Migrant, Mover, Nonwhite) are constructed as population shares using the 2021 ONS Census. Environment and Housing represent the Living Environment and Barriers to Housing and Services deprivation domains, respectively, sourced from the 2019 English Indices of Deprivation. The sample used covers only the period with visible star ratings (May 2015 to December 2019). Low (high) is defined as LSOAs below (above) the median with respect to the relevant demographic characteristic. RD estimates use local linear regressions with triangular kernels and include controls for practice age, number of reviews, and cutoff fixed effects. We follow Calonico et al. (2019) and Calonico et al. (2014) to select bandwidths, include covariates, and calculate standard errors. Standard errors clustered at the GP practice level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B-2
Effect of Star Ratings on GP Demand
RD Heterogeneity: Greater London versus the Rest of England

	Greater London			England Excluding Greater London		
	Income	Age	Disability	Income	Age	Disability
Above × Low	0.181**	0.150**	0.118*	0.130***	0.134**	0.009
	(0.078)	(0.071)	(0.064)	(0.050)	(0.057)	(0.056)
Above × High	0.009	−0.059	−0.020	0.021	0.035	0.099**
	(0.073)	(0.086)	(0.113)	(0.050)	(0.044)	(0.044)
High/Low Diff P-Value	0.046	0.033	0.206	0.068	0.108	0.132
Bandwidth	0.25	0.25	0.25	0.25	0.25	0.25
N	516,867	500,236	500,236	1,104,800	1,073,410	1,073,410

Notes: Dependent variable is the change in enrollment at the GP-LSOA-quarter level. The sample used covers only the period with visible star ratings (May 2015 to December 2019). Low (high) is defined as LSOAs below (above) the median with respect to the relevant demographic characteristic. RD estimates use local linear regressions with triangular kernels and include controls for practice age, number of reviews, and cutoff fixed effects. We follow Calonico et al. (2019) and Calonico et al. (2014) to select bandwidths, include covariates, and calculate standard errors. Standard errors clustered at the GP practice level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Figure B-2
Density Tests of the Review Index Around Star-Rating Thresholds



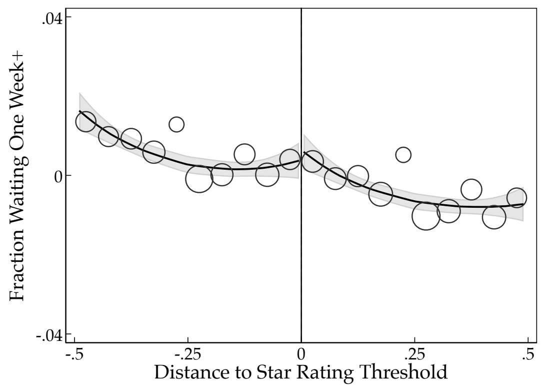
Notes: The plot shows a histogram and polynomial density estimate for practices above and below star rounding thresholds using our RD analysis sample. The unrestricted robust bias-corrected t-statistic following Cattaneo et al. (2018) is shown in the upper-left corner.

Table B-3
Effect of Star Ratings on GP Demand
RD Estimates Using the Period without Visible Star Ratings as the Control Group

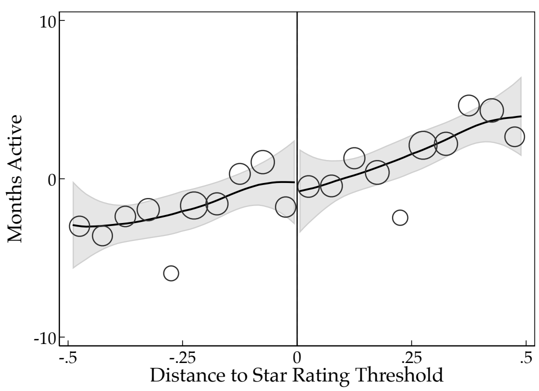
	Full	RD Effect Heterogeneity				
	Sample	Income	Education	English	Age	Health
Above Threshold × Visible Stars	0.078** (0.034)					
Above × Low × Stars		0.144*** (0.042)	0.119*** (0.042)	0.120*** (0.046)	0.133*** (0.045)	0.127*** (0.042)
Above × High × Stars		0.015 (0.041)	0.032 (0.045)	0.028 (0.041)	0.016 (0.039)	0.032 (0.042)
High/Low Diff P-Value		0.009	0.115	0.110	0.025	0.061
Bandwidth	0.25	0.25	0.25	0.25	0.25	0.25
N	2,022,032	1,621,745	1,573,713	1,573,713	1,573,713	1,621,745

Notes: Dependent variable is the change in enrollment at the GP-LSOA-quarter level. The treatment period is the period with visible star ratings (through December 2019). The control group is the later period without visible star ratings, restricted to April 2020 onward. Low (high) is defined as LSOAs below (above) the median with respect to the relevant demographic characteristic. RD estimates use local linear regressions with triangular kernels and include controls for practice age, number of reviews, and cutoff fixed effects. We follow Calonico et al. (2019) and Calonico et al. (2014) to select bandwidths, include covariates, and calculate standard errors. Standard errors clustered at the GP practice level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

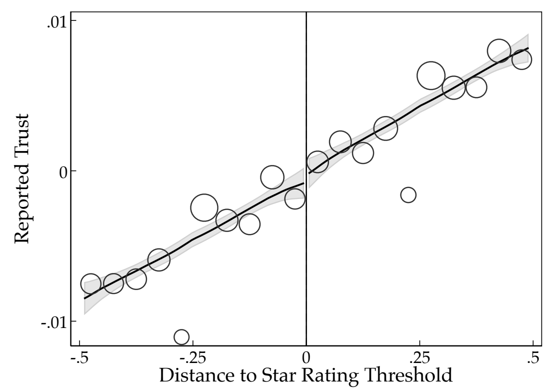
Figure B-3
Smoothness of GP-level Covariates Around Threshold



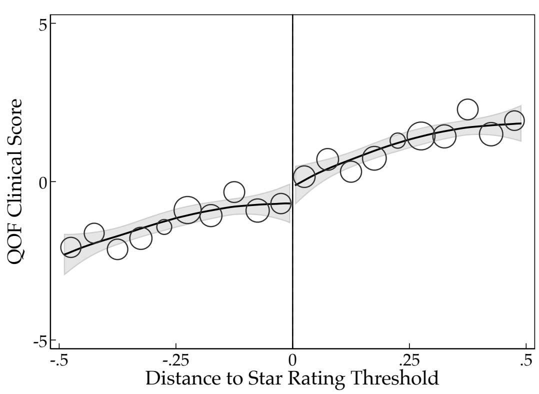
a. Waiting Times



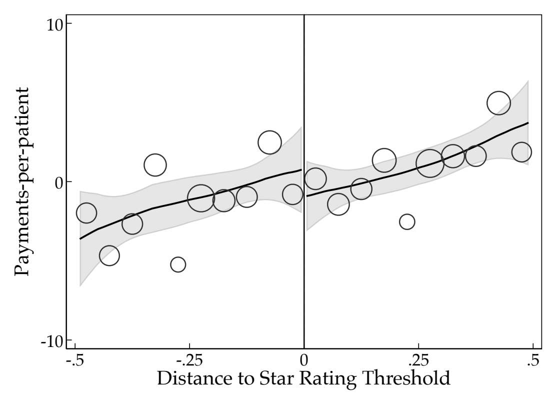
b. Months Active



c. Confidence and Trust (Survey)



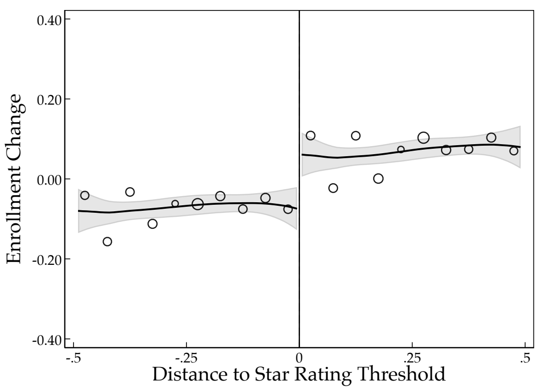
d. QOF Clinical Score



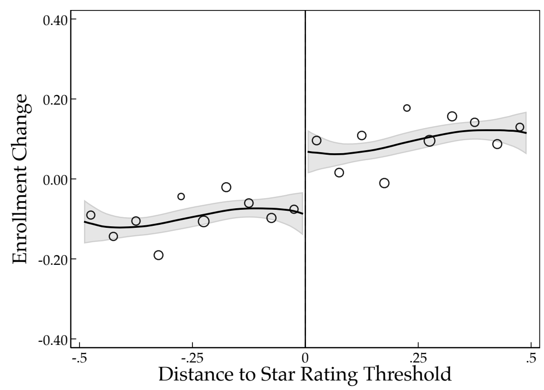
e. Payments Per-Patient

Notes: Changes at the rounding threshold for GP-level observables. The size of the circles corresponds to the number of observations in each bin. The fitted line is from a local linear regression using a triangular kernel. Shaded area shows 95% confidence interval.

