

THE WELFARE EFFECTS OF RESTRICTED HOSPITAL CHOICE IN THE US MEDICAL CARE MARKET

KATHERINE HO*

Economics Department, Columbia University, New York, USA

SUMMARY

Managed care health insurers in the USA restrict their enrollees' choice of hospitals to within specific networks. This paper considers the implications of these restrictions. A three-step econometric model is used to predict consumer preferences over health plans conditional on the hospitals they offer. The results indicate that consumers place a positive and significant weight on their expected utility from the hospital network when choosing plans. A welfare analysis, assuming fixed prices, implies that restricting consumers' choice of hospitals leads to a loss to society of approximately \$1 billion per year across the 43 US markets considered. This figure may be outweighed by the price reductions generated by the restriction. Copyright © 2006 John Wiley & Sons, Ltd.

Received 29 September 2004; Revised 13 September 2005

1. INTRODUCTION

Managed care health insurers in the US medical care market restrict their enrollees to visiting hospitals within specific networks. The network offered by each insurer affects consumer welfare, hospital profits and the incentives faced by hospitals to invest in new capacity, new technology and quality. However, there is very little literature on the allocation and impact of plans' hospital networks, constrained largely by a lack of data on plan contracts. This paper introduces a new dataset that lists the hospital networks of every managed care plan in 43 markets across the USA, making possible an analysis of the phenomenon. The dataset demonstrates that there is significant variation across both plans and markets in the extent to which plans exclude major hospitals from their networks. On average 17% of potential plan–hospital pairs in my data fail to arrange contracts. The proportion varies from 0% in some markets to as much as 40% in others. I define selective markets as those in which at least four of the five major plans fail to reach agreement with at least one major hospital: by this definition roughly 20% of observed markets are selective.

In this paper I investigate the effects of restricted hospital choice on consumer welfare. This requires deriving an estimate of consumer demand for health plans conditional on the network of hospitals they offer. The analysis is conducted in three steps. First I estimate a discrete choice model of demand for hospitals, taking into account consumer attributes such as location, diagnosis and income. Identification comes from the variation in individuals' hospital choice sets across markets. The second step is to use the estimated parameters from this demand system to find each type of consumer's expected utility from the hospital network offered by each plan in his or her market. Finally, this expected utility variable is included as one input to a discrete choice model

* Correspondence to: Katherine Ho, Economics Department, Columbia University, 1037 International Affairs Building, 420 West 118th Street, New York, NY 10027, USA. E-mail: kh2214@columbia.edu

for health plans. Here identification comes from the variation in plan choice sets across markets and from variation in the quality of the hospitals offered by each plan both within and across markets.

I use the demand estimates to predict the total welfare generated by the set of health plans in each market under two scenarios: first, that plans offer their observed hospital networks, and second, that all plans contract with every hospital in the market. I find, assuming fixed prices, that a change from the first to the second scenario would lead to a median equivalent variation (the dollar spending needed to compensate consumers for a move back to the observed networks) of \$15.70 per privately insured consumer per year. This translates to a total gain of \$1.04 billion for the 43 markets in the data. The variation across markets is high: there are seven markets in which the equivalent variation is over \$100 per person per year and 13 in which it is under \$1 per person per year. Producer surplus effects are small and negative. The change to unrestricted hospital choice would result in a median producer surplus reduction of just \$19 000 per market per year, translating to a \$0.80 million loss across the markets in the sample. The consumer surplus effects clearly dominate.

This paper models the demand side of the market only. I predict the change in consumers' choices when plans change their hospital networks and use these estimates to generate welfare predictions. I do not consider the supply side in any detail: the reasons why plans choose to restrict the set of hospitals they offer, and the benefits that these restrictions might bring, are not modeled in my analysis. Several potential benefits are clear from the existing health economics literature. First, plans may choose to limit choice in order to manage supply-side moral hazard problems, denying access to hospitals and doctors with questionable billing practices.¹ Second, an awareness of demand-side moral hazard issues, caused by the existence of generous health insurance and exacerbated by the subsidies implicit in the tax code, may lead consumers to accept restricted choice, higher deductibles and other coverage limits in exchange for lower premiums.² Finally, since patients generally stay within insurers' networks of hospitals, plans can use the threat of exclusion as a bargaining lever to contain hospital prices.³ For all these reasons, a move to unrestricted hospital networks would be likely to lead to significant increases in hospital costs and prices and in health plan premiums. My analysis holds premiums fixed. It therefore estimates just one effect of selective contracting: the loss to society from restricted hospital choice, ignoring price effects. This is a piece of the picture that has not been considered in the previous literature but it should not be interpreted as defining the full welfare effect. In fact the benefits of selective contracting may well outweigh the costs. This is discussed further in the final section of this paper.⁴

I am not the first to estimate demand for hospitals or for health plans. My analysis is closely related to the innovative papers of Town and Vistnes (2001) and especially Capps *et al.* (2003). Both of these papers use logit demand models to estimate consumer preferences over hospitals; these models are very similar to the first stage of my analysis.⁵ Capps *et al.* use the demand coefficients to calculate the value in utils that patients receive from having access to a network of

¹ See Newhouse (2002) among others for a discussion of these moral hazard problems.

² See, for example, Feldstein (1973) and Pauly (1986).

³ Cutler *et al.* (2000) and Capps *et al.* (2003) both discuss these issues.

⁴ One further caveat should be noted. The demand analysis yields no information on whether the full-choice outcome is in fact an equilibrium for profit-maximizing plans.

⁵ Tay (2003), Capps *et al.* (2001) and Gaynor and Vogt (2003) also estimate hospital choice models using encounter-level data and maximum likelihood techniques.

hospitals. The second stage of my analysis follows this calculation closely. However, my paper differs from their work in the method by which utils are translated into dollars for use in policy analysis. Capps *et al.* assume a simple bargaining model which implies that each hospital receives profits proportionate to the utility that it adds to the network; they estimate the dollar value of utils using a simple regression analysis. I make the more realistic assumption that, if a plan offers a higher-utility network to consumers, it will generate higher premiums and profits. I use the observed relationship between networks and premiums in the data, together with observed plan market shares, to infer the dollar value of utils. This requires estimating health plan demand as well as hospital demand but is closer to the first principles of consumer demand and therefore more realistic than the approach taken by Capps *et al.*

The plan choice literature is less fully developed than that on hospital demand. Several papers consider the effect of premium changes or quality data on employer and/or consumer choices of health plans;⁶ most of these use reduced-form analysis. I advance this literature by estimating a structural model of health plan demand that explicitly connects the three key players in this sector of the market: consumers, plans, and hospitals.⁷ The new dataset mentioned above, collected specifically for this paper, makes this analysis possible.⁸

The results reported here have implications for policy. They provide new evidence regarding the effect of hospital choice on welfare. Further research is clearly needed both to confirm the results of this paper and to pull together analysis of the demand and supply sides of the market. A complete welfare analysis would require a full model covering the bargaining process between insurers and hospitals and the effect of network changes on prices. I take a first step towards such a model in Ho (2005).⁹

This paper also offers a framework for analyzing other related demand systems. For example, the effect of physician network size, or the number of drugs included in a formulary, on demand for health plans could be modeled in an analogous way. Similar demand systems also exist in other industries, where downstream firms such as distributors or retailers contract with networks of upstream firms and may restrict consumer choice in order to reduce costs or avoid competition. The methods used in this paper may prove useful in these other situations.

This paper continues as follows. In the next section I describe relevant aspects of the industry and summarize the variation in networks across plans and markets. The demand estimation is outlined in Section 3; Section 4 describes the dataset; and Section 5 gives details on the estimation strategy. Demand results are given in Section 6. Section 7 analyzes the welfare effects of selective contracts and Section 8 concludes.

⁶ Buchmueller and Feldstein (1997), Beaulieu (2002) and Chernew *et al.* (2004a, 2004b) are examples.

⁷ Previous papers consider just two of the players and make assumptions or use reduced-form analysis to bring in the third. For example, the models in Town and Vistnes (2001) and Capps *et al.* (2003) both include consumers and hospitals explicitly but plans are added only in a reduced-form analysis. They find that hospital profits increase with value to the consumer, implying that plan willingness-to-pay increases with this measure. In order to investigate not prices but the existence of contracts, I model the impact of consumer demand for the hospital on plan demand explicitly.

⁸ Other important players include primary care physicians, who significantly impact the consumer's choice of hospital given his plan, and employers, who select the menu of plans from which the consumer chooses. Data limitations imply that their contribution cannot be modeled explicitly.

⁹ Ho (2005) employs the results of the demand model estimated in this paper as one input to a model of supply. First I use the demand estimates, together with data on hospital costs, to predict the producer surplus generated when each insurer chooses any potential hospital network in its market, holding other plans' networks fixed. Then I use these predictions, together with data on insurers' observed choices of hospital networks and on the characteristics of insurers and providers, to estimate a reduced-form function for the profits secured by different types of hospitals.

2. INDUSTRY BACKGROUND

2.1. Firm and Consumer Decisions

Each year, every privately insured consumer in the USA chooses a health plan, generally from a menu offered by his employer,¹⁰ and pays that plan a monthly premium in return for insurance coverage. The insurer contracts with hospitals and physicians to provide any care needed during the year. When the consumer requires medical care, he may visit any of the providers listed by the health plan, and receive services at zero charge or after making a small out-of-pocket payment. There is some variety in the restrictiveness of different types of managed care plan. If an individual is insured by a Health Maintenance Organization (HMO) he may visit only the hospitals in that plan's network. Point of Service (POS) plan enrollees can visit out-of-network hospitals but only if referred to them by a Primary Care Physician. Preferred Provider Organizations (PPOs) and indemnity plans are the least restrictive insurers: enrollees do not need a PCP referral to visit an out-of-network hospital, although PPOs may impose financial penalties for doing so, for example in the form of increased copayments or deductibles. The focus of this paper is on HMO and POS plans since their network choices have the strongest effect on both consumers and hospitals. Fifty-three percent of the privately insured population were enrolled in an HMO/POS plan in 2000.

Throughout the analysis I assume that hospital and plan decisions regarding the type and quality of products and services to offer, and also the network of hospitals offered by each plan, are exogenous to the consumer's choice of insurer and provider. Two additional assumptions simplify the estimation. First, I reduce the two-stage process by which employers choose a menu of plans and then consumers choose a plan from that menu into a single stage representing a 'joint' employer/consumer choice. There is a small literature that investigates how employers make their decisions, and how these decisions relate to consumer preferences,¹¹ but since my data do not identify employers I am forced to ignore this question. Second, I also assume that health plan choices are made at the individual, not the family, level. I have access only to aggregate plan data; while I could have included the distribution of family size in my estimation, this would have complicated the model substantially.

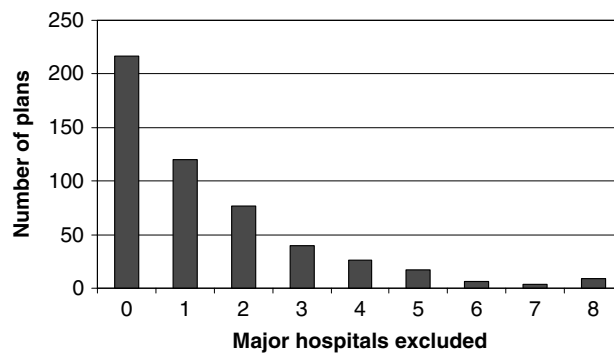
2.2. Variation in Size of Hospital Networks

The new dataset introduced in this paper, which lists the hospital network of every HMO and POS plan in 43 markets across the USA, demonstrates significant variation across both markets and plans in the extent to which plans exclude major hospitals from their networks. The data were collected from individual plan websites; missing data were filled in by phone. Figure 1 illustrates the variation. I categorize markets on a scale from 1 to 5, where 1 is the least selective, indicating that each of the five largest plans (by enrollment) contracts with all eight largest hospitals (by number of admissions). In markets ranked 5, at least four of the largest plans exclude at least one major hospital; the other categories lie between these extremes. Markets are fairly evenly spread across the five categories: 16 markets are ranked 1 or 2 (not selective) and 21 are ranked

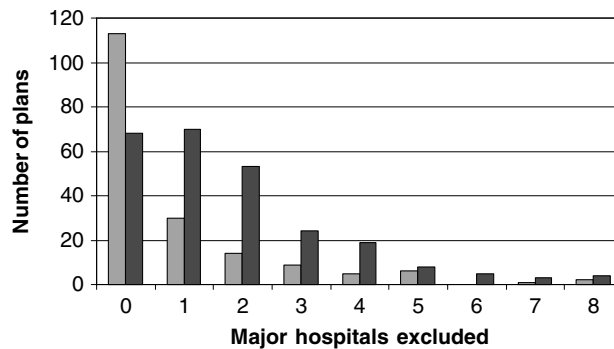
¹⁰ Fifty-eight percent of the population is insured through an employer, while only 5% purchase insurance independently (see the website www.statehealthfacts.kff.org).

¹¹ See, for example, Chernew *et al.* (2004a), which uses a dataset listing the health plans available to and offered by employers across markets to examine the effect of quality information on employers' choices of health plans.

Category	Definition	Number of markets	Examples
1	The five largest plans (by enrollment) contract with all eight largest hospitals (by number of admissions)	6	Baltimore, MD; Atlanta, GA
2	One plan excludes at least one hospital	10	Boston, MA; Columbus, OH
3	Two plans exclude at least one hospital or three plans exclude exactly one hospital each	6	Detroit, MI; San Francisco, CA
4	Three plans exclude at least one hospital; one of them excludes more than one	13	Houston, TX; Miami, FL
5	Four or more plans exclude at least one hospital each	8	Portland, OR; New Orleans, LA



Graph 1. Number of major hospitals excluded by each plan



Graph 2. Number of major hospitals excluded by each plan in selective markets (dark bars; categories 4-5 in the table above) compared to unselective markets (pale bars; categories 1-2 in the table)

Figure 1. Variation in plan networks across and within markets. This figure summarizes the variation in selectivity of plans' hospital networks both across and within markets. Markets are categorized on a scale from 1 to 5, where 1 is the least selective. Markets are fairly evenly distributed across the categories as shown in the table

4 or 5 (very selective).¹² The figure also shows the distribution of plans by the number of major hospitals excluded, and the variation in this distribution across types of market. Plans' selective behavior varies widely: 217 plans exclude no major hospitals, but 62 plans exclude at least four of the eight major hospitals in their markets. Plans on average exclude more hospitals in selective markets than elsewhere, but even within markets there is considerable variation in plan networks. This variation will be an advantage in terms of identification when I come to estimate the health plan choice equation. For example, as discussed below, it will be important to include market fixed effects to control for cross-market variation in the quality of the outside option. I will be able to do this, while still identifying the coefficient on expected utility from the hospital network, because of the variation in hospital choice sets across plans within each market.

