Free to Choose?
Reform, Choice, and Consideration Sets in the English National Health Service

Martin Gaynor	Carol Propper	Stephan Seiler
Carnegie Mellon University	Imperial College	Stanford University
University of Bristol	University of Bristol	
NBER	CEPR	

ONLINE APPENDIX

A Timing of the Choice Reform

The reform did not happen in a discrete way on a certain date for cardiac care. There were two distinct trial/phase-in periods which we need to take into account when defining the pre- and post-reform dates. The 2006 choice reform was preceded by a choice pilot for cardiac patients who were experiencing particularly long waiting times (over 6 months). Between July 2002 and November 2003 such patients were allowed to change their provider to get treatment earlier. Since we do not observe which patients were actually offered the choice to switch providers, it is difficult to analyze this situation explicitly. At the same time, because patients eligible for this scheme had to have waited for a minimum of six months, this situation is quite different from the full choice reform (in which choice was mandated at the point of the referral) as well as from the situation of no-choice pre-reform. Second, choice was first introduced in a limited way in April 2005 and only fully rolled out in January 2006. In the introductory period, choice between only 2 hospitals was offered to patients and decisions were taken locally as to which choice to offer. In order to keep our analysis as clean as possible we therefore exclude both of these phase-ins of the reform from our analysis. We also allow for some time (one year) for the reform to settle in and therefore use only data from January 2007 onwards when analyzing the post-reform time period. These restrictions mean we use the period January 2004 to March 2005 as the pre-reform period and January 2007 to March 2008 as the post-reform period.

B Mortality Rate Adjustment

B.1 Regression Framework

In this section we describe a framework for risk adjusting hospital mortality rates, and assess whether risk adjustment has an impact on our estimation. To do this, we specify a linear probability model of patient mortality in which we regress an indicator for whether the patient died after the surgery on a set of hospital-time period fixed effects. Let the mortality of patient i in period t at hospital j be determined as follows,

$$M = JT\psi + (\gamma H + \eta) \tag{1}$$

where the last two terms in parentheses denote two components of the econometric error term. M is a vector of indicator variables. An entry corresponds to a particular patient i receiving a CABG in time period t and is equal to one if the patient died after receiving treatment in hospital j. JT is a matrix of hospital-time period dummy variables and ψ is a vector of coefficients. We define a time period to be a quarter. H represents the patient's health status. η is a vector of iid error terms.

We use the regression above to estimate the causal impact of visiting a particular hospital on the patient's probability of dying. However, simply estimating the relationship by OLS does not necessarily allow us to uncover the true causal relationship. Hospital choice will likely be correlated with patient health status, which will be subsumed in the empirical error term. If sicker patients choose systematically different hospitals, health status is predictive of which hospital dummy is "switched on." Therefore, any arbitrary column of the hospital dummy matrix, JT will correlated with the error term. This endogeneity problem is very closely related to the fact that the hospital's mortality rate is "contaminated" by differences in patient case-mix. In fact, when running the above regression by OLS, the fitted hospital fixed effects $\widehat{\psi}$ will be equal to the hospital/quarter-specific mortality rates.[1] The linear probability regression therefore recasts the issue of case-mix affecting the mortality rate as an endogeneity problem.

In order to uncover the causal effect on mortality from visiting a specific hospital, we need to instrument the hospital dummies JT, which we do using distance to the hospital (D_{ij}). Since we allow the hospital fixed effects to vary over time, we need to instrument $(J_t - 1)$ variables in each time period (a set of hospital dummies minus a constant). In order to do this, we need at least as many instruments. We choose to use the distance to each hospital and a set of dummies equal to one for the closest hospital. This yields a total of $(2 \cdot J_t)$ instruments for each time period (quarter). The identifying assumption that allows us to obtain a causal effect on patient survival is the exogeneity of patients' locations with respect to their health status. Under this assumption, we can use distance as an instrument for choice. It is a relevant instrument in the sense that distance is highly predictive of choice, as our demand model estimates show. It is a valid instrument under the assumption that patients' health status is not correlated with their locational choice and in particular with the distance of where they live to different hospitals that offer CABGs. Formally, in the actual IV regression this corresponds to the assumption that distance (to each hospital) is uncorrelated with the error term of the mortality regression (i.e. patient health status or the patient-specific probability of survival independent of hospital choice). The estimated hospital fixed effects in the IV regression can be interpreted as a case-mix adjusted mortality rate due to the fact that the instruments remove any impact of patients selecting according to severity.