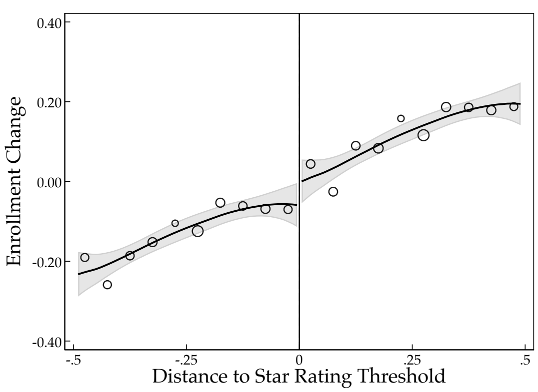
Figure B-4
Effect of Star Rating Threshold on GP Enrollment
by Income Quartile



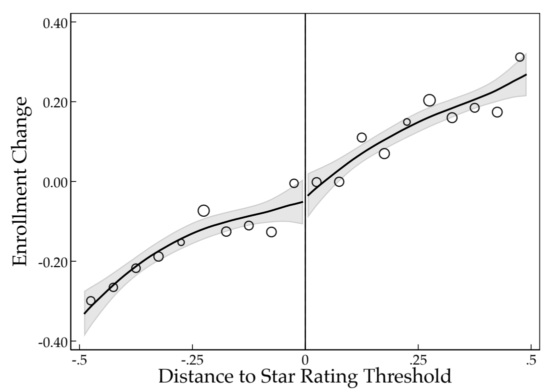
a. First Quartile
(Lowest Income)



b. Second Quartile



c. Third Quartile



d. Fourth Quartile
(Highest Income)

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

Table B-4
RD Estimates by Income and Age
Robustness Addressing Heterogeneous Website Usage by Age

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

Notes: The dependent variable is quarterly enrollment change for a GP–LSOA. Young and old LSOAs are defined by whether the fraction of individuals ages 20-44 is below or above the median. The sample used covers only the period with visible star ratings (May 2015 to December 2019). Low (high) is defined as LSOAs below (above) the median with respect to the relevant demographic characteristic. RD estimates use local linear regressions with triangular kernels and include controls for practice age, number of reviews, and cutoff fixed effects. We follow Calonico et al. (2019) and Calonico et al. (2014) to select bandwidths, include covariates, and calculate standard errors. Standard errors clustered at the GP practice level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B-5
RD Estimates by Income
Robustness Addressing Potential Capacity Constraints

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

Notes: The first two columns limit the sample to high-income LSOAs within 1 km of a low-income LSOA and to low-income LSOAs within 1 km of a high-income LSOA, implying that low- and high-income groups face a similar choice set. Columns 3 and 4 limit the sample to exclude GPs that appear to be at capacity, defined as total enrollment not changing between at least two consecutive months within a year. The dependent variable is quarterly enrollment change for a GP-LSOA. The sample used covers only the period with visible star ratings (May 2015 to December 2019). Low (high) is defined as LSOAs below (above) the median with respect to the relevant demographic characteristic. RD estimates use local linear regressions with triangular kernels and include controls for practice age, number of reviews, and cutoff fixed effects. We follow Calonico et al. (2019) and Calonico et al. (2014) to select bandwidths, include covariates, and calculate standard errors. Standard errors clustered at the GP practice level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table B-6
Effect of Star Ratings on GP Demand for Movers

	Low Income	High Income	Low Age	High Age	Low Disability	High Disability
RD estimate (jump)	0.127^{***}	-0.034	0.070	0.029	0.040	0.052
	(0.047)	(0.051)	(0.052)	(0.046)	(0.051)	(0.047)
Rating (slope)	-0.025	0.258^{***}	0.060	0.164^{***}	0.162^{***}	0.065
	(0.051)	(0.051)	(0.053)	(0.048)	(0.053)	(0.049)
Outcome Mean	1.86	1.94	2.03	1.78	1.99	1.81
Observations	345,962	345,174	341,113	350,023	340,860	350,276

Notes: Sample is individuals who moved to a new LSOA in the period of May 2015 to December 2019, during the period with visible star ratings. RD estimates use a bandwidth of 0.1. The slope coefficient is estimated in a separate regression using the full bandwidth. The unit of observation is a GP-LSOA-quarter. The dependent variable is the number of movers to the LSOA during that quarter registering with the GP practice. Low (high) is defined as LSOAs below (above) the median with respect to the relevant demographic characteristic. Standard errors clustered at the GP practice level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

B.2 Panel Fixed-Effects Regressions

As a complementary robustness exercise, we estimate a two-way fixed-effects specification that exploits within-practice changes in rounded star ratings over time:

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

where ℓ indexes LSOAs, s_{jt} is the rounded star rating, α_j are GP fixed effects, and δ_t are quarter-by-year fixed effects. The vector X_{jt} includes the continuous review index r_{jt}, practice age, practice age squared, and practitioners per practice. The outcome is quarterly enrollment change at the GP–LSOA level, and standard errors are clustered by GP practice.

Appendix Table B-7 shows that higher rounded star ratings are associated with higher enrollment growth, with larger effects for low-income LSOAs. The estimates are consistent with the RD results, although they are somewhat smaller in magnitude and rely on the stronger assumption that changes in rounded star ratings are uncorrelated with other changes in patients' information about practice quality.

Table B-7
Effect of Star Ratings on Enrollment Change
Panel Regression Estimates

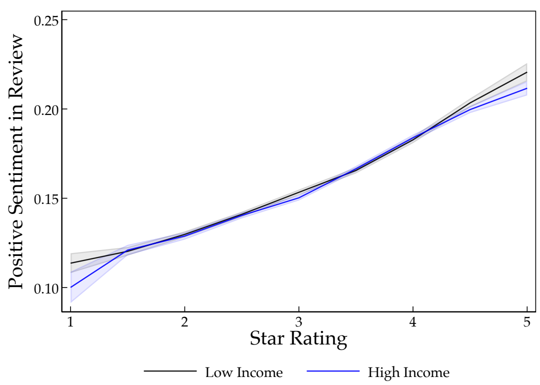
	Overall	Income	Education	English	Age	Health	All Interactions
Stars	0.010	0.003	0.007	-0.007	-0.005	0.004	-0.011
	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
Stars × Low Income		0.015^{***}					0.004^{***}
		(0.001)					(0.001)
Stars × Low Education			0.004^{***}				-0.004^{**}
			(0.002)				(0.002)
Stars × Low English				0.028^{***}			0.016^{***}
				(0.002)			(0.002)
Stars × Younger					0.025^{***}		0.017^{***}
					(0.001)		(0.001)
Stars × Low Health						0.013^{***}	0.004^{***}
						(0.001)	(0.001)
GP FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Quarter FEs	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Observations	8,472,817	8,472,817	8,154,497	8,154,497	8,154,497	8,472,817	8,154,497

Notes: The unit of observation is the quarterly enrollment change for a GP-LSOA. All specifications control for the continuous review index r_{jt}, practice age, number of reviews, and cutoff fixed effects. The sample used covers only the period with visible star ratings (May 2015 to December 2019). Standard errors clustered at the GP practice level are shown in parentheses. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