The dataset contains no exclusive contracts (either hospitals reaching agreement exclusively with a single insurer or vice versa), and very few vertically integrated organizations. Many hospitals and health plans attempted vertical integration in the 1990s but this has become increasingly rare in recent years; the literature implies that the breadth of skills needed to run both a hospital and an insurer is too large for the vertically integrated model to be viable except in very specific circumstances.¹³ The key exception to this pattern is Kaiser Permanente, a dominant HMO in California and elsewhere that owns a large number of hospitals but contracts with few outside its own organization. I include Kaiser health plans and hospitals in the demand estimation and the welfare calculations: they are important members of the plan and hospital choice sets, particularly in California, and any change in their hospital networks could have significant effects on consumer surplus.¹⁴

3. SUMMARY OF THE ESTIMATION APPROACH

My main objective in estimating demand is to understand to what extent consumer utility from health plans is affected by the set of hospitals offered by each plan in the market. The analysis requires me to take three steps: these are summarized in Figure 2.¹⁵ First, I estimate demand for hospitals using a standard multinomial logit model and allowing for observed differences across individuals.¹⁶ With some probability consumer i becomes ill at time t in market m . His utility from visiting hospital h given diagnosis l is given by

$$u_{ihlt} = u(x_{ht}, \eta_h, v_{il} | \alpha, \beta) \quad (1)$$

where x_{ht} , η_h are vectors of observed and unobserved hospital characteristics respectively, v_{il} are observed characteristics of the consumer such as diagnosis and location, and (α, β) are the

¹² The variation in the extent of selective contracting across markets is not easily explained using market-level variables. Selective markets do not have significantly smaller populations, higher managed care penetration, more hospitals or more beds per capita than unselective markets and are not clustered geographically. There are no significant differences in demographic characteristics such as median income, average age of the population, race or education. These issues are discussed in further detail in Ho (2005).

¹³ See, for example, Burns and Pauly (2002) and Burns and Thorpe (2001).

¹⁴ I repeated the welfare analysis holding Kaiser networks fixed as a robustness test. The results are reported in Section 7.

¹⁵ In principle it would be preferable to estimate the two sets of parameters jointly. Unfortunately this is not possible owing to the scale of the problem: convergence would not be achieved within reasonable time limits.

¹⁶ This model was first proposed in McFadden (1973).

Step of estimation	Data inputs	Outputs
1. Hospital demand	MEDSTAT: encounter-level data on indemnity and PPO patients' characteristics and choice of hospital in 11 markets, 1997–8 AHA: hospital characteristics	Estimated effect of hospital characteristics on consumer utility given age, gender, diagnosis and location
2. Expected utility	New dataset listing network of each HMO/POS plan in 43 markets, Q1 2003 Estimated parameters from Step 1	Expected utility of every consumer type in the market from every plan's network of hospitals
3. Plan demand	AIS: market share and characteristics of HMO/POS plans in 43 markets, 2002 NCQA: clinical quality and consumer assessment data, 2000 Weiss: characteristics of HMO/POS plans, 2002 Expected utility from Step 2	Estimated parameters of plan demand equation

Figure 2. Overview of the three-stage estimation process. This table summarizes the three steps of the estimation process. The key input to the plan demand equation is the expected utility variable which measures the utility each type of consumer could expect to receive from each plan's hospital network given his age, gender and location. It is calculated using parameter estimates from a separate analysis (the hospital choice model) that uses a different dataset (MEDSTAT data). This linking of datasets is necessary because no single source simultaneously identifies the patient's insurer and corresponding hospital choice set, the chosen hospital and the patient's characteristics

coefficients on the specification.¹⁷ No outside option is needed in the hospital choice equation: the data include only patients who are sick enough to go to hospital for a particular diagnosis.¹⁸ Consumers choose hospitals to maximize their utility, so that if consumer i with diagnosis l chooses hospital h then for all other hospitals h' in the market

$$u_{ihl} = u(x_h, \eta_h, v_{il} | \alpha, \beta) \geq u_{ih'l} = u(x_{h'}, \eta_{h'}, v_{il} | \alpha, \beta) \quad (2)$$

This maximization produces the set A_h of v that choose hospital h . Thus shares are given by

$$s_h(x, \eta | \alpha, \beta) = \Pr(v \in A_h) \quad (3)$$

¹⁷ Individual i 's choice set is defined by his market m ; however, no market subscript is needed on individual-specific variables since individual characteristics i include location. Subscript t is used to define years. Observed hospital characteristics were permitted to vary by year; the estimation was simplified by assuming that unobserved quality and the estimated coefficients were constant across the two years of available data. The time subscript is omitted for the remainder of the paper for ease of exposition.

¹⁸ The consumer's expected utility from the set of hospitals offered by the plan incorporates the probability of admission to hospital for each diagnosis, conditional on age and sex, separately from the hospital demand estimates.

Details of the estimation process are set out in Section 5.1. The next step is to take the estimated parameters α , β and use them, together with predictions of the probabilities of diagnoses for each type of consumer, to predict expected utilities provided by each plan's hospital network. The exact methodology is described in Section 5.2. Finally the resulting variable, individual i 's expected utility from the set of hospitals offered by plan j in market m , which I denote EU_{ijm} , is used as an input to the health plan demand model. In this case I use a methodology similar to that first proposed by Berry *et al.* (1995) (BLP), and later used by Nevo (2000, 2001a), Petrin (2001) and others. The utility of individual i from enrolling in plan j in market m is given by

$$w_{ijm} = w(z_{jm}, \xi_{jm}, \text{prem}_{jm}, EU_{ijm}, y_i | \vartheta, \gamma) \quad (4)$$

where z_{jm} , ξ_{jm} are vectors of observed and unobserved plan characteristics, respectively, prem_{jm} is the vector of plan premiums (prices charged to consumers), y_i is the income of individual i , and (ϑ, γ) are the coefficients of this specification. The outside option is choosing to remain uninsured. Exactly the same reasoning applies as for the hospital demand model, so that if B_{jm} is the set of consumer types that choose plan j in market m , then plan j 's share in market m is

$$s_{jm}(z, \xi, \text{prem} | \vartheta, \gamma) = \Pr[(EU, y) \in B_{jm}] \quad (5)$$

Details on the empirical approach to estimating this model are given in Section 5.3.

Preferences over characteristics in the hospital choice model are identified using variation in the individual's hospital choice set across markets. The effect of each hospital's characteristics is uncovered by considering the impact of changes in the choice set on hospital market shares: for example, how does the share of hospital A, which has a high number of nurses per bed, change when hospital B, with fewer nurses but more doctors per bed, is added to the choice set? (The choice set does not vary within markets since, as discussed in the next section, I consider indemnity and PPO enrollees only.) Identification in the plan choice model comes from variation in the individual's plan choice set across markets, and also from differences in plans' hospital networks, both within and across markets. This allows me to estimate the impact of the key variable, the individual's expected utility from the plan's hospital network, on his plan preferences.

4. DATA

My analysis employs three sets of data. The first, which covers 11 US markets, includes all the information needed to estimate the consumer utility equation for hospitals. The second identifies the hospitals in the network of every managed care plan in a different sample of 43 US markets; this is used to predict the utility each individual can expect to gain from the hospital network offered by each plan in his or her market. The third dataset contains other plan characteristics and plan market shares for the same 43 markets; this is the final input needed to estimate the consumer utility equation for health plans.

4.1. Hospital Dataset

The core of this dataset is the MEDSTAT MarketScan Research Database for 1997–98. It is constructed from privately insured paid medical claims data provided by approximately 50

employer databases across the USA and gives encounter-level data on all hospital admissions of the relevant enrollees during this two-year period. For each admission, the data includes the patient diagnosis and characteristics, the identity of the hospital and the type of plan. Patient income is not included in the MEDSTAT data; I approximate this using the median income of families in the Zip Code Tabulation Area (ZCTA), taken from Census 2000 data. This very detailed dataset enables me to pin down quite accurately the effect of individual patient characteristics on their choice of hospitals.

My final analysis will investigate consumers' choice of managed care plans. Thus in an ideal world I would estimate consumers' hospital choices using data for managed care enrollees and identifying preferences using each individual's choice of hospital given the choice set specified by his or her plan. However, this is not feasible because the MEDSTAT data do not identify the hospital networks offered by each managed care plan: that is, the hospital choice set of managed care enrollees is unobserved. Instead I examine the choices made by indemnity and PPO enrollees, whose hospital choice sets are unrestricted. Two significant assumptions are required to apply the equation estimated here as an input to the plan choice equation estimated later. The first is that indemnity plan/PPO enrollees have the same preferences over hospitals as HMO/POS enrollees conditional on their diagnosis, income and location. This assumption has been made several times in the existing literature and may not be unreasonable.¹⁹ The average fee-for-service plan enrollee probably has different preferences over hospitals from the average managed care enrollee before he knows his diagnosis: for example, he may have a stronger desire for choice. However, when informed that he has a specific disease, he might well choose the same hospital as the average managed care enrollee of the same age and living in the same zip code.²⁰

The second assumption regards prices. PPO enrollees may be required to pay additional copays or deductibles if they choose to go out-of-network. These financial penalties, and the hospitals in the PPO network, are not identified in the dataset; that is, the 'price' of the hospital at the point of service is unobserved. I therefore assume that out-of-pocket prices charged to patients on the margin are zero for both PPO and indemnity patients. This may be reasonable, particularly where prices take the form of increased deductibles, since many of these patients are likely to have spent past their deductible before making their decision.^{21,22}

¹⁹ For example, Town and Vistnes (2001) use data on the hospital selection decisions of Medicare enrollees, assuming that the Medicare population's valuation of hospitals is a reasonable proxy for that of HMO enrollees. Capps *et al.* (2003) make a similar assumption to justify considering patients insured by Medicare, Medicaid, Blue Cross/Blue Shield and indemnity plans.

²⁰ I test this assumption by estimating the hospital choice model separately using MEDSTAT data for HMO/POS enrollees in Boston, MA, a market in which I observe that the vast majority of plans contract with all hospitals. (I observe this fact for 2003; I assume it also to have been true in 1997/8.) The estimated coefficients are not identical, but are broadly similar, to those estimated using PPO/indemnity enrollee data for Boston, MA only. Where HMO/POS enrollees derive positive (negative) utility from a hospital or characteristic, PPO/indemnity enrollees do the same, often with similar coefficient magnitudes. Only 3 out of 36 hospital dummy coefficients and 2 out of 32 interaction coefficients are different in sign across the two models and both significant at $p = 0.1$. While not overwhelming evidence, I take this to be sufficient to support the assumption, particularly given that no other approach is possible with the available data.

²¹ The average copay for PPO enrollees in my data was \$289 for an average stay of 4.8 days. At only around 3% of the average cost per admission, this is probably small enough to justify the zero-price assumption. I tested the assumption further by re-estimating the hospital choice model using data for indemnity enrollees only (roughly half the total sample). The results were similar to those for the main specification. No interaction coefficients had different signs in the two models and were both significant at $p = 0.10$. Not surprisingly, the standard errors were higher in the robustness test.

²² Three further assumptions are also implicit in the methodology. First, I assume that managed care enrollees have the same probability of admission to hospital conditional on age and gender as the overall population. Second, patients choosing managed care plans are assumed to place as high a value on their health as those choosing indemnity plans.

Table I. Descriptive statistics for hospitals, MEDSTAT dataset

	Mean	SD
Number of beds	286	193
Teaching status	0.20	0.40
For-profit	0.06	0.25
Registered nurses per bed	1.24	0.46
Cardiac services	0.72	0.37
Imaging services	0.42	0.26
Cancer services	0.60	0.41
Birth services	0.82	0.38

Notes: $N = 434$ hospital-years. Cardiac, imaging, cancer and birth services refer to four summary variables defined in Appendix A. Each hospital is rated on a scale from 0 to 1, where 0 indicates that no procedures in this category are provided by the hospital and a higher rating indicates that a less common service is provided.

Further details on the hospital data, and the other datasets used for the demand estimation, are given in Appendix A. Tables I and II set out summary statistics for the hospital dataset. I consider the 11 largest markets in the MEDSTAT data (those with over 1000 observations per market-year): the dataset therefore includes 217 hospitals, 434 hospital-years and 28 666 encounters in total. The hospitals have 286 beds and 1.24 registered nurses per bed on average; 20% are teaching hospitals; 51% of patients are enrolled in a PPO rather than an indemnity plan; 63% are women. The most common diagnoses are labor (17% of encounters) and cardiac admissions (11%). The employers providing data are not identified; however, the most common industry sectors are services (including hotels, personal services and health care; these comprise 46% of observations) and durable goods manufacturing (31% of observations). Other industry sectors represented include non-durable goods manufacturing, transport, communications and utilities, and state and local government.

4.2. Plan Networks

The link between the hospitals and the HMO/POS plans in each market is provided by a uniquely constructed dataset that defines, for every HMO/POS plan in the 43 markets covered by the plan characteristics data, the network of hospitals offered to enrollees in March/April 2003.²³ The dataset includes 516 HMO/POS plans and 665 hospitals in total (on average 12 plans and 15 hospitals in each market). The data are summarized in Section 2.2 above.

4.3. Plan Characteristics

The final dataset contains aggregate data on health plan characteristics, including plan enrollment, for the 516 HMO/POS plans covered by the plan networks dataset. Each potential plan choice is

Finally, I assume that employees in the MEDSTAT dataset are representative of all privately insured consumers in the market, conditional on age and gender.

²³ I consider all the markets covered by the Atlantic Information Services data described in Section 4.3 except New York. The markets are listed in Appendix A. I assume that enrollees in POS plans, like those in HMOs, are restricted to hospitals within their insurer's network.