One might also wonder whether the recovered LATE effect of the IV regression constitutes the relevant quality measure. This might be a concern if the LATE was very different from the average treatment effect. In our case, the LATE might differ from the average treatment effect if hospitals' ability to perform a CABG differs across patient types. For instance, it is conceivable that visiting a high quality hospitals makes more of a difference to the survival probability for a difficult case relative to a more standard one. The question then is whether distance has a differential effect on choice for different parts of the severity distribution, which is something we can test using the hospital choice data used to estimate the demand model. In order to explore whether the LATE is likely to differ from the average treatment effect we run a simple multinomial-logit regression using as covariates distance as well as an interaction of distance with the severity of the case. We run several different specifications using either linear distance or a closest hospital dummy or both as measures of travel distance. In all cases we find that severity does not alter the coefficient on distance in the utility function that predicts hospital choice. The results are consistently insignificant and of small magnitude. Based on these regressions we conclude that the recovered LATE effect is likely to be very similar to the average treatment effect.

[1] In a linear regression model without a constant, the fixed effects are equal to the hospital-specific means of the dependent variable $M_{it}(j)$. The average of $M_{it}(j)$ for a particular hospital j (and time-period t) is therefore simply equal to the number of deaths divided by the total number of admissions, i.e. the mortality rate.

B.2 Test for Exogeneity

Our primary focus of the IV regression is to establish whether case-mix differences did play an important role and hence generated an endogeneity problem in our linear regression model. The IV regression provides us with a straightforward test for this. Specifically, when running the regression as OLS without instrumenting, the estimated values of hospital/quarter fixed effects $\widehat{\psi}$ will be equal to the hospital/quarter-specific (unadjusted) mortality rates. Instead, the IV allows us to isolate the contribution of the hospital to the probability of patient survival independent of case-mix. Therefore, if case-mix did differ across hospitals, the estimated coefficient on the hospital dummies will be significantly different between the OLS and IV regression.

Before testing for differences between the OLS and IV estimates, we first establish the strength of our first stage. As a simple measure for the strength of the instruments, we compute the F-statistic for each of the individual first stage regressions (each hospital/quarter dummy is an endogenous regressors, hence there are 284 first stages for each of the 285 hospital/quarter dummies minus a constant). Doing so, we find an average F-statistic across all first stage regressions of 160.9 and a median F-statistic of 100.7 (the lowest F-statistic still has a relatively large value of 15.6). This is unsurprising, as our demand model clearly shows that distance is a strong predictor of hospital choice and this is reflected in a strong first stage in the case-mix adjustment regression.

Finally and most importantly, we run a Durbin-Wu-Hausman test to test for endogeneity of the hospital dummies. We fail to reject the null hypothesis of the test, that the hospital dummies are exogenous. We find an F-stat of 0.9735 and corresponding p-value of 0.6138. In other words, the OLS and IV estimates are not statistically different from each other and hence the case-mix adjusted mortality rate (obtained from the IV-regression) and the unadjusted rate (obtained from the OLS) are not significantly different from each other. Because distance is a strong predictor of hospital choice, it is not the case that we fail to reject the null due to a lack of statistical power.