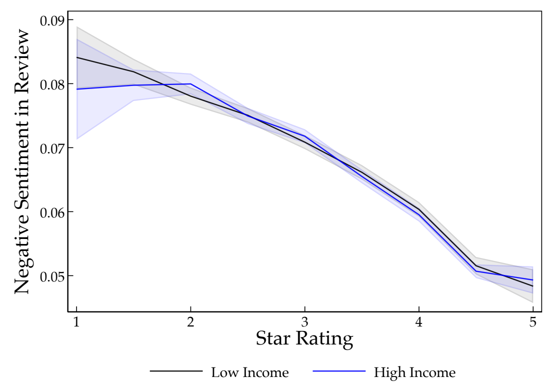
C Mechanisms and Survey

C.1 Reporting and Review Content by Income

Figure C-1
Review Text Sentiment by Income Group and Individual Review Rating



a. Frequency of Positive Sentiment



b. Frequency of Negative Sentiment

Notes: Figure plots the share of positive and negative words in review text against the individual star rating assigned in the review. Review text covers September 2018 to February 2022. Words are classified using the sentiment lexicon of Mohammad and Turney (2013). Low (high) is defined as LSOAs below (above) the median with respect to the relevant demographic characteristic.

Table C-1
Mean Patient Survey Responses by Income

	Income			
	Low	Middle	High	High − Low
Panel A. Raw mean				
Easy getting through to GP	70.18	73.32	74.57	4.38
Receptionist was helpful	90.22	91.63	92.57	2.36
Able to get appointment	69.15	69.51	69.90	0.76
Overall experience good	83.94	86.06	87.61	3.67
GP treated with care & concern	87.41	89.19	90.52	3.11
GP involved you	92.47	94.48	95.59	3.12
Confidence and trust in GP	95.09	96.28	97.05	1.96
Panel B. Within practice (GP fixed effects)				
Easy getting through to GP	-0.10	0.13	0.37	0.47
Receptionist was helpful	-0.22	0.07	0.31	0.52
Able to get appointment	0.60	-0.17	-0.13	-0.73
Overall experience good	-0.25	0.05	0.49	0.75
GP treated with care & concern	-0.27	0.08	0.53	0.80
GP involved you	-0.59	0.20	0.54	1.13
Confidence and trust in GP	-0.33	0.10	0.37	0.71

Notes: Entries are shares, reported in percentage points, of respondents answering "good" or "very good" to each GP Patient Survey question. Columns classify respondents by income tercile. Panel A reports raw means. Panel B residualizes responses using GP-practice fixed effects, so the remaining differences compare respondents in different income groups who attend the same practice.

C.2 Survey Evidence on Information Sources

Table C-2
Survey of GP Choice Information Sources

	Income				Health			
	Low Income	High Income	Diff	p-value	Poor Health	Good Health	Diff	p-value
Importance of clinical quality (1-10)	8.12	8.32	−0.20	0.224	7.98	8.53	−0.55	0.001
Importance of distance (1-10)	7.67	7.80	−0.13	0.493	7.66	7.85	−0.20	0.305
Importance of wait times (1-10)	8.17	8.21	−0.04	0.793	8.08	8.34	−0.27	0.090
Importance of friendliness (1-10)	7.81	7.91	−0.10	0.586	7.67	8.11	−0.43	0.017
Hours Research GP	3.26	4.35	−1.09	0.523	2.96	4.96	−2.00	0.242
Num GPs called/visited	2.08	2.62	−0.54	0.067	1.78	3.11	−1.33	0.000
Websites visited	3.10	7.04	−3.94	0.376	2.75	8.31	−5.55	0.214
Indivs consulted	2.30	3.53	−1.23	0.035	2.33	3.75	−1.42	0.016
Social media/forums consulted	1.10	1.47	−0.37	0.166	0.74	2.02	−1.28	0.000
GPs aware of	3.28	3.36	−0.08	0.735	3.21	3.47	−0.26	0.270
GPs considered	1.91	2.22	−0.31	0.053	1.73	2.51	−0.79	0.000
Website listed as source	0.71	0.71	−0.00	0.948	0.68	0.76	−0.08	0.053
Number sources listed	3.27	3.18	0.08	0.810	2.90	3.63	−0.74	0.033
Friends care about quality	7.32	7.32	−0.01	0.975	7.08	7.62	−0.54	0.005
Conversations with others	3.60	3.63	−0.04	0.941	3.37	3.93	−0.56	0.273
Friends/family are healthcare professionals	1.76	2.63	−0.86	0.026	1.70	2.88	−1.18	0.002
Neighbors registered at same GP	2.92	3.43	−0.51	0.194	3.05	3.37	−0.32	0.409
Observations	209	240			251	198		

Notes: Results from a convenience sample of 452 residents of England conducted through Amazon Mechanical Turk from May to November 2024. We exclude incomplete responses, respondents who completed the survey in less than 30 seconds, and respondents who reported being students. High-income is defined as self-reported monthly income after taxes above £5,000. Good health is defined as self-reported health (on a 1-10 scale) greater than the median.

C.3 GP Choice Survey Methodology and Questions

We administered a short, one-time online survey to adults aged 18 or older on Amazon Mechanical Turk (AMT) between May and November 2024. Eligibility required current registration with a GP practice in England. The instrument took under five minutes to complete, was anonymous, and concluded with a completion code for AMT credit. Participation was voluntary, with informed consent presented at the start of the questionnaire. We excluded incomplete responses, responses completed in less than 30 seconds, and respondents who reported being students, whose GP registration decisions may reflect transitory residence or institutional arrangements. The final sample includes 452 respondents.

Table C-3
GP Choice Survey Questions

Construct	**Survey question (response format)**
Demographic characteristics	
Postcode	What is the postcode of your current UK residence? (e.g., SW2 or BD11). [Text]
Household income	What is your household's monthly income after taxes, in pounds (£), approximately? [Number]
Education	What is the highest level of education you have completed? [Categorical: Primary; Secondary/high school; College/university; Postgraduate; Other]
Student status	Are you currently a student? [Yes/No]
Born in UK	Were you born in the UK? [Yes/No]
English as first language	Is English your first language? [Yes/No]
Self–rated health	On a scale of 1–10, how healthy are you compared to others your age and gender? (1 = very unhealthy, 10 = very healthy). [1–10]
Outcomes used in the analysis	
Importance of clinical quality (1–10)	On a scale of 1–10, how important is *clinical quality* in your decision of which GP to register with? (1 = you don't care, 10 = you care a lot). [1–10]
Importance of distance (1–10)	On a scale of 1–10, how important is *distance to your residence* in your decision of which GP to register with? [1–10]
Importance of wait times (1–10)	On a scale of 1–10, how important are *waiting times for appointments* in your decision of which GP to register with? [1–10]
Importance of friendliness (1–10)	On a scale of 1–10, how important is *friendliness and approachability* in your decision of which GP to register with? [1–10]

Continued on next page

GP Choice Survey Questions (continued)

Construct	Survey question (response format)
Hours researching GPs	How many hours did you spend researching GPs online when deciding which GP to register with? [Number]
GPs called/visited	How many GP practices did you call or visit for inquiries when deciding which GP to register with? [Number]
GP practice websites visited	How many official GP clinics' websites did you visit when deciding which GP to register with? [Number]
Individuals consulted	How many individuals (friends, family, neighbors, etc.) did you consult when deciding which GP to register with? [Number]
Social media / community forums consulted	How many social media groups or community forums did you consult for feedback on GP practices when deciding which GP to register with? [Number]
GPs aware of	How many GP practices within 3 km of your residence are you currently aware of? [Number]
GPs considered	How many GPs did you consider joining before choosing the GP you registered with? [Number]
Online sources indicator	*Derived from a multi–select item:* "Which of these sources did you use when deciding which GP practice to register with?" We code this as 1 if the respondent selected any online source (e.g., GP practice websites, Social Media, Google Maps); 0 otherwise. [Indicator]
Number of sources used	*Derived from the same multi–select item:* count of distinct sources selected. [Count]
Friends care about quality (1–10)	On a scale of 1–10, how much do your friends care about how good their GP is? (1 = don't care, 10 = care a lot). [1–10]
Conversations about healthcare	In the last month, how many conversations have you had with your friends or neighbours about GPs, hospitals, or healthcare? [Number]
Share of friends who are healthcare professionals	What percent of your friends, acquaintances, and neighbours are healthcare professionals (doctors, nurses, etc.)? [Number, percent]
Share of neighbours at same GP	What percent of your neighbours are registered at the same GP as you? [Number, percent]

D Empirical Model and Estimation

D.1 Estimation Sample, Choice Sets, and Demographic Principal Components

Estimation sample and choice sets. For structural estimation, we use movers age 25 and older who register with a new GP in Greater London between May 2015 and December 2019, as described in Section 2.