Table II. Patient descriptive statistics, MEDSTAT dataset

	Mean	SD
<i>Diagnosis</i>		
Neurological	0.01	0.10
Cardiac	0.11	0.32
Labor	0.17	0.38
Baby	0.07	0.26
Digestive	0.09	0.28
Cancer	0.08	0.27
<i>Age</i>		
0–17	0.13	0.34
18–34	0.22	0.42
35–44	0.13	0.33
45–54	0.19	0.39
55–64	0.29	0.45
Over 64	0.03	0.17
<i>Industry</i>		
Manufacturing (durable)	0.31	0.46
Manufacturing (nondurable)	0.06	0.24
Transport, Communications, Utilities	0.02	0.14
Finance, Insurance, Real Estate	0.01	0.08
Services	0.46	0.50
State and Local Government	0.03	0.17
<i>Working status</i>		
Full time	0.74	0.44
Part time	0.002	0.05
Early retiree	0.19	0.39
Retiree	0.04	0.19
Female	0.63	0.48
PPO enrollee	0.51	0.50
Emergency admission	0.05	0.21
Distance to chosen hospital (miles)	12.40	9.67
Distance to all hospitals (miles)	22.78	11.85

$N = 28\,666$ encounters.

defined as an insurer–product (HMO/POS) market. Thus Aetna HMO is a separate choice from Aetna POS plan in Boston. Data are brought together from several sources. Atlantic Information Services provide data on enrollment and characteristics (premiums earned, number of enrollees, tax status and other information) for the third and fourth quarters of 2002, respectively. The *Weiss Ratings' Guide to HMOs and Health Insurers* includes additional plan characteristics for over 500 HMOs in fall 2002. Finally, data on plan performance are taken from the *Health Employer Data and Information Set* (HEDIS) and the *Consumer Assessment of Health Plans* (CAHPS) 2000 data, both of which are published by the National Committee for Quality Assurance (NCQA). These data measure clinical performance and patient satisfaction in 1999, and are taken from the first year in which NCQA required plans to hire third-party firms to audit the data.²⁴

²⁴ Individuals choosing health plans in 2002 would in fact have been informed by NCQA 2001 data, to which I do not have access. I therefore choose the HEDIS and CAHPS 2000 variables that are most highly correlated with their 1999 counterparts, assuming that their correlation with the 2001 data will also be high. All variables used have correlations with the 1999 data of over 0.65.

Table III. Descriptive statistics for HMO/POS plans

Variable	Definition	N	Mean	SD
Market share	Plan share of non-elderly market	516	0.03	0.04
Enrollment	Number of enrollees (thousands)	516	66.22	110.1
Premium pmpm (\$)	Premiums earned per member per month	478	140.75	44.27
Physicians per 1000 population	Number of physician contracts per 1000 pop. in markets covered by plan	418	1.56	1.51
Breast cancer screening	% of women aged 52–69 who received a mammogram within last 2 years	352	0.73	0.05
Cervical cancer screening	% of adult women who received pap smear within last 3 years	352	0.72	0.07
Check-ups after delivery	% of new mothers receiving a check-up within 8 weeks of delivery	351	0.72	0.11
Diabetic eye exam	% of adult diabetics receiving eye exam within last year	350	0.45	0.11
Adolescent immunization 1	% of children receiving all required doses of MMR and hep. B vaccines before 13th birthday	346	0.31	0.16
Adolescent immunization 2	% of children receiving all required doses of MMR, hep. B and VZV vaccines before 13th birthday	313	0.15	0.11
Advice on smoking	% of adult smokers advised by physician to quit	213	0.63	0.07
Mental illness check-up	% of members seen as outpatient within 30 days of discharge after hospitalization for mental illness	307	0.68	0.15
Care quickly	Composite measure of member satisfaction re: getting care as soon as wanted	304	0.75	0.05
Care needed	Composite measure of member satisfaction re: getting authorizations for needed/desired care	304	0.72	0.06
Age 0–2	Dummy for plans aged 0–2 years	516	0.01	0.08
Age 3–5	Dummy for plans aged 3–5 years	516	0.06	0.23
Age 6–9	Dummy for plans aged 6–9 years	516	0.17	0.37
Aetna	Plan fixed effect	516	0.15	0.36
CIGNA	Plan fixed effect	516	0.10	0.31
Kaiser	Plan fixed effect	516	0.03	0.16
Blue Cross Blue Shield	Dummy for ownership by BCBS	516	0.16	0.36
POS plan	Dummy for POS plan	516	0.35	0.49

Variable definitions and summary statistics for the plan dataset are given in Table III. The average market share of the HMO/POS plans in the dataset is 3%.²⁵ Premiums average \$141 per member per month; 35% of insurers are POS plans; 76% have been in existence for over 10 years. HEDIS scores vary widely, from an average rating of 0.15 (for the percentage of children receiving all required doses of MMR, hepatitis B and VZV vaccines before their 13th birthday) to an average of 0.73 (the proportion of women aged 52–69 who had received a mammogram within the previous two years). The two most frequently occurring plans are Aetna and CIGNA, with 15% and 10% of observations, respectively.

Several assumptions are needed to link the datasets. In order to apply the estimated parameters from the 11 markets in the hospital dataset to the 43 markets used for the plan choice model I assume that hospital preferences given demographics and diagnosis are fixed across markets and over time.²⁶ In addition, because patient diagnosis and severity of illness influence hospital choice

²⁵ Shares are measured as percentage of the non-elderly population in the market. See Appendix A for details of the methodology used to generate this variable and robustness tests of the resulting data.

²⁶ A comparison of the two sets of markets shows no significant difference in population, managed care penetration or demographic characteristics such as age, education, race or income. The characteristics of the hospitals in the two sets of markets are also similar except that the percentage of for-profit hospitals in the market is higher, in a test of size 0.05, in

but are not observed in the health plan data, I assume invariance of both factors across markets and over time. I then predict diagnosis probabilities in different markets based on age and sex and assume that the predicted choices from an equation that excludes severity can be interpreted as average choices across the severity distribution.²⁷

5. DEMAND: DETAILS ON EMPIRICAL APPROACH

5.1. Demand for Hospitals

The specification of the utility function giving rise to PPO/indemnity enrollees' demand for hospitals is given by

$$u_{ihl} = \delta_h + x_h v_{il} \beta + \varepsilon_{ihl} \quad (6)$$

where ε_{ihl} captures unobserved idiosyncratic tastes which are assumed to be iid distributed according to a Type 1 extreme value distribution. The hospital-specific variable δ_h is given by

$$\delta_h = \eta_h + x_h \alpha \quad (7)$$

where η, x are as defined in Section 3. This formulation implies that the share equation can be written as

$$s_h = \sum_{i,l} \frac{N_{il}}{N} \left(\frac{\exp(\delta_h + x_h v_{il} \beta)}{\sum_{p \in H} \exp(\delta_p + x_p v_{il} \beta)} \right) \quad (8)$$

where N_{il} is the number of individuals in consumer-type i who are hospitalized with diagnosis l , N is the number of individuals admitted to hospital in the market, and H is the set of hospitals in the market.

To ensure consistency of the coefficient estimates, I include hospital fixed effects in my estimation of equation (8): these give me the predicted $\hat{\delta}$'s.²⁸ Since all the variables in equation (8) are observed, the estimation can be performed using MLE. A number of interaction terms are also included. Previous studies have shown that distance traveled to hospital has a significant negative effect on utility: the distance between hospital and patient residence zip codes, distance squared, and distance interacted with market dummies and a dummy indicating an emergency admission are included to account for this. The other interactions are between patient characteristics (the seven diagnosis categories listed in Appendix A, income, a dummy for emergency admissions and

the 43 markets used for the plan choice model. I control for this difference by including hospital ownership details in the hospital demand model and the expected utility calculation.

²⁷ Linking the various sources of plan data also requires an assumption: since the network dataset was compiled after the AIS enrollment data, I assume that plans' hospital networks were stable between 2001, when individuals choosing plans for Q4 2002 made their decisions, and Q1–2 2003 when the network data were observed. It is reassuring to note that I gathered 25% of the plan network data twice: first in Q4 2002 and second in Q2 2003. There was very little change in networks between these two dates.

²⁸ The alternative would be to include uninteracted hospital characteristics (the x 's) in the specification: that is, to estimate α directly. However, this would imply ignoring unobserved quality η ; the estimates of α and β would therefore be inconsistent. I tried this alternative approach as a robustness test: the estimates of β changed only a little, but the $\hat{\alpha}$ were significantly different, often a factor of 10 smaller than the main specification.

a PPO dummy) and hospital characteristics (teaching status, tax status, the number of nurses per bed and four variables that summarize the services offered by each hospital).²⁹ Interactions that should have no effect (for example, a cancer diagnosis interacted with provision of birth services) are restricted to be zero.

Since the next step is to use the estimated coefficients in markets and years outside my sample, I need to predict the value of $\hat{\delta}$ using variables included in the plan dataset. Assuming that μ_h is the sampling error in δ_h , $\hat{\delta}_h = \delta_h + \mu_h$, I estimate the following equation:

$$\hat{\delta}_h = x_h\alpha + \eta_h + \mu_h \quad (9)$$

where η_h is unobserved hospital quality as defined above. I estimate equation (9) by regressing the estimated hospital dummy coefficients on 31 hospital characteristic variables taken from the AHA 1997 and 1998 data. Market fixed effects are also included.³⁰ To account for the heteroscedasticity introduced by μ , I adjust the standard errors of the OLS regression using White's (1980) heteroscedastic consistent standard error estimator.³¹

The estimated coefficients (the α 's in equation (9)) are used to predict the $\hat{\delta}$'s in equation (8) when the analysis is extended out-of-sample. The specification for predicted utility used in the subsequent analysis is therefore given by the following:

$$\hat{u}_{ihl} = \eta_h + x_h\hat{\alpha} + x_h\nu_{il}\hat{\beta} + \varepsilon_{ihl} \quad (10)$$

where $\hat{\alpha}$ and $\hat{\beta}$ are estimated in the two-stage process described above, ε is assumed to be distributed iid according to the Type 1 extreme value distribution, and η is unobserved hospital quality.

5.2. Expected Utility from Each Plan's Hospital Network

5.2.1 Computing the Expected Utility Variable

The analysis now moves from the hospital dataset to the plan data, taking the estimated coefficients from equations (8) and (9) and using them to predict the utility that each type of individual

²⁹ The hospital service variables are described in detail in Appendix A. Market and time fixed effects cannot be included since there is no outside option in the hospital choice equation. Removing emergency admissions from the dataset had little effect on the results. I also removed newborn babies: this had no effect on the estimated coefficients up to the second decimal place.

³⁰ The variables included are: number of beds; distance from City Hall; distance from City Hall squared; registered nurses per bed and nurses per bed squared; doctors per bed; dummy variables for JCAHO accreditation, cancer program approved by ACS, residency training program, medical school, member of Council of Teaching Hospitals of the Association of American Medical Colleges, Independent Practice Association, Foundation, Indemnity Fee for Service Plan, osteopathic hospital, and operating subsidiary corporations; Control/ownership dummies (control by county, church and for-profit partnership); and a list of service dummies (for neonatal intensive care, angioplasty, cardiac catheterization laboratory, computed tomography scanner, positron emission tomography, single-photon emission computed tomography, oncology services, obstetric services, emergency department, breast cancer screening, burn care, and alcohol/drug abuse inpatient care). Market fixed effects are needed to normalize the baseline hospitals—that is, the randomly chosen hospitals, one in each market, whose dummies are excluded from the choice model—across markets. The fixed effects are set to zero when the analysis is extended out of sample. I repeated the analysis without the market fixed effects as a robustness test: the final results were very similar to those from the main specification.

³¹ White's estimator does not adjust for serial correlation. However, the off-diagonal elements of $E(\mu\mu')$, the variance-covariance matrix for $\hat{\delta}$, are of order one hundredth the magnitude of the diagonal elements. I assume that $E(\eta\eta') = \sigma^2 I$, so by assumption there is no serial correlation in η . It therefore seems reasonable to assume no serial correlation in $\eta + \mu$.

expects to receive from the hospital network of each HMO/POS plan in his market.³² Individual i 's expected utility from the hospitals offered by plan j in market m is given by

$$EU_{ijm} = \sum_l p_{il} E_\varepsilon(\max_{h \in H_{jm}}(\hat{u}_{ihl})) \quad (11)$$

where p_{il} is the probability that individual i will be hospitalized with diagnosis l and H_{jm} is the set of hospitals offered by HMO/POS plan j in market m .^{33,34} Ben-Akiva (1973) shows that, under the assumption of Type 1 extreme value errors, this formula reduces to

$$EU_{ijm} = \sum_l p_{il} \log \left\{ \sum_{h \in H_{jm}} \exp(\eta_h + x_h \hat{\alpha} + x_h v_{il} \hat{\beta}) \right\} \quad (12)$$

Six diagnoses (l) are included here: all the categories described in Appendix B except newborn babies.³⁵

5.2.2 Methodologies to Account for Unobserved Hospital Quality η

The rest of the methodology would be straightforward if η was known; unfortunately I do not observe this variable directly. However, some analysis of the magnitude of η is possible using the available data. If we take draws η^* from the distribution of η and assign one draw as each hospital's unobserved quality (measured with error), then the error in the resulting expected utility variable $EU_{ijm}^{\eta^*}$ (the input to the plan demand equations) is

$$\text{Error} = EU_{ijm}^\eta - EU_{ijm}^{\eta^*} = \sum_l p_{il} \log \left\{ \sum_{h \in H_{jm}} C_{hl} \exp(\eta_h) \right\} - \sum_l p_{il} \log \left\{ \sum_{h \in H_{jm}} C_{hl} \exp(\eta_h^*) \right\} \quad (13)$$

where $C_{hl} = \exp(x_h \hat{\alpha} + x_h v_{il} \hat{\beta})$. This implies that the expected value of the error is zero, since η and η^* are iid draws from the same distribution. In a linear context, such as the logit demand

³² This is the point in the analysis when I begin to use the dataset listing the hospital network of each HMO/POS plan in 43 US markets. Up until now I have considered only indemnity/PPO enrollees in the 11 markets considered from the MEDSTAT dataset, whom I assume have access to all hospitals in the market. I now switch to HMO/POS enrollees whose choice is limited.

³³ The expectation over values of ε implies an assumption that each individual's ε is unknown when he chooses his plan. The alternative assumption, that he knows his ε when making the choice, leads to a different expression for expected utility: $EU_{ijm} = \sum_l p_{il}(\max(\hat{u}_{ihl}))$. I estimate this quantity as a robustness check by taking draws of ε from the Type 1 extreme value distribution, and assigning a draw to each simulated individual in each ZCTA-age-sex cell, for every possible diagnosis. Using this new expected utility variable in the health plan choice model had little effect on the final results.

³⁴ The expected utility variable is the equivalent of the expectation over the probabilities of different diagnoses of Capps *et al.*'s (2003) $V^{IU}(G, Y_i, Z_i, \lambda_i)$. My methodology therefore incorporates the 'option demand' calculations in that paper.