As a final robustness check, we estimate our demand model with the adjusted mortality rate rather than the unadjusted one. The results from this regression are reported in Table (D2). The point estimates on mortality as well as other variables are very similar to our baseline specification using the unadjusted mortality rate. This is perhaps and simply another way to confirm that the two mortality rates are not very different from each other.[2]

C Kernel-smoothed Frequency Estimator

The simple frequency estimator described in the main text suffers from discontinuities in the likelihood. Specifically, in our context, a small movement in a parameter which influences the consideration set process might not alter the simulated consideration set for any consumer / simulated draw. In this case the choice probabilities and the corresponding likelihood are unchanged. In order to smooth the choice probabilities we employ a kernel-smoothed frequency estimator (D. McFadden (1989)). Note that the smoothing only happens at the "upper-level" of the consideration set process. Patient choices (conditional on a given consideration set) do not need to be simulated and therefore do not suffer from discontinuities.

In order to implement the smoothing, we compute for each draw the simulated consideration set $\widetilde{CS}_{s_i}$ (as defined in the main text) as well as a secondary consideration set $\widetilde{CS}^{+}_{s_i}$ that contains one additional option. Specifically, we add the highest utility option among the hospitals *not included* in the consideration set to form this secondary set. We then compute the choice probabilities conditional

[2] Note that the coefficient estimate on mortality is slightly smaller when using the adjusted mortality rate. However, the case-mix adjusted mortality rate has a higher standard deviation (1.46 times larger standard deviation than the raw mortality rate) and in terms standard deviations of the underlying variable, the point estimates are therefore more similar.

on both the main and the secondary simulated consideration set and take a weighted average between the two. The choice probability for any given set of draws s_i (which enter physician utility and therefore the consideration set process) is thus given by

$$\begin{aligned} \widetilde{Pr}^{CON}_{s_i k}(\Omega_{patient}, \Omega_{physician}) &= 1(k \in \widetilde{CS}^{+}_{s_i})[w_i(\Omega_{physician}) Pr_i(k|\widetilde{CS}_{s_i}, \Omega_{patient}) \\ &\quad +(1 - w_i(\Omega_{physician}))Pr_i(k|\widetilde{CS}^{+}_{s_i}, \Omega_{patient})], \end{aligned}$$

where $w_i(\Omega_{physician})$ is the weight associated to the primary consideration set. Note that if hospital k is contained in $\widetilde{CS}_{s_i}$, it is also included in $\widetilde{CS}^{+}_{s_i}$ (because by construction $\widetilde{CS}_{s_i} \subset \widetilde{CS}^{+}_{s_i}$).[3] The simple frequency simulator corresponds to the expression above with $w_i(\Omega_{physician}) = 1$. Summing over draws yields the simulated choice probability for individual i:

$$\begin{aligned} \widetilde{Pr}^{CON}_{ik}(\Omega_{patient}, \Omega_{physician}) &= \frac{1}{S_i} \sum_{s_i} \widetilde{Pr}^{CON}_{s_i k}(\Omega_{patient}, \Omega_{physician}) \\ &= \frac{1}{S_i} \sum_{s_i} 1(k \in \widetilde{CS^{+}}_{s_i})[w_i(\Omega_{physician}) Pr_i(k|\widetilde{CS}_{s_i}, \Omega_{patient}) \\ &\quad +(1 - w_i(\Omega_{physician}))Pr_i(k|\widetilde{CS^{+}}_{s_i}, \Omega_{patient})]. \end{aligned}$$