The set of inside options J_{it}^{I} consists of GP practices whose catchment areas include the destination LSOA. Since patients may register outside their catchment area during our sample period, the model also includes a composite outside option $j = 0$. If a catchment contains more than 30 practices, we retain the 30 closest practices as inside options and absorb the remainder into the outside option.

Appendix Table D-1 reports summary statistics for the resulting estimation sample and choice sets. The average mover faces roughly 10 inside options with substantial variation in provider quality, as measured by the review index.

Income and the disability share are assigned from the mover's destination LSOA, while age comes from the mover record and is measured in five-year bins. We also use a broader set of neighborhood demographic characteristics to construct four principal components. Appendix Table D-3 reports the corresponding loadings.

Table D-1
Summary of GP Choice Set

	Mean	SD
Demographics:		
Income	−0.16	0.10
Age (individual)	39.10	14.50
Disability	0.16	0.05
Principal component 1	1.41	3.72
Principal component 2	1.92	3.74
Principal component 3	0.53	2.15
Principal component 4	0.14	1.91
Choice Set:		
Distance to GP (km)	2.85	2.23
GP rating	3.25	0.95
GP num practioners	5.54	4.62
GP capacity	0.72	2.18
GPs in Choice Set	9.82	6.44
Chosen GP:		
Distance to GP (km)	1.47	1.68
GP rating	3.33	0.93
GP num practioners	7.77	5.62
GP capacity	0.73	1.30
Number of individuals	1,029,396	
Observations	1,266,759	

Notes: Summary statistics for the choice sets faced by movers in Greater London. The choice set is defined by GP catchment areas. Income, disability, and other neighborhood demographics are constructed at the destination-LSOA level, where an LSOA is a small geographic area with about 700 households (2,000 individuals). Age comes from the mover record. The sample used covers only the period with visible star ratings (May 2015 to December 2019).

Table D-2
Summary of Demographic Principal Components

	Eigenvalue	Difference	Proportion	Cumulative
Principal Component 1	6.013	2.615	0.401	0.401
Principal Component 2	3.398	1.690	0.227	0.627
Principal Component 3	1.708	0.747	0.114	0.741
Principal Component 4	0.962	0.107	0.064	0.805
Principal Component 5	0.854	0.157	0.057	0.862
Principal Component 6	0.697	0.284	0.046	0.909
Principal Component 7	0.413	0.097	0.028	0.936
Principal Component 8	0.316	0.140	0.021	0.957
Principal Component 9	0.176	0.020	0.012	0.969
Principal Component 10	0.156	0.050	0.010	0.980

Notes: We perform a principal component analysis on LSOA demographic variables measuring socioeconomic status (SES), income, health, age, English as a first language, foreign-born share, and education. The table reports each component's eigenvalue (variance), the difference from the next eigenvalue, the proportion of total variance explained, and the cumulative proportion. We retain the first four components.

Table D-3
Demographic Principal Components Loadings

	Principal Component 1	Principal Component 2	Principal Component 3	Principal Component 4
SES	−0.923	−0.312	−0.048	0.120
Income	−0.910	−0.324	−0.108	0.064
Health	−0.761	−0.455	−0.125	−0.079
Age	−0.699	0.417	−0.329	−0.296
English is first language	−0.684	0.628	0.022	−0.022
Born outside UK	0.567	−0.751	−0.089	−0.038
Education	−0.384	−0.660	−0.136	−0.192
Housing	0.303	−0.329	−0.399	−0.487
Environmental deprivation	0.444	−0.298	−0.165	−0.307
Labor force participation	0.066	−0.549	0.752	−0.026
Fulltime employed rate	−0.433	−0.365	0.726	−0.052
Unemployed rate	0.891	0.177	0.172	−0.063
Student rate	0.287	−0.339	−0.426	0.684
Disability rate	0.770	0.510	0.121	−0.062
Fraction minority	0.614	−0.619	−0.195	0.009

Notes: Principal components analysis on the correlation matrix of patient demographic characteristics. Entries are loadings $L_{jk} = v_{jk}\sqrt{\lambda_k}$, i.e., $\mathrm{corr}(x_j, \mathrm{PC}_k)$. Larger $|L_{jk}|$ indicates a stronger association with component k. Signs are arbitrary.

D.2 Additional Details for the Empirical Model Setup

This subsection gives the belief-updating formulas and implementation details used in Section 5. The empirical model differs from the stylized model in Section 3 because the public signal is a rounded half-star rating based on a finite number of reviews.

Posterior mean implied by the rounded public signal. Patients start from the prior $\tilde{r}_j \sim \mathcal{N}(m_0, s_0)$ and observe the private signal

$$\tilde{z}_{ij} = \tilde{r}_j + \sigma_i \varepsilon_{ij}, \qquad \varepsilon_{ij} \overset{i.i.d.}{\sim} \mathcal{N}(0,1).$$

Combining the prior and private signal gives

$$\tilde{r}_j \mid \tilde{z}_{ij} \sim \mathcal{N}(m_{ij}, v_{ij}), \qquad v_{ij} = \left(\frac{1}{s_0} + \frac{1}{\sigma_i^2}\right)^{-1}, \qquad m_{ij} = v_{ij}\left(\frac{m_0}{s_0} + \frac{\tilde{z}_{ij}}{\sigma_i^2}\right).$$

The continuous review index satisfies

$$r_{jt} \mid \tilde{r}_j, n_{jt} \sim \mathcal{N}(\tilde{r}_j, \varsigma_{jt}^2), \qquad \varsigma_{jt}^2 = \frac{\sigma_r^2}{n_{jt}}.$$

The website displays the rounded half-star rating s_{jt}, which reveals that $r_{jt} \in \mathcal{B}(s_{jt})$, where

$$\mathcal{B}(s) \equiv [\ell_s, u_s), \qquad \ell_s \equiv \max\{1, s - \tfrac{1}{4}\}, \qquad u_s \equiv \min\{5, s + \tfrac{1}{4}\}.$$

Conditional on the private signal, the predictive distribution of the review index is

$$r_{jt} \mid \tilde{z}_{ij}, n_{jt} \sim \mathcal{N}(m_{ij}, \Sigma_{ijt}^2), \qquad \Sigma_{ijt}^2 \equiv v_{ij} + \varsigma_{jt}^2.$$

Define

$$L_{ijt} \equiv \frac{\ell_{s_{jt}} - m_{ij}}{\Sigma_{ijt}}, \qquad U_{ijt} \equiv \frac{u_{s_{jt}} - m_{ij}}{\Sigma_{ijt}}, \qquad \lambda_{ijt} \equiv \frac{\phi(L_{ijt}) - \phi(U_{ijt})}{\Phi(U_{ijt}) - \Phi(L_{ijt})}.$$

The mean of the truncated predictive distribution is

$$\mathbb{E}\left[r_{jt} \mid \tilde{z}_{ij}, s_{jt}, n_{jt}\right] = m_{ij} + \Sigma_{ijt}\lambda_{ijt}.$$

Since $(\tilde{r}_j, r_{jt})$ is jointly normal conditional on $\tilde{z}_{ij}$,

$$\mathbb{E}\left[\tilde{r}_j \mid \tilde{z}_{ij}, r_{jt}, n_{jt}\right] = m_{ij} + \rho_{ijt}(r_{jt} - m_{ij}), \qquad \rho_{ijt} \equiv \frac{v_{ij}}{v_{ij} + \varsigma_{jt}^2}.$$

Iterated expectations therefore imply

$$b_{ijt} \equiv \mathbb{E}\left[\tilde{r}_j \mid \tilde{z}_{ij}, s_{jt}, n_{jt}\right] = m_{ij} + \rho_{ijt}\Sigma_{ijt}\lambda_{ijt}. \tag{D-1}$$

This is the empirical analogue of equation (1). The public signal is determined by the star rating bin $r_{jt} \in \mathcal{B}(s_{jt})$ and its precision varies with n_{jt}.