³⁵ Newborn infants are included in the hospital choice model but not the expected utility calculation. Inclusion in the hospital model implies that the baby is considered by the hospital to be a separate unit from its mother. In contrast, I assume that the health plan does not consider the baby as a separate entity, whose preferences should be considered separately from those of its mother, until after the birth episode. (The mother considers her preferences regarding the hospital in which to give birth when she chooses her health plan; she does not choose a plan for her baby until after its birth.) As noted in Section 5.1, removing newborn infants from the hospital choice model had very little effect on the results.

framework used as a benchmark model below, we can therefore adjust for the measurement error using an instrument correlated with $EU_{ijm}^{\eta*}$ but not with the error in measuring this variable. I use EU_{ijm}^{η} , the predicted expected utility from the plan's hospital network found using the methodology above but ignoring η , as an instrument for $EU_{ijm}^{\eta*}$.³⁶ I use information on the distribution of $\eta + \mu$ (the residuals from estimation of equation (9)) and assume that $\mu = 0$ when creating EU_{ijm}^{η} ; the test therefore provides an upper bound for the impact of η on the model's predictions.³⁷ The results are encouraging: the test's logit model predictions are very similar to those for the main model, implying that the assumption that $\eta = 0$ does not lead to significant bias. For this reason I assume $\eta = 0$ in the main analysis.

I conduct one robustness test for this assumption. I use a methodology first introduced by Hanushek (1974) and developed by Lewis (2000) to estimate the variance of η . The idea is to use OLS regression to estimate the variance of $\eta + \mu$, and then use the estimate of $E(\mu\mu')$ obtained in the first-stage choice model to back out a variance estimate for η . Details are given in Appendix B. I find that the variance of η is negligible compared to that of μ , implying that only a very small proportion of the 'lack of fit' in estimating equation (9) is due to unobserved hospital quality. Unfortunately this test is not very powerful in this application. For example, when I remove the number of nurses per bed (a plausible measure of hospital quality) from the specification the variance of η is still reported to be negligible. Even with this caveat, however, the combined results of this section are sufficient to justify setting $\eta = 0$ for the main analysis.

5.2.3 *Creating the Expected Utility Variable: Other Issues*

The coefficients on market–distance interactions, emergency–distance interactions, and interactions with PPO dummies and newborn infant dummies in the hospital utility equation are assumed to be zero in the expected utility calculation.³⁸ I assume that patients in POS plans, like those in HMOs, are restricted to hospitals within their insurer's network (that is, I ignore the possibility that their PCP might refer them to an out-of network hospital). I have no choice about making this assumption: my plan data are at the aggregate level so I have no way to observe which hospitals enrollees in particular POS plans actually visit. The expected utility of consumers choosing indemnity plans or no insurance is also needed. Consumers with indemnity plans can in general go to any hospital in the market and their expected utility is predicted accordingly.³⁹ I assume that uninsured consumers expect to gain zero utility from hospitals.⁴⁰

5.3. Demand for Health Plans Conditional on the Hospital Network Offered

The final step is to take the predicted expected utility from the hospital network of each plan, for each type of individual, and include it as an input to the plan demand equation. I begin with a

³⁶ The expected utility variable used in the logit framework, which as described below is effectively an average over individuals within the most populated ZCTA in the market, also has the property that $E(\text{error}/x, v) = 0$, implying that the instrumental variables methodology is valid.

³⁷ Assuming that μ and η are not negatively correlated.

³⁸ I assume that consumers do not consider the hospital they would choose in an emergency when choosing a health plan. My assumption that preferences over hospital characteristics, given diagnosis and demographics, are fixed across markets implies no need to include market–distance interactions: I assume the coefficient on distance is that estimated for Boston, my 'base case' market.

³⁹ Hospitals can choose not to contract with an indemnity or PPO plan if the plan's reimbursement rates are too low. I may therefore be overstating the number of hospitals from which these consumers can choose.

⁴⁰ This is not quite right: uninsured consumers can be admitted to public hospitals as emergency cases. The implications of this are discussed later in the paper.

simple logit model, ignoring the distribution of individual characteristics within markets, before moving on to a fuller model where variation in individual attributes is taken into account.

5.3.1 The Benchmark Model: Logit Formulation

The logit framework assumes common coefficients for all individuals. It has the advantage of generating an equation that is easy to estimate, making it a useful benchmark model. However, estimating such a model requires a plan-level, rather than plan–individual-level variable representing expected utility from the hospital network. I therefore define a representative agent in each market as an individual living in the most populated ZCTA and having the weighted average probability of diagnoses and of hospital admission of people resident in that ZCTA. I define the plan-level expected utility as that of the representative agent in the relevant market, that is:

$$EUrep_{jm} = \sum_l p_l^{\text{rep}} \log \left\{ \sum_{h \in H_{jm}} \exp(\eta_h + x_h \hat{\alpha} + x_h v_l^{\text{rep}} \hat{\beta}) \right\} \quad (14)$$

where p_l^{rep} is the weighted average probability of diagnoses of individuals in the most populated ZCTA, and v_l^{rep} is the vector of other characteristics of an individual in that ZCTA (income and location, the only individual characteristics other than diagnosis that are not set to zero, are both defined at the ZCTA level). I control for the unobserved η first by assuming $\eta = 0$ and then by using the instrumental variables methodology already described. The utility of consumer i from choosing plan j in market m is therefore given by

$$w_{ijm} = \xi_{jm} + z_{jm} \vartheta + \gamma_1 EUrep_{jm} + \omega_{ijm} \quad (15)$$

where premium is included in the observed plan characteristics z , and I assume that the ω_{ijm} is distributed iid according to a Type 1 extreme value distribution.⁴¹ Normalizing the utility of all consumers from the outside good (good 0) to be zero, I obtain the standard equation for estimation:

$$\log(s_{jm}) - \log(s_{0m}) = \xi_{jm} + z_{jm} \vartheta + \gamma_1 EUrep_{jm} \quad (16)$$

where s_{0m} is the share of the outside good in market m .

The logit model can be estimated using a simple two-stage least squares methodology (instrumenting for premiums since these are likely to be correlated with the unobserved quality variable ξ). The basic plan characteristics included in z are: premium per member per month,⁴² number of physicians per 1000 population in the market and age of the plan (coded into four dummies: less

⁴¹ Both the logit and the full demand specifications assume that consumer utility from the hospital network (which is a function of the probabilities of different possible diagnoses) is additively separable from other aspects of plan quality in the plan demand equation. In terms of a utility equation where expected health is the key input, this implies that the individual's expectation of the effect of a hospital on his health depends on his diagnosis, but his expectation of the effect of the plan is independent of diagnosis. This makes sense given that many plan functions are preventive whereas most if not all hospital interventions are specific to a diagnosis.

⁴² Premium is defined as average plan revenue per enrollee (including both employer and employee contributions). The measure is therefore somewhat noisy: this may lead to attenuation bias in the estimated premium coefficient. I would ideally adjust this variable for the severity of illness of the average enrollee. However, a regression of premiums on observed plan characteristics had an R^2 of just 0.34, implying that the adjustment would add too much additional noise to be helpful.

than 3 years, 3–5 years, 6–9 years and over 10 years of age). The HEDIS measures used are the breast cancer and cervical cancer screening rates; the rate of check-ups after live deliveries; the proportion of diabetic patients with annual eye exams; the proportion of adolescents receiving final immunizations before their 13th birthday; the proportion of smokers advised by their physician to quit; and the proportion of patients seen on an outpatient basis within 30 days of discharge from a mental illness admission. The CAHPS measures are ‘getting needed care’ and ‘getting care quickly’. Each is an aggregation of responses to several CAHPS questions. The CAHPS measures of plan performance are highly correlated with one another; so are the HEDIS measures. Missing variable dummies are included in all specifications. Insurer fixed effects are included for insurers that are active in at least 10 of the major markets defined by AIS,⁴³ and market dummies are also included in some specifications.

The model is completed by defining the outside good. The simplest definition to implement would be a composite of non-managed care private coverage and uninsured (I exclude Medicare by considering only the non-elderly population, and exclude Medicaid by assumption; see Appendix A for details). However, indemnity coverage and no coverage are at opposite ends of the spectrum in terms of price and many aspects of quality so this outside good would be non-homogeneous. Instead I define the outside good as ‘choosing to be uninsured’ and create a separate choice in each market, defined as ‘choosing indemnity or PPO insurance’, and assumed to be homogeneous within each market.⁴⁴ None of the data sources provides information on non-managed care coverage, so assumptions must be made to complete the dataset. Indemnity plans are assumed to be over 10 years old; to have premiums equal to the highest managed care premium in the relevant market; and to offer a physician network size equal to the largest offered by a managed care plan in the market. Indemnity plan performance ratings (both HEDIS and CAHPS) are assumed to equal the average of managed care plans in the market.⁴⁵ Average quality for uninsured consumers is not identified in the plan choice model (unless I make more assumptions or normalize one of the ‘inside’ goods) so I normalize it to zero.

The instruments used for the premium variable, in addition to the usual set of plan characteristics (the z 's), are the average hourly hospital wage and the average weekly nurse wage across the markets in which each health plan is observed to be active. Hospital wage data come from the Centers for Medicaid and Medicare Services (CMS) 1999; nurse wage data are from the Bureau of Labor Statistics 1999. The assumption required for these to be valid instruments is that health plan costs are correlated with premiums but not with unobserved health plan quality. The choice of instruments is a difficult issue. I use average wages across the plan's markets, rather than same-market wages, to generate variation across plans within each market. However, this requires an additional assumption that plans set their premiums

⁴³ These are Aetna, CIGNA, Coventry, Health Net, Kaiser Permanente, One Health, PacifiCare, United, UNICARE and Blue Cross Blue Shield plans.

⁴⁴ Assuming homogeneity of this option, and making the assumptions needed to define its characteristics, are problematic. The quality of this option will affect consumer choices of HMO and POS plans. Data limitations prevent me from modeling this option more completely; instead I include market fixed effects in the demand specification to account for cross-market differences in both this and the outside option.

⁴⁵ An alternative would be to assume that indemnity plan ratings = 0 for all HEDIS and CAHPS measures. I repeated the analyses under this assumption with little change in results.

partly centrally, in response to cost changes across all markets served, which may not be realistic.^{46,47}

The last issue is the need to adjust the estimated standard errors for the three-step estimation process being used here.⁴⁸ I need to take into account the fact that the expected utility variable is constructed from estimated parameters. This has no impact on the consistency of the results but, since the estimator is not adaptive, will affect the standard error estimates.

To incorporate the effect of these estimated parameters into the estimates of the variance–covariance matrix, I take advantage of the GMM structure of the estimation procedure. Incorporating estimated parameters into a GMM estimator is fairly straightforward. An extension of Pakes (1997) shows that if there exist first-stage parameter estimates β such that

$$\sqrt{N_1}(\beta_n - \beta_0) = L_1 \sqrt{N_1^{-1}} \sum_{h,i} f(x_h, v_i) + o_p(1) \quad (17)$$

and second-stage parameter estimates α such that

$$\sqrt{N_2}(\alpha_n - \alpha_0) = L_2 \sqrt{N_2^{-1}} \sum_h m(x_h) + o_p(1) \quad (18)$$

then, under standard regularity conditions, the third-stage parameter estimates θ_n are distributed asymptotically normal as

$$\sqrt{n}(\theta_n - \theta_0) \rightarrow_d N[0, (\Gamma'A^{-1}\Gamma)^{-1}\Gamma'A^{-1}VA^{-1}\Gamma(\Gamma'A^{-1}\Gamma)^{-1}] \quad (19)$$

where Γ is the derivative of the moment condition with respect to the parameters, A is the weight matrix and V is given by

$$V = \text{avar} \left(\frac{1}{\sqrt{n}} \sum_{j,m} \{g_{jm}(z, x, v, \alpha_0, \beta_0, \theta_0) + \sqrt{n}M_1(\beta_n - \beta_0) + \sqrt{n}M_2(\alpha_n - \alpha_0)\} \right) \quad (20)$$

$$= E \left(gg' + \frac{n}{N_1} M_1 v c_1 M_1' + \frac{n}{N_2} M_2 v c_2 M_2' \right) + 6 \text{ covariance terms}$$

where g_{jm} are the third-stage moments, M_1 is a matrix of derivatives of $g(z, x, v, \alpha_0, \beta_0, \theta_0)$ with respect to the elements of β_0 , M_2 is a matrix of derivatives of $g(z, x, v, \alpha_0, \beta_0, \theta_0)$ with respect to the elements of α_0 , and vc_1 and vc_2 are the variance–covariance matrices from the

⁴⁶ If the instruments were in fact correlated with unobserved quality this would imply an upward bias on the premium coefficient, making it less negative than the true value. This would add to the problems caused by attenuation bias in the premium coefficient.

⁴⁷ I do not instrument for the expected utility variable. This implies an assumption that plans' hospital networks are not correlated with unobserved plan characteristics such as the generosity of prescription drug coverage. I control for this issue in part by including large plan fixed effects. The fact that the correlation between expected utility and observed plan attributes is small—less than 0.09 for cancer screening, physician network size and the CAHPS variables—is also reassuring.

⁴⁸ The first step is the estimation of the hospital choice model including hospital fixed effects; the second is the regression of the fixed effect coefficients on hospital characteristics; the third is estimation of the health plan choice model including the expected utility variable.

first two stages. The covariance terms can be reduced to an expression depending solely on the expectations of products of M_1 , M_2 , L_1 , L_2 , and the individual moment conditions. Further details of the methodology used are given in Appendix C.

5.3.2 Full Demand Specification

Since the focus of the analysis is on the weight consumers attach to the expected utility variable in the plan demand equation, the obvious extension to the logit model is to allow a richer specification of this expected utility term. The full demand specification includes the full variable EU_{ijm} rather than the summary measure used in the logit formulation and therefore accounts for the impact of the distribution of individual locations, income and demographics within each market on plan market shares. This has the additional advantage of avoiding the logit model's well-known unattractive implication: the imposition of the independence of irrelevant alternatives (IIA) substitution pattern for an individual patient's choice of plan. The IIA assumption implies that cross-price effects are a function solely of plan shares and are independent of plans' relative positions in the characteristic space. This is clearly inaccurate to the extent that consumers substitute more readily between plans that are 'closer' in terms of characteristics (for example, a consumer switching from a high-premium, high-choice plan is more likely to choose another high-premium plan with similar qualities than a low-premium competitor that offers restricted choice of providers).