We define the weight as follows

$$w(\Omega_{physician}) = \frac{1}{1 + \kappa \exp(V_{ij^{+}t} - [max_{j \in J}(V_{ijt}) - \lambda_i])}$$

where $V_{ij^{+}t}$ denotes patient utility for hospital j^{+} which is the highest utility hospital not included in the primary simulated consideration set (but which is included in the secondary consideration set). $(V_{ij^{+}t} - [max_j(V_{ijt}) - \lambda_i])$ denotes the distance in utility-space of this hospital to the threshold. Note that $V_{ij^{+}t} < max_j(V_{ijt}) - \lambda_i$, i.e. the distance is measured as a negative number, because otherwise j^{+} would be included in the primary consideration set $\widetilde{CS}_{s_i}$. The further away $V_{ij^{+}t}$ is from the inclusion threshold $(max_j(V_{ijt}) - \lambda_i)$, the closer the weight is going to be to one. When the distance is smaller, the weight decreases and the influence of the secondary consideration set becomes larger. κ is a scaling parameter that governs the degree of smoothing. For small κ the estimator is closer to the simple frequency estimator. Due to a large number of draws (20 draws per individual across 32,714 patients) we are able to set $\kappa = 5$ which implies only a small amount of smoothing.

References

McFadden, D. 1989. "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration." *Econometrica*, 57: 995–1026.

[3] For the one option that is only contained in $\widetilde{CS}^{+}_{s_i}$, but not $\widetilde{CS}_{s_i}$, the term $Pr_i(k|\widetilde{CS}_{s_i}, \Omega_{patient})$ is equal to zero (because the primary consideration set $\widetilde{CS}_{s_i}$ does not contain this option).

D Additional Tables

Table D1: Descriptive Statistics — Mortality Rate and Waiting Times at the Quarter Level

		Waiting Times (Days)		Mortality Rate	
	Number of Hospitals	Mean	S.D.	Mean	S.D.
2004q1	28	113.7	36.3	1.14	0.94
2004q2	28	106.1	26.8	1.60	0.89
2004q3	28	102.5	26.6	1.39	1.10
2004q4	29	100.5	23.6	1.60	1.70
2005q1	29	93.4	21.6	1.12	0.90
2007q1	28	66.7	19.0	1.62	1.45
2007q2	28	66.2	18.5	0.67	0.93
2007q3	29	65.3	22.7	0.95	1.03
2007q4	29	63.9	23.9	1.19	2.33
2008q1	29	66.1	23.3	1.12	1.45

Table D2: Robustness Check: Demand Estimation with Unadjusted and Case-Mix Adjusted Mortality Rates[1]

		(1) Unadjusted Mortality Rate		(2) Case-Mix Adjusted Mortality Rate	
		Coefficient	Standard Error	Coefficient	Standard Error
Patient Preferences	Distance	-6.983	0.211	-7.026	0.218
	Closest Hospital Dummy	1.341	0.052	1.311	0.053
	Mortality Rate	-7.883	2.229	-3.828	1.490
	Mortality Rate * High Severity	-5.419	2.467	-3.661	1.660
	Mortality Rate * High Income	3.832	2.320	1.446	1.672
	Waiting Times	-1.528	1.887	-2.623	1.657
	Waiting Times * High Severity	-1.584	1.140	-1.580	1.062
	Waiting Times * High Income	6.262	1.196	5.972	1.297
Physician Preferences	Distance	-4.985	0.207	-5.046	0.561
	Closest Hospital Dummy	1.734	0.110	1.721	0.231
	Within-PCT Dummy	1.309	0.308	1.285	0.268
Choice Constraint Parameters	Constant	0.000	0.119	0.021	0.492
	High Severity	1.011	0.178	0.963	0.809
	High Income	0.000	0.113	0.018	0.653

[1] The first column corresponds to the baseline specification of the paper. The second columns presents results from the same model, using the case-mix adjusted mortality rate rather than the unadjusted one. All other aspects of the estimation are identical.

E Data Sources

Patient Choice Data	Hospital Episodes Statistics (HES) dataset. (http://www.hscic.gov.uk/hes) Administrative discharge dataset that covers all patients that underwent treatment in an NHS hospital.
Index of Multiple Deprivation	UK Census (https://www.gov.uk/government/collections/english-indices-of-deprivation). The measure is defined at the Middle Layer Super Output Area (MSOA). There are about 6,800 MSOAs in England with an average population of 7,200.