Local pass-through of the private signal. For the likelihood approximation, define the local pass-through of the private signal into posterior expected quality by

$$\omega_{ijt}(\tilde{z}) \equiv \frac{\partial b_{ijt}(\tilde{z})}{\partial \tilde{z}}.$$

Because

$$\frac{\partial m_{ij}}{\partial \tilde{z}_{ij}} = \frac{v_{ij}}{\sigma_i^2},$$

we differentiate the truncated-normal update with respect to its location parameter. Let

$$\Lambda(L, U) \equiv 1 + \frac{L\phi(L) - U\phi(U)}{\Phi(U) - \Phi(L)} - \left(\frac{\phi(L) - \phi(U)}{\Phi(U) - \Phi(L)}\right)^2.$$

This is the variance of a standard normal variable conditional on lying in $[L, U]$. Differentiating equation (D-1) gives

$$\omega_{ijt}(\tilde{z}_{ij}) = \left(\frac{v_{ij}}{\sigma_i^2}\right)\left[(1 - \rho_{ijt}) + \rho_{ijt}\Lambda(L_{ijt}, U_{ijt})\right]. \tag{D-2}$$

All objects in equation (D-2) are evaluated at the argument $\tilde{z}_{ij}$ through m_{ij}. In estimation we evaluate this derivative at the expansion point $\tilde{z}_{ij} = \tilde{r}_j$ and write $\omega_{ijt} \equiv \omega_{ijt}(\tilde{r}_j)$.

Local approximation used in the likelihood. The exact posterior mean is nonlinear in the private signal. For estimation, we use the first-order approximation

$$b_{ijt}(\tilde{z}_{ij}) = b_{ijt}(\tilde{r}_j) + \omega_{ijt}\big(\tilde{z}_{ij} - \tilde{r}_j\big) + R_{ijt}.$$

Using $\tilde{z}_{ij} = \tilde{r}_j + \sigma_i \varepsilon_{ij}$ and dropping the second-order remainder in the likelihood gives

$$b_{ijt}(\tilde{z}_{ij}) \approx b_{ijt}(\tilde{r}_j) + \sigma_i \omega_{ijt} \varepsilon_{ij}.$$

Thus the likelihood uses a linearized belief representation, while the welfare calculations use the exact posterior mean in equation (D-1).

Utility and observable heterogeneity. Under full information, mover i's utility from practice j in quarter t is

$$u_{ijt}^{\mathrm{FI}} = \beta_i^q \tilde{r}_j + X_{ijt}'\beta + \xi_j + \nu_{ij}, \qquad \nu_{ij} \sim \mathcal{N}(0, \sigma_\nu^2).$$

We normalize $\sigma_\nu^2 = \pi^2/6$, matching the normalization in a standard logit demand model. Expected utility under incomplete information is obtained by replacing $\tilde{r}_j$ with the posterior mean:

$$u_{ijt} = \beta_i^q \, b_{ijt}(\tilde{z}_{ij}) + X_{ijt}'\beta + \xi_j + \nu_{ij}.$$

Substituting the local approximation yields

$$u_{ijt} \approx \delta_{ijt} + \kappa_{ijt}\varepsilon_{ij} + \nu_{ij},$$

where

$$\delta_{ijt} = \beta_i^q \, b_{ijt}(\tilde{r}_j) + X_{ijt}'\beta + \xi_j, \qquad \kappa_{ijt} = \beta_i^q \sigma_i \omega_{ijt}.$$

The composite shock is therefore independent across options but heteroskedastic, with variance $\kappa_{ijt}^2 + \sigma_\nu^2$.

Observable heterogeneity enters through the marginal utility of quality,

$$\beta_i^q = \exp\big(Z_i'\theta_{\mathrm{qual}}\big),$$

the variance of private-signal noise,

$$\sigma_i^2 = \exp\left(Z_i'\theta_{\text{prec}}\right),$$

and preferences over non-quality attributes,

$$X_{ijt}'\beta = \text{dist}_{ijt}\, Z_i'\beta^d + \text{cap}_{ijt}\, Z_i'\beta^c + \text{outside}_{ijt}\, Z_i'\beta^0.$$

Here dist_{ijt} is distance from the mover's destination LSOA to practice j, cap_{ijt} is log practitioners per 1,000 patients, and $\text{outside}_{ijt} = 1\{j = 0\}$. Practice fixed effects apply only to inside options, with $\xi_0 = 0$. For the outside option we set $b_{i0t} = 0$ and $\omega_{i0t} = 0$.

Constructed inputs, latent quality, and expansion point. Let $r_{jm} \in \{1, \ldots, 5\}$ denote the mth individual star review for practice j, and let N_j be the total number of observed reviews for practice j. Define

$$\hat{r}_j^{LR} \equiv \frac{1}{N_j} \sum_{m=1}^{N_j} r_{jm}.$$

We construct

$$m_0 = \frac{1}{\sum_j N_j} \sum_j \sum_{m=1}^{N_j} r_{jm}, \qquad s_0 = \text{Var}_j(\hat{r}_j^{LR}), \qquad \sigma_r^2 = \text{Var}_{j,m}(r_{jm} - \hat{r}_j^{LR}).$$

The expansion point is

$$\tilde{r}_j = \hat{r}_j^{LR}.$$

The fixed effect ξ_j absorbs persistent non-quality demand shifters not captured by observed covariates or by this review-based quality proxy.

Alternative specifications for Z_i. The three specifications in the main text differ only in the contents of Z_i:

$$\text{(i)} \quad Z_i = (1, z_i^{\text{income}})', \qquad \text{(ii)} \quad Z_i = (1, z_i^{\text{income}}, z_i^{\text{age}}, z_i^{\text{disability}})',$$

$$\text{(iii)} \quad Z_i = (1, z_i^{\text{pc1}}, z_i^{\text{pc2}}, z_i^{\text{pc3}}, z_i^{\text{pc4}})'.$$

Each nonconstant element is standardized as a z-score. In specification (ii), income and disability are destination-LSOA variables, and age is measured in the mover record. Appendix D.1 and Appendix Tables D-3 and D-2 describe the underlying demographic variables and principal components.

D.3 Estimation and Inference Details

This subsection records the auxiliary moments used in the penalized likelihood estimator, the normalization of the penalty weights, and the bootstrap procedure. The notation matches the objective in equation (11).

Auxiliary RD targets. Let

$$\hat{\tau}^{data} = \left(\hat{\tau}_1^{data}, \ldots, \hat{\tau}_{K_{mom}}^{data}\right)'$$

collect the reduced-form RD estimates from Section 4: threshold jumps and within-bin slopes, including the subgroup splits used in the empirical analysis. For a candidate parameter vector θ, we compute model choice probabilities $P_{ij}(\theta)$, aggregate them to the same auxiliary cells as in the reduced-form sample, and re-estimate the same RD specifications to obtain

$$\hat{\tau}^{model}(\theta) = \left(\hat{\tau}_1^{model}(\theta), \ldots, \hat{\tau}_{K_{mom}}^{model}(\theta)\right)'.$$

In this mapping, we allow for a single multiplicative normalization so that the mean of the model-implied auxiliary outcome matches the corresponding mean in the data. This normalization makes the auxiliary block target the shape of the reduced-form response around thresholds rather than the unconditional level of the outcome.