The utility of consumer i choosing plan j in market m in the full demand specification is given by

$$w_{ijm} = \xi_{jm} + z_{jm}\vartheta + \gamma_1 EU_{ijm} + \gamma_2 \frac{\text{prem}_{jm}}{y_i} + \omega_{ijm} \quad (21)$$

where prem_{jm} is plan j 's premium in market m , y_i is the median family income of individual-type i (defined by ZCTA), and the other variables are as specified above. This is similar to the model introduced in BLP (1995) in which random coefficients, which are functions of demographic variables taken from market-level census data, are interacted with product characteristics. The difference is that, in this equation, both consumer-specific terms (EU_{ijm} and y_i) are observed, either at the ZCTA level (in the case of income) or at the ZCTA-age-gender level (for the expected utility variable). Therefore simulation methods are not needed to evaluate the estimation algorithm.⁴⁹ The share equation reduces to

$$s_{jm} = \sum_i \left(\frac{n_i}{n_m} \right) s_{ijm}(\vartheta, \gamma) \quad (22)$$

where n_i is the number of individuals in consumer-type i , n_m is the number in the market, and $s_{ijm}(\vartheta, \gamma)$, the share of type- i individuals choosing plan j in market m , is defined by

$$s_{ijm}(\vartheta, \gamma) = \frac{\exp \left(\xi_{jm} + z_{jm}\vartheta + \gamma_1 EU_{ijm} + \gamma_2 \frac{\text{prem}_{jm}}{y_i} \right)}{1 + \sum_{k \in P_m} \exp \left(\xi_{km} + z_{km}\vartheta + \gamma_1 EU_{ikm} + \gamma_2 \frac{\text{prem}_{km}}{y_i} \right)} \quad (23)$$

⁴⁹ The ZCTA-level income measure, and consumer cell-level expected utility measure, are themselves assumed to be integrals over the types of people within the cell. The methodology used here therefore effectively integrates out over individuals, as in BLP.

The presence of the unobserved quality measure ξ implies that MLE cannot be used to estimate the model (as it was for the plan demand equation).⁵⁰ Instead, the contraction mapping introduced in BLP (1995) is used to transform equation (22) into a linear equation for ξ and the coefficients estimated using a GMM methodology, again as in BLP (1995).⁵¹ The variables included in z are the same as those used in the logit specification. I instrument for premium for the same reasons as in the logit framework; I add two instruments to the hospital and nurse wage variables used before. These are the average expected utility in the most populated ZCTA in the market (the $EUrep_{jm}$ variable defined for the logit model) and the average income across the other markets in which the plan is observed to be active. It is clear that $EUrep_{jm}$ should be correlated with premiums and the set-up of the model already implies that it is uncorrelated with unobserved quality ξ as required for a valid instrument. Some additional assumptions are needed to include the average income across the rest of the plan's markets as an instrument. First, the average income in each market must affect aspects of the plan's unobserved quality, such as promotional activity, in that market but not in others, and this activity must affect plan costs. Second, plans must set premiums in market j taking account of costs in other markets. Put simply, the assumption is that plans determine variables such as promotional activity locally but set premiums (at least partly) centrally.⁵²

Standard error adjustments are again needed to account for the variance introduced in the first two steps of estimation; the methodology used is exactly analogous to that described for the logit model.

6. DEMAND RESULTS

6.1. Hospital Demand: Results

Tables IV and V set out the results of the hospital demand specification. Table IV shows the results of the first stage: the estimation of the hospital choice model using MLE and including hospital fixed effects. The results are very much in line with the previous hospital choice literature, and are intuitive. If a hospital moves an additional mile away from a patient's home, this reduces the probability that the patient will choose it by 21%. Emergency admissions are even more sensitive to distance than non-emergencies.⁵³ The non-interacted effect of teaching hospitals on patient utility is subsumed in the fixed effects; however, the interaction coefficients show that patients with the most complex conditions (neurological diagnoses) attach the highest positive weight to these types of hospital. Cancer patients have very strong preferences for hospitals with a large number of nurses per bed. Women in labor are the only types of patient attaching a positive weight

⁵⁰ If plan fixed effects were included they would subsume the unobservable ξ and MLE could be used. However, the dataset is too small to successfully estimate all 516 plan dummy coefficients in addition to the other parameters.

⁵¹ I use heteroscedasticity-robust standard errors which also allow for correlated errors across the three stages of estimation. The weight matrix used in this analysis is the two-stage least squares weight.

⁵² The assumptions needed are somewhat troubling. However, the instrument does seem to be valid: the estimated coefficients change very little, and the standard errors go down, when the instrument is added. The Difference-in-Sargan or C-statistic for this instrument (Hayashi, 2000, p. 218) was 1.37, implying that the null that income is a valid orthogonal instrument cannot be rejected even in a test of size 0.24.

⁵³ The average distance from an individual to a same-market hospital in the data is 22.8 miles. The average distance to the chosen hospital is 12.4 miles. This distance increases to 12.9 miles when we consider only emergency cases. This may imply that some emergencies in the data are in fact non-urgent cases admitted through the emergency room. However, it may simply be that urgent cases are admitted to the closest hospital that offers the necessary services rather than the closest hospital overall. In either case, it is reassuring that the hospital demand results are robust to including or excluding emergency admissions.

Table IV. Hospital demand results, ML estimation

Interaction Terms	Variable	Estimated coefficient
	Distance (miles)	-0.215** (0.004)
	Distance squared	0.001** (0.000)
	Emergency* distance	-0.008** (0.004)
Interactions: Teaching	Cardiac	0.090 (0.060)
	Cancer	0.192** (0.069)
	Neurological	0.546** (0.175)
	Digestive	-0.145** (0.062)
	Labor	0.157** (0.048)
	Newborn baby	0.038 (0.075)
	Income (\$000)	0.007** (0.001)
	PPO enrollee	-0.067 (0.050)
Interactions: Nurses per bed	Cardiac	-0.096 (0.070)
	Cancer	0.445** (0.079)
	Neurological	0.130 (0.200)
	Digestive	-0.028 (0.076)
	Labor	-0.002 (0.063)
	Newborn baby	0.071 (0.087)
	Income (\$000)	0.005** (0.001)
	PPO enrollee	-0.099* (0.056)
Interactions: For-profit	Cardiac	-0.164 (0.181)
	Cancer	-0.197 (0.202)
	Neurological	0.229 (0.379)
	Digestive	0.195 (0.150)
	Labor	0.300** (0.107)
	Newborn baby	0.194* (0.122)
	Income (\$000)	-0.001 (0.003)
	PPO enrollee	-0.036 (0.090)
Interactions: Cardiac services	Cardiac	1.222** (0.134)
	Income (\$000)	0.001 (0.001)
	PPO enrollee	0.080 (0.088)
Interactions: Imaging services	Cardiac	-0.188** (0.094)
	Cancer	-0.052 (0.107)
	Neurological	-0.084 (0.287)
	Digestive	-0.182* (0.105)
	Labor	-0.071 (0.084)
	Newborn baby	0.398** (0.129)
	Income (\$000)	0.004** (0.001)
	PPO enrollee	-0.061 (0.072)
Interactions: Cancer services	Cancer	0.073 (0.082)
	Income (\$000)	-0.005** (0.001)
	PPO enrollee	0.087 (0.056)
Interactions: Labor services	Labor	3.544** (0.391)
	Newborn baby	3.116** (0.487)
	Income (\$000)	-0.003* (0.002)
	PPO enrollee	0.045 (0.077)
	Hospital fixed effects	Yes
	Pseudo-R ²	0.43

Notes: Maximum likelihood estimation of demand for hospitals using a multinomial logit model. Specification includes hospital fixed effects. $N = 28\,666$ encounters. Standard errors are reported in parentheses. ** Significant at $p = 0.05$; * Significant at $p = 0.1$.

Table V. Results from regression of hospital dummy coefficients on characteristics

Variable	Coefficient estimate
Neonatal intensive care	-1.79 (1.46)
Angioplasty	-1.51 (1.61)
Cardiac catheterization laboratory	5.90** (1.82)
Computed tomography scanner	6.53* (3.88)
Positron emission tomography	4.55** (1.57)
Single-photon emission computerized tomography	-3.23** (1.09)
Oncology services	2.90 (2.08)
Obstetric services	-1.93 (1.69)
Emergency department	-4.29* (2.20)
Breast cancer screening/mammograms	-4.44* (2.49)
Burn care	2.10 (1.88)
Alcohol/drug abuse inpatient care	0.51 (1.24)
Number of beds	0.01** (0.004)
Distance from City Hall	0.63** (0.19)
Distance from City Hall squared	-0.02** (0.01)
Registered nurses per bed	28.18** (4.97)
Nurses per bed squared	-9.74** (1.76)
Doctors per bed	3.78** (1.79)
JCAHO accreditation	6.83* (3.45)
Cancer program approved by ACS	4.32** (1.72)
Residency training program	-4.46** (1.46)
Medical school	4.72** (1.36)
Member of Council of Teaching Hospitals of the Association of American Medical Colleges	-0.29 (1.86)
Independent Practice Association—hospital	5.27** (1.15)
Foundation	-6.79** (2.04)
Indemnity Fee for Service Plan—hospital	2.56 (2.27)
Primarily osteopathic hospital	1.56 (3.72)
Operates subsidiary corporations	2.24** (1.05)
Controlled/owned by county	-9.51** (3.80)
Controlled/owned by church	-4.49** (1.41)
Controlled/owned by for-profit partnership	19.69** (5.06)
Constant	-32.10** (5.57)
Market fixed effects	Yes
R^2	0.44

Notes: Regression of estimated hospital fixed effect coefficients from multinomial logit model on hospital characteristics. $N = 434$ hospital-years. Robust standard errors are reported in parentheses; ** significant at $p = 0.05$; * significant at $p = 0.1$.

to for-profit hospitals. Finally, and reassuringly, cardiac diagnoses place a strong positive weight on hospitals with good cardiac services; cancer patients on hospitals with good cancer services (although this coefficient is not significant at $p = 0.1$), and women in labor and newborn babies have a strong preference for hospitals with good labor services.

The interaction terms between PPO dummies and hospital characteristics are almost all insignificant. This provides support for the assumption that indemnity and PPO enrollees have identical preferences over hospitals conditional on demographics and diagnosis. Most of the interactions between patient income and hospital characteristics are either positive or insignificant. This probably implies that higher-income individuals have a stronger preference for high-tech hospitals and more personalized care, rather than that they are willing to pay more for their services, although the latter may also be relevant for PPO enrollees choosing out-of-network

hospitals. The fit of the overall model is reasonable given that only a cross-section of data is used: the pseudo- R^2 is 0.43.⁵⁴

The results of the regression of the predicted hospital dummy coefficients on hospital characteristics are shown in Table V. Many of the hospital characteristics included (particularly the hospital service variables, each of which was set to 1 if the hospital offered the relevant service, and 0 if not) are correlated with each other, making the individual results difficult to interpret. However, it is clear from the results that consumers place a positive value on the number of nurses per bed, the number of doctors per bed, and overall on hospital accreditation. Hospitals at a higher distance from the city center are more popular with patients, probably because distance from the patient's home is negatively correlated, on average, with distance from the city. The fit of this model is similar to the first stage: the R^2 is 0.44.

6.2. Plan Demand: Results

Table VI sets out the results of the logit demand specification under the assumption that $\eta = 0$. Results are reported for three specifications. Specification 1 has no fixed effects. Specification 2 includes large insurer fixed effects which measure consumer preferences for specific national plans. Specification 3 includes both large insurer and market fixed effects; the latter control for

Table VI. Plan demand results, logit specification

	No fixed effects	Large plan fixed effects	Large plan and market fixed effects
Premium (\$00 pmpm)	-1.26 (3.15)	-1.09 (1.67)	-0.92 (1.10)
Expected utility from hospital network ($EUrep_{jm}$)	0.14 (0.14)	0.22* (0.11)	0.55** (0.14)
Physicians per 1000 population	0.30** (0.13)	0.23** (0.08)	0.21** (0.07)
Breast cancer screening	4.77 (4.66)	-1.71 (3.21)	-0.36 (2.48)
Cervical cancer screening	4.66** (1.83)	4.19** (1.69)	4.46** (1.75)
Check-ups after delivery	-0.53 (1.64)	0.26 (1.07)	0.14 (1.03)
Diabetic eye exams	0.39 (1.68)	-0.83 (1.19)	-1.20 (1.08)
Adolescent immunization 1	-0.77 (1.29)	-2.19* (1.08)	-4.11** (1.16)
Adolescent immunization 2	-1.74 (1.83)	2.19* (1.47)	3.16** (1.40)
Advice on smoking	-7.07** (2.76)	2.75* (1.90)	6.20** (1.80)
Mental illness check-ups	-0.34 (2.46)	2.02 (1.79)	2.67** (1.25)
Care quickly	6.64 (6.10)	4.55 (4.47)	0.75 (3.93)
Care needed	3.77 (6.58)	-1.85 (4.30)	0.81 (3.60)
Plan age: 0-2 years	-1.30 (0.98)	0.52 (1.17)	1.33 (0.94)
Plan age: 3-5 years	-2.31* (1.43)	-0.97* (0.53)	-0.63 (0.42)
Plan age: 6-9 years	-1.63* (0.83)	-0.26 (0.24)	-0.25 (0.22)
POS plan	-1.35** (0.22)	-1.10** (0.13)	-1.11** (0.13)
Constant	-9.38 (7.47)	-6.75* (3.78)	-10.94** (2.89)
Large plan fixed effects	No	Yes	Yes
Market fixed effects	No	No	Yes
R^2	0.362	0.592	0.671

Notes: Logit estimates of demand for health plans. $EUrep_{jm}$ is as defined in Section 5.3.1. Large plan fixed effects are included for insurers active in at least 10 of the markets considered. $N = 559$ insurers (516 HMO/POS plans and 1 indemnity/PPO option for each of the 43 markets). Standard errors (adjusted for the three-stage estimation process as described in Section 5.3.1) are reported in parentheses. ** Significant at $p = 0.05$; * significant at $p = 0.1$.

⁵⁴ The pseudo- R^2 is defined as $1 - (L1/L0)$, where $L0$ is the constant-only log likelihood and $L1$ is the full model log-likelihood. The model chi-squared(261) = 74 810, with p -value = 0.0000.

unobserved differences in the outside option across markets. Including large insurer fixed effects improves the fit of the model; adding market fixed effects improves it still further (the R^2 increases from 0.36 to 0.59 to 0.67 when both types of fixed effect are included).

The coefficient on premium is negative but not significant in any of the specifications; its magnitude suggests an insurer-perspective elasticity of -1.24 and that a \$5 increase in premiums per member per month would reduce the probability that a plan is chosen by 0.0012.⁵⁵ These figures are at the low end of the range found in previous studies. As noted in Section 5.3.1, the low figures may be caused partly by attenuation bias due to noise in the premium measure.⁵⁶

The key variable, $EUrep_{jm}$, has a positive coefficient in all three specifications, implying that consumers care about hospital networks: a plan's market share would be predicted to decrease if it excluded hospitals. The magnitude of the coefficient increases when market fixed effects are added;⁵⁷ it is significant at $p = 0.05$ in Specification 3. The most straightforward way to interpret the result is to compare the marginal effect of $EUrep_{jm}$ to that of plan premiums. The results imply that a one standard deviation increase in expected utility from the hospital network is equivalent to a \$39 reduction in premium per member per month (a change in premium of a little less than one standard deviation). So the effect of the hospital network on plan demand is important but not too large. The coefficient on the number of physicians per 1000 population is also positive and significant, implying that consumers prefer plans with large physician networks. The coefficients on HEDIS and CAHPS variables are difficult to interpret because the variables are fairly highly correlated with each other. However, most of the coefficients that are significant are positive as expected.

Consumers seem to significantly prefer HMO to POS plans. At first sight this is surprising, particularly since POS plans offer more flexibility to use hospitals outside the network. The reason is probably twofold. First, premiums set by POS plans are in general higher than their HMO equivalent; unfortunately the available data in general report the same premium amount for both HMO and POS plans in a given organization. Second, there is a selection problem: plans would prefer to attract HMO enrollees, who have more restricted benefits, rather than more expensive POS enrollees, and probably take unobserved steps (e.g. through promotions) to do so. This issue is beyond the scope of this paper; I assume no selection. Both these complicating factors will lead to a negative bias on the POS dummy coefficient.^{58,59}

⁵⁵ The insurer-perspective elasticity is based on the full premium rather than the out-of-pocket prices faced by enrollees.

⁵⁶ Examples of the previous literature include Cutler and Reber's (1996) estimate of -2 for the insurer-perspective elasticity. Royalty and Solomon (1998) estimate insurer-perspective elasticities between -1.02 and -3.5 . The elasticity estimates in Strombom *et al.* (2002) range from -0.8 to -5.2 . Buchmueller (2005) estimates an elasticity of -3.3 and that a \$5 premium increase reduces the probability that a plan is chosen by 0.002 to 0.008. If my results are affected by attenuation bias this will lead to an overestimate of the loss in consumer surplus from restricted networks. This issue is discussed further in the section on welfare below.

⁵⁷ The reason for the increase is probably that consumers are more likely to choose the outside option (being uninsured) in markets where hospitals on average offer higher utility. This effect will be absorbed into and bias down the coefficient on $EUrep_{jm}$ unless market fixed effects are included. The fixed effects account for differences in the outside option across markets, leaving the expected utility coefficient to measure the effect of differences in $EUrep_{jm}$ across plans within each market. For this reason, Specification 3 gives the most relevant results to the question being considered and will be used for the remainder of the paper.

⁵⁸ If POS plan premiums were \$118 (84%) higher than those of HMOs on average, this would explain the entire POS coefficient without any assumption regarding selection. Recent industry publications suggest a premium difference of around 30% (see, for example, the Ohio Chamber of Commerce Task Force Report on Provider Access, 2000), implying that the selection effect may also be important.

⁵⁹ The estimated coefficients for the missing variable dummies, and for the insurer and market fixed effects in the relevant specifications, are not reported here. All but one of the missing variable dummies have insignificant coefficients in

Several robustness checks are used to test the stability of the results to changes in assumptions and in the underlying data. First, the methodology is repeated using just the plans whose enrollment is reported by AIS at the PMSA level.⁶⁰ Second, I conduct the test of the assumption that unobserved hospital quality $\eta = 0$ described in Section 5.2.2. Finally, the analysis is repeated using the formula in footnote 33 to deal with the error term ε . All three robustness checks yield results that are extremely similar to those reported.

The results from the full demand specification are reported in Table VII. Estimates from Specification 3 (including both large insurer and market fixed effects) are reported; the equivalent estimates from the logit framework are also reported here for ease of comparison. The two sets of results are very similar. The estimates imply an insurer-perspective elasticity of -1.18 , consistent with the logit results but again at the low end of the range of previously published estimates. The coefficient on expected utility is higher in the full model: this makes sense because the distribution of consumer characteristics is now being taken into account. The coefficient on premium/income is very small and not significant: higher-income consumers seem not to have a different price elasticity of demand for health plans from their lower-income neighbors. Finally, some of the standard errors from the full specification are higher than those in the logit model, implying that too much may be being asked of the limited data available. Since the magnitudes of the coefficients are so similar, however, I take the full demand estimation results to be reasonable and use them in what follows.

7. WELFARE IMPLICATIONS OF SELECTIVE CONTRACTING

The demand estimates provide information on the trade-offs consumers are willing to make when choosing health plans: that is, on the extent to which hospital choice affects consumers' welfare and choice of plans compared to other plan characteristics such as premium, size of the physician network and clinical quality. The results indicate that consumers significantly prefer plans that offer larger hospital networks. The next step is to quantify the welfare benefits attached to an increase in network size. I consider the impact of a move, holding prices fixed, from the observed set of networks in each market to a hypothesized equilibrium in which every plan offers every hospital.⁶¹ The two dimensions of welfare are considered in turn: first consumer surplus and then provider and insurer profits.

7.1. A Simple Measure of Consumer Surplus

The methodology used in this section follows that discussed by Nevo (2001b) and based on McFadden (1981). Consumer i 's expected gain from a change in a plan's hospital network, assuming that no other plan characteristics (including its premium) change, is

$$\Delta_i = u_i^t - u_i^{t-1} \quad (24)$$

Specification 3. All the insurer fixed effects have negative coefficients, many of which are significant. Since insurer fixed effects are included only for the largest plans, the implication is that, all else equal, consumers prefer to enrol in local plans rather than those that operate at a national level.

⁶⁰ See Appendix A for details on this issue.

⁶¹ There is only one market in my sample in which all plans already offer an unrestricted choice of hospitals: this is Buffalo, NY. This market will therefore exhibit a zero welfare increase from the network change being considered.

Table VII. Plan demand results, logit and full specifications

	Logit specification	Full demand specification
Premium (\$00 pmpm)	-0.92 (1.10)	-0.94 (1.13)
Expected utility from hospital network ($EUrep_{jm}$ or EU_{ijm})	0.55** (0.14)	0.59** (0.21)
Premium (\$00 pmpm)/income (\$000 per year)	—	0.002 (43.9)
Physicians per 1000 population	0.21** (0.07)	0.21** (0.09)
Breast cancer screening	-0.36 (2.48)	-0.38 (2.66)
Cervical cancer screening	4.46** (1.75)	4.40** (2.09)
Check-ups after delivery	0.14 (1.03)	0.18 (1.38)
Diabetic eye exams	-1.20 (1.08)	-1.19 (1.60)
Adolescent immunization 1	-4.11** (1.16)	-4.11** (1.17)
Adolescent immunization 2	3.16** (1.40)	3.08 (3.76)
Advice on smoking	6.20** (1.80)	6.17** (2.08)
Mental illness check-ups	2.67** (1.25)	2.70** (1.30)
Care quickly	0.75 (3.93)	0.78 (5.63)
Care needed	0.81 (3.60)	0.85 (3.99)
Plan age: 0–2 years	1.33 (0.94)	1.36 (0.97)
Plan age: 3–5 years	-0.63 (0.42)	-0.64 (1.97)
Plan age: 6–9 years	-0.25 (0.22)	-0.25 (0.58)
POS plan	-1.11** (0.13)	-1.11** (0.13)
Constant	-10.94** (2.89)	-10.50* (5.65)
Large plan fixed effects	Yes	Yes
Market fixed effects	Yes	Yes

Notes: Logit and full (BLP) estimates of demand for health plans. $EUrep_{jm}$ and EU_{ijm} are as defined in Section 5.3. Large plan fixed effects are included for insurers active in at least 10 of the markets considered. $N = 559$ insurers (516 HMO/POS plans and 1 indemnity/PPO option for each of the 43 markets). Standard errors (adjusted for the three-stage estimation process as described in Section 5.3) are reported in parentheses. ** Significant at $p = 0.05$; * significant at $p = 0.1$.

where u_i^t and u_i^{t-1} are defined by

$$u_i^t = E_\omega \max_j (w_{ijm}^t) = E_\omega \max_j (V_{ijm}^t + \omega_{ijm}) \quad (25)$$

with w_{ijm}^t defined as in equation (21), so that

$$V_{ijm}^t = \xi_{jm} + z_{jm}\vartheta + \gamma_1 EU_{ijm}^t + \gamma_2 \frac{\text{prem}_{jm}}{y_i} \quad (26)$$

Note that this is the expected welfare gain from the perspective of the econometrician given the available data. A dollar-valued measure of welfare can be obtained using the method suggested by Hicks (1939) to create the equivalent variation (EV). The EV is the change in consumer wealth that would be equivalent to the change in consumer welfare due to the network modification. McFadden (1981) shows that

$$EV_{it} = \frac{1}{\alpha_i} (u_i^t - u_i^{t-1}) \quad (27)$$

where α_i is the negative of the coefficient on premium in the plan utility equation. Integrating analytically over the extreme value distribution of ω and summing over types of individual implies

that

$$EV_m = \sum_i \frac{n_i}{\alpha_i} \left[\ln \sum_{j \in m} \exp(V_{ijm}^t) - \ln \sum_{j \in m} \exp(V_{ijm}^{t-1}) \right] \quad (28)$$

where n_i is the population in ZCTA-age-sex cell i and the difference between V_{ijm}^t and V_{ijm}^{t-1} comes solely from the change in the hospital network offered by each plan.⁶² The expression for EU_{ijm}^{t-1} is calculated using the specification in equation (12) and considering a network that includes all hospitals in the market. EU_{ijm}^t takes the value already calculated using equation (12) and the observed networks.

I also adjust for hospital capacity constraints. When a hospital is predicted to be over 85% of its maximum capacity I reallocate patients randomly to non-capacity constrained hospitals in the market. The adjustment affects patients' hospital choices and therefore their values of EU_{ijm} ; these values are used to adjust down each patient's utility from each plan.⁶³

The results of this exercise are presented in Figure 3. Column 1 of the Table and Graph 1 give details on the EV per privately insured person per year: the dollar spending needed to compensate the average consumer in the market for a move back from unselective to observed networks. The variation in this measure across markets is high: from $-\$29$ in Las Vegas, NV to a very high $\$722$ per person year in Austin, TX. The median value is $\$15.70$ per person per year. Aggregating over consumers, this implies a total EV of $\$1.04$ billion per year for the 43 major markets in the data.^{64,65}

Two caveats should be mentioned here. First, the imprecise estimate of the premium coefficient implies some uncertainty about the welfare estimates. The noisy measure of premiums may also lead to attenuation bias and an overestimate of the consumer surplus loss from restricted networks. I investigate the impact of these issues by recalculating the welfare estimates setting the premium coefficient at different levels.⁶⁶ Increasing the elasticity to -3 , a value more consistent with the previous literature, reduces the predicted consumer surplus loss to just $\$0.27$ billion (or $\$4.03$

⁶² The utility offered by the outside option (being uninsured) is treated the same as that from any other choice: having normalized V_{i0m} to zero for each market, we include a term $e^0 = 1$ in each log sum expression.

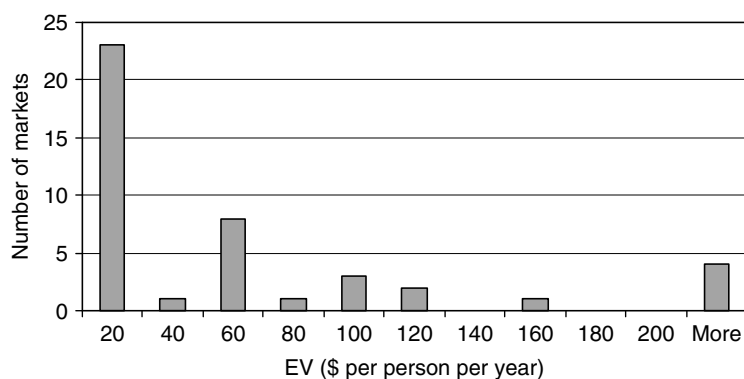
⁶³ The adjustment does not affect patients' choices of plan; I assume that consumers expect to have access to every hospital on the plan's list when they make their plan choice. Instead I account for the fact that some patients find, when they get sick, that they cannot access their preferred hospital and therefore receive less utility than expected from their chosen plan. The median equivalent variation falls only slightly as a result of this adjustment.

⁶⁴ Median values are used to reduce the impact of outliers. The negative value in Las Vegas, NV is caused because the move to unselective contracts displaces some high-valuation patients from capacity-constrained hospitals in favor of patients who value the hospital less. Only three of the 43 markets have negative EV per person values. Graph 1 of Figure 2 makes clear that Austin, TX, which has an EV per person of $\$722$ per year, is an extreme outlier. It includes a plan that is significantly different from others in the data. Scott and White Health Plan is managed by physicians, has strong links with the community, and is heavily focused on preventive rather than acute care. The demand model's ability to identify the effect of this plan's characteristics on consumer utility is limited: the specification implies that all attributes other than premiums and the hospital network affect all consumers equally, whereas Scott and White plan's strengths are likely to appeal to some consumers more than others. The predicted average quality of the plan, and therefore the estimated welfare effects for this market, are therefore likely to be biased. However, this is an issue that probably only affects a few markets (Austin, TX is clearly an extreme case), so reporting median results should be sufficient to control for the problem.

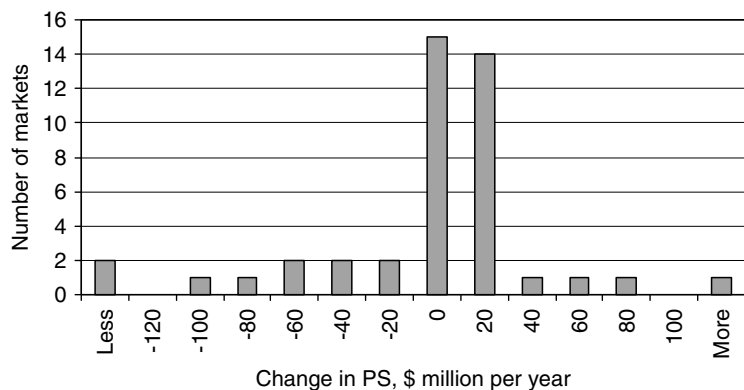
⁶⁵ I repeated the analysis holding Kaiser plan networks fixed and not allowing other plans to agree on contracts with Kaiser hospitals. In this case the total EV fell to $\$0.16$ billion per year.