Fixed-effect smoothness moments. To impose continuity of unobserved non-quality demand shifters at rating thresholds, we apply the same threshold-jump operators to the practice fixed effects and stack the resulting moments in

$$\Delta\xi(\theta) = \left(\Delta\xi_1(\theta), \ldots, \Delta\xi_{K_\xi}(\theta)\right)'.$$

Each element is the estimated discontinuity at a rating threshold when the dependent variable is the practice fixed effect rather than demand. Penalizing $\Delta\xi(\theta)$ toward zero imposes the same continuity logic as the reduced-form placebo tests for observables.

Weight normalization. Define the discrepancy for auxiliary RD target k as

$$m_k(\theta) \equiv \hat{\tau}_k^{model}(\theta) - \hat{\tau}_k^{data}.$$

The criterion in equation (11) can equivalently be written as

$$Q_N(\theta) = -\overline{\mathcal{L}}_N(\theta) + \lambda_0 \left[\sum_{k=1}^{K_{\text{mom}}} \widetilde{\lambda}_k m_k(\theta)^2 + \widetilde{\lambda}_{FE} \left\| \Delta\xi(\theta) \right\|^2 \right].$$

The relative weights within the penalty block are inverse-variance weights:

$$\widetilde{\lambda}_k = \frac{1}{2\widehat{\text{Var}}\left(\hat{\tau}_k^{data}\right)}, \qquad \widetilde{\lambda}_{FE} = \frac{1}{2\bar{\sigma}^2_{\text{jump}}},$$

where

$$\bar{\sigma}^2_{\text{jump}} \equiv \frac{1}{|\mathcal{K}_{\text{jump}}|} \sum_{k \in \mathcal{K}_{\text{jump}}} \widehat{\text{Var}}\left(\hat{\tau}_k^{data}\right),$$

and $\mathcal{K}_{\text{jump}}$ indexes the RD jump coefficients. Thus, more precisely estimated RD targets receive greater relative weight, while the fixed-effect smoothness block is scaled using the average sampling variance of the jump estimates.

Define the average relative weight across the K_{mom} scalar RD targets and the single fixed-effect smoothness block as

$$\widetilde{\lambda}_{\text{avg}} \equiv \frac{\sum_{k=1}^{K_{\text{mom}}} \widetilde{\lambda}_k + \widetilde{\lambda}_{FE}}{K_{\text{mom}} + 1}.$$

The common scaling factor is

$$\lambda_0 \equiv \frac{\kappa}{\widetilde{\lambda}_{\text{avg}}}, \qquad \kappa = 10. \tag{D-3}$$

With this normalization, the fitted model matches the reduced-form RD patterns closely while keeping the practice fixed effects smooth across thresholds. Moderate changes in λ_0 leave the structural estimates and counterfactuals very similar.

Bootstrap inference. Standard errors for the low-dimensional structural parameters are computed with an individual-level nonparametric bootstrap. In bootstrap repetition b, we resample the N movers with replacement and let $w_i^{(b)}$ denote the number of times mover i appears in the bootstrap sample, so that $\sum_{i=1}^{N} w_i^{(b)} = N$. The bootstrap average log-likelihood is

$$\overline{\mathcal{L}}_N^{(b)}(\theta) \equiv \frac{1}{N}\sum_{i=1}^{N} w_i^{(b)} \log P_{i\,y_i}(\theta).$$

The same bootstrap weights are also used on the model side of the penalty block. In each repetition we recompute the model-implied auxiliary outcome using the weighted mover sample, aggregate model choice probabilities to the auxiliary RD cells with weights $w_i^{(b)}$, apply the same mean normalization as in the full sample, and then evaluate the same RD operators to obtain $\hat{\tau}^{model,(b)}(\theta)$. We recompute the fixed-effect smoothness moments in an analogous way.

We re-optimize the objective over the structural parameters while holding the estimated practice fixed effects at their full-sample values. This substantially reduces computation in the presence of a large number of practice fixed effects, essentially treating the fixed effects as high-dimensional nuisance parameters. See Chernozhukov et al. (2018).

D.4 Elasticity with Respect to Quality

This appendix derives the semi-elasticity reported in equation (12) used to examine responsiveness to the ratings-based quality. Under the local utility approximation in Section 5, review counts and star ratings affect choice probabilities through δ_{ijt} and κ_{ijt}, or equivalently through δ_{ijt} and $\sigma^2_{u,ijt} \equiv \kappa^2_{ijt} + \sigma^2_v$.

Choice probability can be written as

$$P_{ij}(\theta) = \int_{-\infty}^{\infty} \frac{1}{\sigma_{u,ijt}} \phi\big(A_{ijt}(x)\big) \prod_{k \in J_{it} \setminus \{j\}} \Phi\big(A_{ikt}(x)\big)\, dx, \tag{D-4}$$

where

$$A_{ikt}(x) \equiv \frac{x - \delta_{ikt}}{\sigma_{u,ikt}}. \tag{D-5}$$

Define

$$G_{ijt}(x) \equiv \frac{1}{\sigma_{u,ijt}} \phi\big(A_{ijt}(x)\big) \prod_{k \in J_{it} \setminus \{j\}} \Phi\big(A_{ikt}(x)\big), \qquad \text{so that} \qquad P_{ij}(\theta) = \int G_{ijt}(x)\, dx. \quad \text{(D-6)}$$

Holding $\sigma_{u,ijt}$ fixed, differentiating under the integral sign yields

$$\begin{aligned} \frac{\partial P_{ij}(\theta)}{\partial \delta_{ijt}} &= \int \frac{\partial G_{ijt}(x)}{\partial \delta_{ijt}}\, dx \\ &= \int G_{ijt}(x) \left[\frac{A_{ijt}(x)}{\sigma_{u,ijt}} \right] dx. \end{aligned} \quad \text{(D-7)}$$

The second line uses

$$\frac{\partial}{\partial \delta_{ijt}} \left[\frac{1}{\sigma_{u,ijt}} \phi\big(A_{ijt}(x)\big) \right] = \frac{A_{ijt}(x)}{\sigma_{u,ijt}} \left[\frac{1}{\sigma_{u,ijt}} \phi\big(A_{ijt}(x)\big) \right].$$

Dividing by $P_{ij}(\theta)$ gives the semi-elasticity with respect to mean utility:

$$\frac{\partial \log P_{ij}(\theta)}{\partial \delta_{ijt}} = \frac{1}{\sigma_{u,ijt}} \frac{\int G_{ijt}(x)\, A_{ijt}(x)\, dx}{\int G_{ijt}(x)\, dx}. \quad \text{(D-8)}$$

Define the normalized weighting density over x,

$$d\mu_{ijt}(x) \equiv \frac{G_{ijt}(x)}{P_{ij}(\theta)}\, dx. \quad \text{(D-9)}$$

The semi-elasticity is therefore the expectation of $A_{ijt}(x)$ under this normalized weighting density.

$$\frac{\partial \log P_{ij}(\theta)}{\partial \delta_{ijt}} = \frac{1}{\sigma_{u,ijt}} \mathbb{E}_{x \sim \mu_{ijt}} \left[A_{ijt}(x) \right]. \quad \text{(D-10)}$$

Because a one-unit increase in posterior expected quality raises mean utility by β_i^q, the quality semi-elasticity is

$$H_{ijt}(\theta) = \beta_i^q \frac{\partial \log P_{ij}(\theta)}{\partial \delta_{ijt}} = \frac{\beta_i^q}{\sigma_{u,ijt}} \mathbb{E}_{x \sim \mu_{ijt}} \left[A_{ijt}(x) \right]. \quad \text{(D-11)}$$

Equation (D-11) gives the semi-elasticity with respect to posterior expected quality. To obtain the semi-elasticity with respect to latent quality $\tilde{r}_j$, holding fixed the public-information cell (s_{jt}, n_{jt}), define

$$\psi_{ijt} \equiv \frac{\partial b_{ijt}}{\partial \tilde{r}_j}.$$

Under the local approximation in Appendix D.2, $\psi_{ijt} = \omega_{ijt}$. Hence

$$H^q_{ijt}(\theta) \equiv \frac{\partial \log P_{ij}(\theta)}{\partial \tilde{r}_j} = \psi_{ijt}\, \beta^q_i \frac{\partial \log P_{ij}(\theta)}{\partial \delta_{ijt}} = \psi_{ijt} H_{ijt}(\theta).$$

For a subgroup g, the reported elasticity is

$$\epsilon^g_q(\theta) \equiv \frac{\sum_{i\in g} \sum_{j\in J^I_{it}} P_{ij}(\theta)\, H^q_{ijt}(\theta)}{\sum_{i\in g} \sum_{j\in J^I_{it}} P_{ij}(\theta)}. \tag{D-12}$$

When g is omitted, equation (D-12) is evaluated on the full sample.

The integral in equation (D-11) is one-dimensional and is evaluated with the same quadrature routine used for the likelihood.

D.5 Welfare and Counterfactual Implementation

Counterfactuals alter the public information available to movers and, in some exercises, the underlying quality distribution. In each environment E, we therefore recompute beliefs evaluated at the environment-specific public signal (s^E_{jt}, n^E_{jt}) and the mover's private-signal draw

$$\tilde{z}^E_{ij} = \tilde{r}^E_j + \sigma_i \varepsilon_{ij}.$$

Given a draw of ε_i, anticipated utility is

$$u^E_{ijt}(\varepsilon_i) = \beta^q_i\, b^E_{ijt}(\varepsilon_{ij}) + X'_{ijt}\beta + \xi_j + \nu_{ij}.$$

For notational convenience, write $b^E_{ijt}(\varepsilon_{ij}) \equiv b^E_{ijt}(\tilde{r}^E_j + \sigma_i \varepsilon_{ij})$. The realized-versus-anticipated quality gap in utility units is

$$d^E_{ijt}(\varepsilon_{ij}) \equiv \beta^q_i \big(\tilde{r}^E_j - b^E_{ijt}(\varepsilon_{ij})\big).$$

Hence consumer surplus in environment E can be written as

$$CS_i^E = \mathbb{E}_{\varepsilon_i}\left[IV_i^E(\varepsilon_i) + \sum_{j \in J_{it}} p_{ij}^E(\varepsilon_i)\, d_{ijt}^E(\varepsilon_{ij})\right],$$

where

$$IV_i^E(\varepsilon_i) \equiv \mathbb{E}_\nu\left[\max_{j \in J_{it}} u_{ijt}^E(\varepsilon_i)\right].$$

The inner expectation over ν is the inclusive value associated with the normal utility shocks and is evaluated with the same one-dimensional quadrature used for choice probabilities. The outer expectation over ε_i is approximated by Monte Carlo integration.

To express welfare in economically interpretable units, we convert surplus in utils into distance-equivalent kilometers using the mover-specific marginal disutility of distance,

$$CS_{i,\mathrm{km}}^E = \frac{CS_i^E}{-\beta_i^d}.$$

Aggregate welfare is obtained by summing $CS_{i,\mathrm{km}}^E$ over the relevant set of movers, and the reported policy effect is $\Delta CS = CS^{\text{policy}} - CS^{\text{baseline}}$. Note that $\beta_i^d \equiv Z_i' \beta^d$.

E Additional Structural Results

This appendix collects supplementary structural results referenced in Section 5.

E.1 Naive Model Assuming Full Information

As a naive benchmark, we estimate a conventional discrete-choice demand model that ignores information frictions. To align this exercise with the no-stars baseline used in our counterfactuals, we use the preferred specification to simulate choices in a synthetic no-stars environment and then estimate, on those simulated choices, a simpler demand model that assumes patients observe each practice's latent quality directly. All other features of the demand system follow the preferred specification.

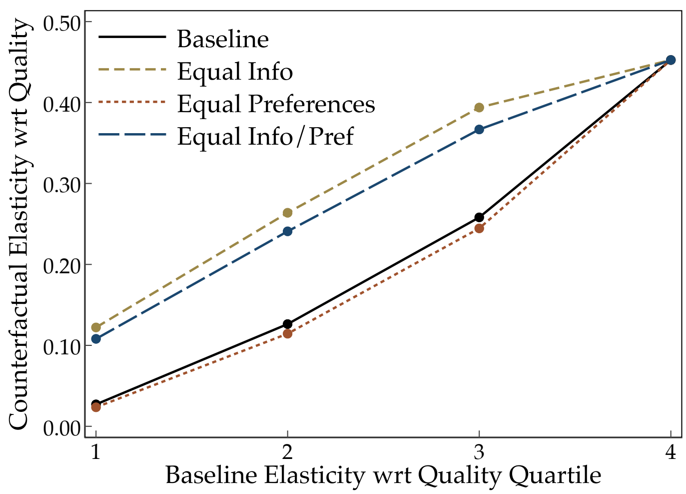
In this model, all heterogeneity is attributed to preferences rather than information.[29] This exercise therefore shows what a standard full-information demand model would infer in a setting without visible star ratings.

E.2 Model Fit and Decomposition

Table E-2 reports the fit of the preferred specification to the targeted RD moments, and Figure E-1 decomposes the role of information and preferences by baseline quality responsiveness.

[29]Because latent practice quality is otherwise time invariant, the coefficient on quality is not separately identified from practice fixed effects in this simple model without additional variation. When constructing the simulated estimation sample for this model, we temporarily add within-practice variation to latent quality using the time-varying component of the review measure. This preserves each practice's average latent quality while generating the within-practice variation needed to separate the quality coefficient from practice fixed effects. The adjustment is used only for estimation of this model and is removed before any baseline or counterfactual outcomes are reported.

Figure E-1
Decomposing Role of Heterogeneous Information and Heterogeneous Preferences
by Quartile of Baseline Elasticity with Respect to Latent Quality



Notes: The figure shows the elasticity with respect to latent quality under counterfactuals that equalize information, preferences, or both. Counterfactual simulations use the specification allowing for heterogeneity by income, age, and disability. The black line reports the elasticity for each quartile of baseline elasticity with respect to latent quality in the no-stars baseline. The equal-information counterfactual sets each individual's information parameters equal to those of the highest-elasticity quartile. The equal-preference counterfactual sets each individual's quality-preference parameters equal to those of the highest-elasticity quartile. The final counterfactual equalizes both.

Table E-1
Structural Model Estimates
Naive Model Assuming Full Information

	Heterogeneity by Income		Heterogeneity by Income/Age/Disability		Heterogeneity by Principal Components	
	Mean	SE	Mean	SE	Mean	SE
Preference for Quality:						
Constant	−0.457	(0.004)	−1.027	(0.007)	−0.652	(0.006)
× Income Score	0.155	(0.002)	0.195	(0.006)		
× Age			−0.377	(0.006)		
× Disability			−0.005	(0.008)		
× Principal Component 1					−0.153	(0.004)
× Principal Component 2					0.056	(0.004)
× Principal Component 3					0.020	(0.005)
× Principal Component 4					0.043	(0.005)
Preference for Distance:						
Constant	−2.107	(0.002)	−2.097	(0.002)	−2.102	(0.002)
× Income Score	−0.036	(0.002)	−0.059	(0.002)		
× Age			0.042	(0.002)		
× Disability			−0.033	(0.003)		
× Principal Component 1					−0.019	(0.002)
× Principal Component 2					−0.014	(0.003)
× Principal Component 3					0.026	(0.004)
× Principal Component 4					0.230	(0.002)
Congestion:						
Constant	0.103	(0.005)	0.162	(0.005)	−0.012	(0.004)
× Income Score	−0.058	(0.003)	−0.082	(0.007)		
× Age			0.075	(0.004)		
× Disability			−0.058	(0.008)		
× Principal Component 1					−0.046	(0.006)
× Principal Component 2					−0.102	(0.007)
× Principal Component 3					−0.051	(0.006)
× Principal Component 4					−0.028	(0.006)
Log-Likelihood	-718,073		-721,808		-720,484	

Notes: The table reports estimates from a model in which all individuals have full information. All other assumptions follow the model in Section 5. For comparability, the model is estimated on simulated choices from the no-stars baseline. Bootstrapped standard errors are shown in parentheses.