⁶⁶ I would ideally have used a Monte Carlo analysis to calculate standard errors for the welfare estimates. Unfortunately the scale of this computation was infeasible: the demand analysis takes approximately 10 hours to run so the simulation would have taken 42 days if 100 draws were taken from the parameter distribution.

	Equivalent value per privately insured person (\$ per year)	Change in producer surplus per market (\$ per year)
Mean	\$59.71	-\$8.61 million
Median	\$15.70	-\$0.02 million
Maximum	\$722.49	\$288.2 million
Minimum	-\$28.84	-\$166.7 million
SD	\$123.60	\$64.6 million
Total gain implied by median values	\$1.04 billion	-\$0.80 million



Graph 1. Equivalent variation per privately insured person, \$ per year



Graph 2. Change in producer surplus per market, \$ million per year

Figure 3. Welfare effects of move to unrestricted hospital choice

per person). Smaller elasticities are required to make the welfare effects larger: for example, an elasticity of just -0.5 (improbably low compared to previous estimates) is needed to generate a consumer surplus loss of \$2.59 billion or \$39.10 per person per year. The second caveat, as noted in the Introduction, is that this analysis simply considers the loss to consumers from restricted hospital

choice, ignoring the benefits (from reduced hospital costs and margins) that these restrictions bring with them.⁶⁷

7.2. Welfare Analysis: Producer Surplus

I do not attempt to distinguish between hospital and plan profits in the producer surplus calculation since this would require estimating the prices paid to hospitals and therefore a full model of the supply side. Instead I consider the effect of the network change on the sum of hospital and plan profits (that is, on total producer surplus). The producer surplus generated by a given network depends on the number of consumers attracted to each plan and the cost of the hospitals they visit. I use the demand estimates to predict consumer flows to both plans and hospitals before and after the network change. These are used to calculate revenues (which are increasing in the number of consumers that choose high-premium plans) and costs (which increase if consumers on average choose more expensive hospitals). The total producer surplus to be divided between plan j in market m and the hospitals in its network is

$$PS_{jm} = \sum_i n_i s_{ijm} (\text{prem}_{jm} - \text{cost}_i) \quad (29)$$

where n_i is again the population in ZCTA-age-sex cell i , s_{ijm} is plan j 's share of type- i people in market m as specified in equation (23), and cost_i is the cost of treating a person of type i .⁶⁸ As before, define p_{il} as the probability that consumer-type i is admitted to hospital for diagnosis l . If s_{ihl} is the probability that a type- i person with diagnosis l will visit hospital h (the term in parentheses in equation (8)) and cost_{hl} is the cost that hospital h incurs by treating a patient with diagnosis l , then cost_i is given by

$$\text{cost}_i = \sum_l p_{il} E(\text{cost}_{il}) = \sum_l p_{il} \sum_{h \in H_{jm}} s_{ihl} \text{cost}_{hl} \quad (30)$$

The change in producer surplus when the plan switches to offering a free choice of hospitals, which will be divided between the plan and all the hospitals in its network, is given by

$$PS_{jm}^{\text{change}} = PS_{jm}^{\text{choice}} - PS_{jm}^{\text{observed}} \quad (31)$$

where $PS_{jm}^{\text{observed}}$ is the producer surplus predicted under the observed contracts and PS_{jm}^{choice} is that predicted when all plans in the market contract with all hospitals. The total change in producer surplus in the market is the sum of PS_{jm}^{change} over all plans j in market m . As in the consumer surplus calculation I take account of hospital capacity constraints by reallocating patients randomly from hospitals that are predicted to be over 85% of maximum capacity to other providers in the market. This reallocation affects the calculation of treatment costs.

⁶⁷ In many cases a premium increase simply represents a transfer from consumers to firms: this should therefore not have a significant effect on the calculation of total social welfare. However, there would be welfare effects in cases where the change prompted consumers to choose to be uninsured. Large premium increases could lead to large increases in the uninsured population.

⁶⁸ I assume that plans do not know consumers' idiosyncratic error terms, ε and ω , when they make their network decisions. They therefore predict consumer flows using share equations (8) and (22), which imply taking an expectation over the error terms.

A number of assumptions are made to simplify the analysis. First, I assume that plans have no variable costs except the payments made to hospitals. Any other variable costs would affect the surplus change calculation, even if they were constant across plans, because adding a new hospital could persuade consumers to switch from the outside option (being uninsured) to being insured, creating new variable costs that are not captured by the analysis. However, I do not have access to data on plan variable costs and therefore cannot measure these effects.⁶⁹ Data on total hospital expenses per admission are taken from the AHA 2001. Diagnosis-specific cost data would be preferable (since patient flows are predicted for each diagnosis separately) but were not available for this study. Finally, as in the consumer surplus calculations, I assume that plan premiums are fixed.

The results of the analysis are set out in Figure 3. The hypothesized network expansion causes a median reduction in producer surplus of just \$0.02 million per market per year. As with consumer surplus, there is considerable variation across markets, from an increase of \$288 million per year to a decrease of \$167 million per year. The median figure implies a total reduction in producer surplus of \$0.80 million per year across the 43 markets in the sample.⁷⁰

8. DISCUSSION AND CONCLUSION

This paper investigates the welfare effects of health plan restrictions on hospital choice. I use a three-stage method to estimate demand for health plans conditional on the hospital networks they offer. The results indicate that consumers place a positive and significant weight on their expected utility from hospital networks when choosing their plans. I find significant welfare effects of a move, at fixed prices, from plans' observed networks to a hypothesized equilibrium where every plan offers access to every hospital in its market. The predicted results are shown in the following table:

Welfare effect	Predicted change per year from move to unselective networks
Consumer surplus	\$1.04 billion
Producer surplus	−\$0.80 million
Total	\$1.04 billion

The predicted gain in consumer surplus outweighs the reduction in producer surplus: overall the results imply a benefit to society of just under \$1.04 billion per year.

The assumption of fixed prices and premiums implies that the results of the analysis represent just one part of the full welfare picture. The existing literature provides evidence on the effect of

⁶⁹ The analysis does allow for the existence of additional fixed costs, since these would cancel out when we consider the surplus change from a change in networks.

⁷⁰ Repeating the analysis holding Kaiser plan and hospital networks fixed generated a reduction in producer surplus of \$0.21 million per year. I also conducted the test of sensitivity to changes in the premium coefficient that I used for the consumer surplus estimates. An elasticity of -3 implies that markets with an increase in producer surplus outweigh those with a decrease: a move to unrestricted networks would increase total producer surplus by \$0.95 million. An elasticity of -0.5 translates to a producer surplus reduction of \$3.31 million from the change. Both these numbers are small compared to the corresponding consumer surplus effects.

managed care on plan costs and therefore on the potential for premium increases after a move back to broad hospital access. For example, Cutler *et al.* (2000) found that HMOs had 20–40% lower expenditures on the treatment of heart disease than indemnity plans. Virtually all of the spending difference came from lower unit prices rather than treatment or outcome differences; premium differentials generally matched the differences in reimbursement. Miller and Luft (1997) survey previous studies: a typical finding in this older literature is that HMOs save about 10% of the costs of indemnity insurance. The typical privately insured individual spends several hundred dollars annually for inpatient hospital care.⁷¹ Thus even the most conservative of these estimates translates to a saving from managed care that outweighs the \$15.70 per person per year welfare loss from restricted access. If selective contracting truly offers managed care plans a key lever to control prices, \$16 per person may not be an excessive price to pay.⁷²

A model of the supply side of the market is needed to understand these effects more fully. The analysis in Ho (2005) is a first step towards estimating such a model. Further research would clearly be useful to add to and confirm these results.

APPENDIX A: DETAILS ON HOSPITAL AND PLAN DATASETS

Hospital Dataset

The MEDSTAT MarketScan Research Database lists the date, the patient's age, sex, zip code and primary diagnosis (defined using ICD-9-CM codes), the identity of the hospital, and the type of plan (managed care; indemnity; PPO etc.) for every admission. I include six diagnosis categories in my analyses: cardiac; cancer; labor; newborn baby; digestive diseases; and neurological diseases. The specific ICD-9-CM codes included in each category are listed in Table VIII. These six categories account for 55% of the full dataset. The seventh diagnosis category, 'other diagnoses', comprises all other diagnoses included in the data. I also identify emergency admissions using the place of service and the type of service for each admission.

I supplement the MEDSTAT data with data from the American Hospital Association (AHA) for 1997 and 1998. This contains hospital characteristics, including details of location, ownership, accreditation, services provided, number of beds and numbers of admissions, for every hospital in the USA. I define distance from the patient's home to the hospital using the five-digit zip codes of each; distance of the hospital from the city center is defined as distance from the hospital's five-digit zip code to that of the City Hall. Four 'service' variables are created from the AHA data to summarize the services offered by each hospital: these are interacted with consumer characteristics in the hospital choice model. The summary variables cover cardiac, imaging, cancer and birth services. Each hospital is rated on a scale from 0 to 1, where 1 implies that the hospital offers the least common of a list of relevant services and 0 implies that it offers none of the services. Details are provided in Table IX.

⁷¹ For example, the National Center for Health Statistics reports that the average annual health care expense for privately insured consumers who had expenses was \$2026 in 2000. The Kaiser Family Foundation (www.statehealthfacts.org) estimates that 36.4% of personal health care expenditures go to hospital care. This implies a hospital cost of approximately \$700 per person per year.

⁷² The imprecise estimate of the premium coefficient implied some uncertainty regarding the precision of the welfare estimates. However, as discussed in Section 7.2, a very low elasticity of -0.5 would be needed to bring the welfare effects up to \$39 per person per year, a number that still seems small compared to the potential cost reductions.

Table VIII. Definition of diagnosis categories

Category	ICD-9-CM codes
Cardiac	393-398; 401-405; 410-417; 420-429
Cancer	140-239
Neurological	320-326; 330-337; 340-359
Digestive	520-579
Labor	644, 647, 648, 650-677, V22-V24, V27
Newborn baby	V29-V39

Notes: Patient diagnoses were defined by matching the principal diagnosis in the MEDSTAT claims data to 1998 ICD-9-CM diagnosis codes. This table lists the codes included in each of the six diagnosis categories used in the final analysis. Fifty-five percent of the total sample is accounted for by these six diagnoses; the remainder are distributed fairly evenly over a large number of other categories.

Table IX. Definition of hospital services

Cardiac	Imaging	Cancer	Births
CC laboratory	Ultrasound	Oncology services	Obstetric care
Cardiac IC	CT scans	Radiation therapy	Birthing room
Angioplasty	MRI		
Open heart surgery	SPECT		
	PET		

Notes: This table sets out the definition of the hospital service variables used to generate interaction terms for the hospital demand equation. Hospitals were rated on a scale from 0 to 1 within four service categories, where 0 indicates that no services within this category are provided by the hospital and a higher rating indicates that a less common (assumed to be higher-tech) service in the category is offered. The categories are cardiac, imaging, cancer and births. The services included in each category are listed in the table. The exact methodology for rating hospitals is as follows. If the hospital provides none of the services its rating = 0. If it provides the least common service its rating = 1. If it offers some service X but not the least common service its rating = $(1 - x)(1 - y)$, where x = the percentage of hospitals offering service X and y = the percent of hospitals offering the least common service.

The set of hospitals operating in each market is defined by the zip codes of consumers considered to reside in the market and the distance they are likely to be willing to travel to hospital. I consider patients whose home zip code is within the Primary Metropolitan Statistical Area (PMSA). Previous papers have considered hospitals within boundaries such as counties or states, those within 30 miles or 50 miles of the city center, or those hospitals within 30 or 50 miles of the individual patient's home zip code.⁷³ I limit the analysis to manageable proportions, while still including a reasonable sample of hospitals, by defining the market to include all hospitals within 30 miles of the city center. I include in the patient choice set every general medical/surgical hospital, other than those owned by the federal government, in the relevant market area.⁷⁴

⁷³ See, for example, Tay (2003).

⁷⁴ According to this market definition, the Boston market contains 37 hospitals; Chicago contains 72. If I included every hospital within 50 miles of the city center, Boston would have 72 hospitals and Chicago would have 86.

I consider the 11 largest markets in the MEDSTAT data (those with over 1000 observations per market-year). The markets are located in Massachusetts, Illinois, Arizona, Washington, Florida, and Michigan states; 5 of the 11 markets are in Michigan.⁷⁵ There are a total of 237 hospitals and 29 657 encounters⁷⁶ in these market-years. A number of observations are lost from the analysis because of missing hospital or individual data: if a variable (such as home zip code) is missing for a given individual, that individual is excluded from the analysis. If a variable (such as services provided or location) is missing for a given hospital, the missing data are filled in using surrounding years of AHA data where possible; otherwise that hospital, and all individuals who chose it, are excluded. The final dataset comprises 217 hospitals, 434 hospital-years and 28 666 encounters in total.⁷⁷

Plan Networks

The 516 plans for which network data were collected comprise all HMO/POS plans in 43 markets, as defined by the Atlantic Information Services data discussed in section 4.3.⁷⁸ The list of potential hospitals comprises all general medical/surgical hospitals listed by AHA 2001 that have more than 150 beds, are not owned by the federal government, and are located in the relevant PMSA. In smaller PMSAs, where there were fewer than 10 such hospitals, facilities with over 100 beds were included.⁷⁹

The expected utility calculation requires the different types of consumers to be defined using variables observed in the aggregate plan data; I therefore define types by sex, age group (0–17; 18–34; 35–44; 45–54; 55–64) and zip code tabulation area (ZCTA) of residence. There are 10 cells per ZCTA (five age groups and two sexes within each), and a total of 6363 ZCTAs across the 43 markets (an average of 148 in each market). The number of people in each ZCTA–age–sex cell is found using Census 2000 data from GeoLytics.⁸⁰ Diagnosis probabilities given age, sex,

⁷⁵ The markets are Boston MA, Chicago, IL, Dayton–Springfield, OH, Orlando, FL, Phoenix, AZ, Seattle, WA, and Detroit, Grand Rapids, Kalamazoo–Battle Creek, Lansing, and Saginaw Bay, all in MI.