Table E-2
Model Fit for Targeted Moments

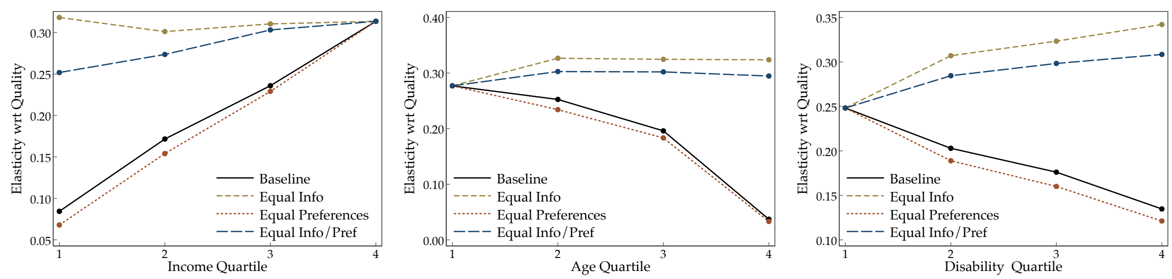
	Model	Target
RD Jump		
Low Income	0.137	0.127
High Income	−0.022	−0.034
Low Age	0.059	0.070
High Age	0.017	0.029
Low Disability	0.041	0.040
High Disability	0.053	0.052
RD Slope		
Low Income	−0.019	−0.025
High Income	0.252	0.258
Low Age	0.067	0.060
High Age	0.165	0.164
Low Disability	0.162	0.162
High Disability	0.074	0.065
FE Smoothness		
Low Income	0.002	0.000
High Income	0.001	0.000
Low Age	−0.002	0.000
High Age	−0.000	0.000
Low Disability	0.000	0.000
High Disability	−0.002	0.000

Notes: The table reports the model-implied RD jump, RD slope, and fixed-effect smoothness moments from the preferred specification (Specification 2 in Table 2) alongside their empirical counterparts.

E.3 Additional Counterfactuals and Structural Robustness

This subsection reports two additional exercises referenced in the main text: equalized access to local choice sets and counterfactuals with endogenous capacity adjustment.

Figure E-2
Decomposing Role of Heterogeneous Information and
Heterogeneous Preferences
with Endogenous Capacity Adjustment



a. Income quartile   b. Age quartile   c. Disability quartile

Notes: The figure repeats Figure 7 allowing capacity to adjust endogenously in each counterfactual. Counterfactual simulations use the specification allowing for heterogeneity by income, age, and disability. Capacity, measured by practitioners per 1,000 patients, is updated using simulated enrollment until the implied enrollment distribution and capacity levels converge to a fixed point. The panels report the elasticity with respect to latent quality for quartiles of income, age, and disability in the no-stars baseline and under counterfactual equalization of information, preferences, or both.

Table E-3
Counterfactual Simulations Equalizing Access

	Full Model					Naive Model Assuming Full Info	
	Baseline (No Stars)	Stars on Website	Full Info	Quality Improvement (No Stars)	Quality Improvement (Full Info)	Baseline	Quality Improvement
Avg Rating	3.27	3.30	3.32	3.57	3.64	3.27	3.58
Avg Rating (Low Info)	3.25	3.29	3.33	3.53	3.65		
Avg Rating (High Info)	3.30	3.30	3.31	3.61	3.63		
Avg Rating (Low Income)	3.26	3.30	3.33	3.55	3.65	3.27	3.56
Avg Rating (High Income)	3.29	3.30	3.31	3.59	3.63	3.28	3.59
Elasticity wrt quality (ϵ_q)	0.182	0.093	0.353	0.205	0.403	0.716	0.779
Elasticity wrt quality (Low Info)	0.092	0.029	0.375	0.101	0.429		0.940
Elasticity wrt quality (High Info)	0.290	0.199	0.331	0.329	0.379		
Elasticity wrt quality (Low Income)	0.135	0.051	0.384	0.151	0.439	0.626	0.683
Elasticity wrt quality (High Income)	0.237	0.157	0.324	0.264	0.371	0.840	0.912
Δ CS		2,386	7,611	70,119	80,087		52,137
Δ CS (Low Info)		2,285	6,427	36,917	45,609		
Δ CS (High Info)		101	1,184	33,202	34,478		
Δ CS (Low Income)		1,900	5,648	37,843	45,360		22,648
Δ CS (High Income)		487	1,965	32,279	34,731		29,486
Rating-Info Correlation	0.06	0.01	-0.02	0.08	-0.02		
Rating-Income Correlation	0.03	-0.00	-0.02	0.04	-0.03	0.02	0.02
ϵ_q-Info Correlation	0.589	0.697	-0.121	0.595	-0.126		
ϵ_q-Income Correlation	0.289	0.420	-0.168	0.293	-0.174	0.291	0.289

Notes: The table reports the same counterfactuals as Table 3 under an equal-access policy that randomly reassigns movers to choice sets. Full model refers to the demand specification allowing for heterogeneity in information and preferences by income, age, and disability in Table 2. Naive model assuming full information refers to the demand specification in Appendix Table E-1, which allows heterogeneity in preferences by income, age, and disability. Quality improvement refers to a counterfactual in which the latent quality of a random 25% of GP practices is increased by 1 star.

Table E-4
Counterfactual Simulations
with Endogenous Capacity Adjustment

	Full Model					Naive Model Assuming Full Info	
	Baseline (No Stars)	Stars on Website	Full Info	Quality Improvement (No Stars)	Quality Improvement (Full Info)	Baseline	Quality Improvement
Avg Rating	3.27	3.30	3.32	3.57	3.64	3.27	3.57
Avg Rating (Low Info)	3.21	3.26	3.29	3.50	3.62		
Avg Rating (High Info)	3.33	3.34	3.35	3.64	3.66		
Avg Rating (Low Income)	3.22	3.26	3.29	3.51	3.62	3.22	3.52
Avg Rating (High Income)	3.33	3.34	3.35	3.63	3.66	3.33	3.62
Elasticity wrt quality (ϵ_q)	0.188	0.093	0.352	0.209	0.407	0.707	0.771
Elasticity wrt quality (Low Info)	0.086	0.027	0.345	0.095	0.397		
Elasticity wrt quality (High Info)	0.314	0.216	0.359	0.357	0.419		
Elasticity wrt quality (Low Income)	0.120	0.043	0.336	0.136	0.384	0.623	0.683
Elasticity wrt quality (High Income)	0.272	0.184	0.372	0.303	0.431	0.838	0.902
Δ CS		2,448	7,859	69,992	80,182		51,874
Δ CS (Low Info)		2,328	6,648	35,712	44,605		
Δ CS (High Info)		119	1,211	34,280	35,577		
Δ CS (Low Income)		1,955	5,931	36,406	44,216		22,803
Δ CS (High Income)		493	1,931	33,588	35,972		29,066
Rating-Info Correlation	0.16	0.11	0.08	0.16	0.05		
Rating-Income Correlation	0.15	0.11	0.09	0.13	0.05	0.14	0.11
ϵ_q-Info Correlation	0.671	0.749	0.032	0.672	0.016		
ϵ_q-Income Correlation	0.431	0.542	0.071	0.426	0.046	0.293	0.282

Notes: The table reports the same counterfactuals as Table 3 but allows capacity, measured by practitioners per 1,000 patients, to adjust endogenously in each counterfactual by iterating to a fixed point. Full model refers to the demand specification allowing for heterogeneity in information and preferences by income, age, and disability in Table 2. Naive model assuming full information refers to the demand specification in Appendix Table E-1, which allows heterogeneity in preferences by income, age, and disability. Quality improvement refers to a counterfactual in which the latent quality of a random 25% of GP practices is increased by 1 star.

Appendix References

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.

Chernozhukov, Victor, Denis Chetverikov, Mert Demirer, Esther Duflo, Christian Hansen, Whitney Newey, and James Robins, "Double/debiased machine learning for treatment and structural parameters," 2018.

Mohammad, Saif M. and Peter D. Turney, "Crowdsourcing a Word-Emotion Association Lexicon," *Computational Intelligence*, 2013, *29* (3), 436–465.