⁷⁶ Each encounter is an individual admission; 51% of these are for PPO enrollees. If a single patient is admitted more than once in the two-year time period, I assume that the admissions represent independent choices.

⁷⁷ This attrition will bias the results if it is non-random. The characteristics of consumers whose data are missing are unlikely to be biased in any direction. While it is possible that smaller hospitals are less likely to report data to AHA, this seems unlikely because the existence of missing data in one year does not predict whether data in the subsequent or previous year will be missing.

⁷⁸ I consider all the markets covered by the AIS data except New York. The markets are: Atlanta, GA, Austin, TX, Baltimore, MD, Boston, MA, Buffalo, NY, Charlotte, NC, Chicago, IL, Cincinnati, OH, Cleveland, OH, Columbus, OH, Dallas, TX, Denver, CO, Detroit, MI, Fort Worth, TX, Houston, TX, Indianapolis, IN, Jacksonville, FL, Kansas City, MO, Las Vegas, NV, Los Angeles, CA, Miami, FL, Milwaukee, WI, Minneapolis, MN, New Orleans, LA, Norfolk, VA, Oakland, CA, Orange County, CA, Orlando, FL, Philadelphia, PA, Phoenix, AZ, Pittsburgh, PA, Portland, OR, Sacramento, CA, St Louis, MO, Salt Lake City, UT, San Antonio, TX, San Diego, CA, San Francisco, CA, San Jose, CA, Seattle, WA, Tampa, FL, Washington, DC, and West Palm Beach, FL.

⁷⁹ This simple definition of the hospitals in the market was not used in the hospital dataset because, in some of the smaller markets included in that data, it seems likely that patients would choose to travel outside the MSA to go to hospital. The 30-mile radius definition increased the hospital choice set in these markets. The plan dataset covers only larger MSAs, for which a 30-mile radius would reduce the hospital choice set in almost all cases.

⁸⁰ The analysis requires ZCTA–age–sex cells to be defined for an entire MSA. ZCTAs, rather than zip codes, were used because Census data from 2000 onwards use the former rather than the latter. Some ZCTAs cross MSA boundaries, implying that this aggregation to MSA level is not exact. However, the difference in definitions accounts for a less than 8% difference between the MSA population defined by the MSA boundary and that defined by the boundaries of its constituent ZCTAs. Since residents of contiguous ZCTAs are likely to have similar characteristics (income, distance from hospitals and age/sex distribution), I assume that ZCTA and MSA boundaries are perfectly aligned.

and admission to hospital are estimated from the MEDSTAT data using probit analysis; probabilities of admission to hospital given age and sex are taken from the National Hospital Discharge Survey 2000.

AHA data for 2001 (the most recent year for which data were available) were used in the expected utility calculation. A number of hospitals have missing data for AHA 2001. To avoid dropping these from the choice set, the missing data were filled in using previous years of AHA data where possible, and if necessary (for 16 hospitals) using information provided in individual hospital websites.⁸¹

Two plans in my sample offer none of the hospitals in the defined choice set but are licensed in and have significant numbers of enrollees in the market. Each of these offers a number of smaller hospitals in the relevant area (too small to be included in the base list of hospitals used to define plans' hospital networks). The expected utility from these plans' networks cannot be calculated using my dataset. I assume that a substantially restricted choice set reduces expected utility compared to other plans in the market; I therefore assign each type of consumer the minimum possible utility from any hospital in the market, given each diagnosis, if he or she chooses one of these two plans.

Plan Characteristics

The two datasets from Atlantic Information Services are the *HMO Enrollment Report* and *HMO Directory 2003*. Both are based on plan state insurance filings. The enrollment data give detailed enrollment for every HMO and POS plan in 40 major markets in the USA.⁸² The characteristic data cover all commercial health plans in the USA.

The unit of observation for the NCQA data is the NCQA plan identifier, which does not correspond exactly to the identifier for the AIS enrollment data. I matched the NCQA and AIS datasets at the insurer-product-market level; in cases where multiple NCQA plans correspond to one AIS plan I used the mean rating over NCQA plans. Similarly, the Weiss data and AIS characteristic data do not correspond perfectly to AIS enrollment data plan identifiers: both contain more aggregated data (for example, characteristics are provided for Aetna Florida rather than Aetna Jacksonville; Aetna Miami etc.) and often covers only HMOs. I matched the two datasets to the AIS enrollment data at the insurer-product-market level where possible, and at the insurer-market level otherwise. I matched aggregate data to all plans within the geographic area, and if no POS data were given separately I matched the plan's HMO characteristics to both HMO and POS plan types.

Missing data represent a significant issue. Of the 516 observations considered, 162 (31.4%) do not have HEDIS data and 212 (41.1%) do not have CAHPS data. Most of the plans without data did not respond to NCQA data requests; many did not provide information for any of the HEDIS and CAHPS categories used in this paper. There are similar problems with the price measure used: premium earned per enrolled member per month. In most cases (354 observations) this measure is calculated from AIS data on both premium and enrollment. Both inputs come from Weiss for 120

⁸¹ Data were taken from hospital websites in 2003. The same data were not used to fill in 1997/98 characteristics since hospitals are likely to have changed the services offered over the intervening five-year period.

⁸² AIS works with individual plans to disaggregate their base data. The data include some Medicare-only and Medicaid-only insurers as well as commercial plans; my analysis excludes the former and examines only plans that accept commercial business. I also exclude plans with fewer than 100 enrollees and/or no hospitals in the relevant market; I assume that these plans primarily serve neighboring areas. AIS publishes the data for 40 markets. I disaggregate this to 43 markets; see later in this section for methodology.

observations (where one or both pieces of information were missing in the AIS data). Price data are missing for 42 plans (8.2% of observations). Dropping plans with missing data (particularly NCQA data) could cause selection bias because the plans that failed to respond to NCQA requests are likely to be smaller or have lower quality than those that provided data. Instead I include these plans and add dummy variables that indicate missing premium and characteristic data.⁸³

I convert the AIS enrollment data to market share using the total non-elderly population of the MSA as defined by the Census Bureau in 2000 as the denominator.⁸⁴ The share uninsured and the share in PPO/indemnity plans in each market are also needed for the analysis. Census Bureau data are used to find the number of non-elderly uninsured; the difference between the total non-elderly population and the sum of uninsured and insured by HMO/POS plans is assumed to be indemnity/PPO coverage. One assumption is implicit in this methodology. The publicly insured, non-elderly (Medicaid) population should ideally be excluded explicitly from these groups. When this was done using Census data some markets had very low or negative implied indemnity/PPO market shares. I therefore ignore the existence of the non-elderly publicly insured. The problem is caused by errors in the AIS enrollment data; I assume that the errors are randomly distributed across plans and markets and therefore will not bias the results.

Plan enrollment in 17 of the 40 markets is reported at the Consolidated Metropolitan Statistical Area (CMSA) rather than the PMSA level. Most of these⁸⁵ are essentially one major city surrounded by very minor or inseparable suburbs: in these cases I perform the analysis as if the CMSA is a PMSA centered in the major city. However, for five CMSAs⁸⁶ the population was sufficiently diffuse to warrant considering the individual PMSA/MSAs separately (for example, consumers in Miami would probably not consider which hospitals were available in Fort Lauderdale when choosing a health plan). I disaggregate enrollment in these five CMSAs to PMSA/MSA level using the populations of the counties where each plan was licensed; plans in the nine largest of these PMSA/MSAs are included in the plan choice model.⁸⁷

Atlantic Information Services report enrollment at the level of detail provided by the insurer.⁸⁸ For 163 out of 516 plans (31.6%), enrollment data are reported at the state or (for three plans) national level rather than the MSA level. I disaggregate the state- or national-level data to MSA level, again using the populations of the counties in which each plan was licensed, and estimate the models using all 516 observations. I perform a robustness test of the effect of this approximation by re-estimating the model using just the 353 plans for which MSA-level data are provided; the missing HMO/POS plans are pulled into the 'indemnity/PPO' option. The results obtained through the two different methodologies are very similar.

I test the final market share data by comparing the implied total HMO share in each market to published data on market shares by state. The results are encouraging. For example, the

⁸³ The other alternative would be to fill in the missing observations with previous years' data. This is not attempted, both because plans with missing data in 2000 often had missing data in previous years and because cross-year correlations in reported data for a given plan are low (much lower than for the hospital data where this approach was used).

⁸⁴ I exclude people aged over 64 from the plan demand equation in order to exclude Medicare enrollees.

⁸⁵ Examples are Cincinnati–Hamilton, Cleveland–Akron, and Sacramento–Yolo.

⁸⁶ These are Dallas–Fort Worth, Los Angeles–Riverside–Orange County, Miami–Fort Lauderdale and San Francisco–Oakland–San Jose.

⁸⁷ One market, New York, was included by AIS but excluded from my analysis because the market boundaries were difficult to define; 43 markets are therefore considered in total.

⁸⁸ This is an advantage compared to some other vendors' market share data, in which researchers are not informed which plans provided data at the MSA level and for which the vendor disaggregated data originally provided at the state or national level.

predicted total HMO share in Boston is 57% of the non-elderly (and by assumption non-Medicaid) population; published data from the Kaiser Foundation (www.statehealthfacts.org) implies that HMOs have a 53% share of the non-elderly non-Medicaid population in Massachusetts.

APPENDIX B: ESTIMATING THE MAGNITUDE OF η

This appendix outlines the methodology used to estimate the variance of unobserved hospital quality η . Recall the regression defined by equation (9):

$$\hat{\delta}_h = x_h\alpha + \eta_h + \mu_h$$

where μ_h is the sampling error in δ_h , defined by

$$\hat{\delta}_h = \delta_h + \mu_h$$

Assume that μ and η are independent, that η has a normal distribution with mean zero and $E(\eta\eta') = \sigma^2 I$, and that

$$E(\mu_i\mu_j) = \begin{cases} \omega_i^2 & \text{if } i = j \\ \omega_{ij} & \text{otherwise} \end{cases}$$

The assumptions imply that the error variance for the regression analysis, defined as $u_h = \eta_h + \mu_h$, will not be homoscedastic. Following Hanushek (1974) and Lewis (2000), begin by running an OLS regression of $\hat{\delta}$ on x and defining \hat{u} as the residual vector. The expectation of the sum of squared residuals from this OLS regression can be written as

$$E\left(\sum_h \hat{u}_h^2\right) = E(u'u) - \text{tr}((X'X)^{-1}X'\Omega X)$$

where Ω is the variance–covariance matrix of the vector of regression residuals u and tr is the trace operator.⁸⁹ It can be shown that this implies

$$\sigma^2 = \frac{1}{H-k} \left[E\left(\sum_h \hat{u}_h^2\right) - \sum_h \omega_h^2 + \text{tr}((X'X)^{-1}X'GX) \right]$$

where H is the number of observations in the sample, k is the number of variables in X , and $G = \Omega - \sigma^2 I$ is the variance–covariance matrix of $\hat{\delta}$ in the first-stage hospital choice model. ($\sum_h \omega_h^2$ is the trace of this matrix.) An unbiased estimate of σ^2 can therefore be found using the following equation:

$$\hat{\sigma}^2 = \frac{1}{H-k} \left[E\left(\sum_h \hat{u}_h^2\right) - \sum_h \omega_h^2 + \text{tr}((X'X)^{-1}X'GX) \right]$$

where $\sum_h \hat{u}_h^2$ is the sum of squared residuals from the OLS regression of $\hat{\delta}$ on x .

⁸⁹ The trace of a square matrix is the sum of its diagonal elements.

I follow this methodology and find that $\hat{\sigma}^2 < 0$. As noted by Lewis, this result is possible in small samples and can be assumed to imply that $\sigma^2 = 0$.

APPENDIX C: ADJUSTED STANDARD ERROR ESTIMATES

This appendix provides further details on the method used to adjust the standard errors for the three-step demand estimation process. Pakes (1997) proves consistency and asymptotic normality of two-step GMM estimators. The methodology used in this paper is an extension to the three-step case. The necessary proofs are straightforward and are not given here; the following instead provides a derivation of the functional forms used in estimation. Consider the three steps of the estimation:

1. In Step 1 the following equation is considered:

$$u_{ihl} = \delta_h + x_h v_{il} \beta + \varepsilon_{ihl}$$

MLE provides a consistent estimator of the true β , which we call β_n . The sample analogue of the moment condition is

$$F_n(\beta) = \frac{1}{N_1} \sum_{i,h,l} f_{ihl}(x, \beta) = \frac{1}{N_1} \sum_{i,h,l} I_{ihl} \frac{\partial \log s_{ihl}(x, \beta)}{\partial \beta}$$

where s_{ihl} is the expression in parentheses in equation (8). Define

$$\Gamma_1 = \frac{\partial F}{\partial \beta}$$

and

$$L_1 = (\Gamma_1' \Gamma_1)^{-1} \Gamma_1$$

where

$$F(\beta) = E f(x, \beta)$$

2. In Step 2, a consistent estimate α_n of the true α in the following equation is found using OLS:

$$\hat{\delta}_h = x_h \alpha + u_h$$

The sample analogue of the moment condition here is

$$M_n(\alpha) = \frac{1}{N_2} \sum_h m_h(x, \alpha) = \frac{1}{N_2} \sum_h x_h (\hat{\delta}_h - x' \alpha)$$

Define

$$\Gamma_2 = \frac{\partial M}{\partial \alpha}$$

and

$$L_2 = (\Gamma_2' \Gamma_2)^{-1} \Gamma_2$$

where, analogous to Step 1

$$M(\alpha) = Em(x, \alpha)$$

3. Finally, consider Step 3. Considering the logit equation, and assuming a unit weight matrix for simplicity, we use two-stage least squares to find a consistent estimate, ϑ_n , of the true parameter vector ϑ in the following equation:

$$y_{jm} = \log(s_{jm}) - \log(s_{om}) = z_{jm}\vartheta + \omega_{jm}$$

where

$$z_{1jm} = EUrep_{jm} = h(z, x, v, \beta_n, \alpha_n)$$

The sample analogue of the moment condition here is

$$G_n(\vartheta, h(\beta_n, \alpha_n)) = \frac{1}{n} \sum_{j,m} g_{jm}(z, x, v, \alpha_n, \beta_n, \vartheta) = \frac{1}{n} \sum_{j,m} W_{jm}(y_{jm} - z_{jm}\vartheta)$$

where W_{jm} is a vector of instruments. Defining Γ_3 and L_3 analogously to the definitions in Steps 1 and 2 above, it can be shown that

$$avar(\sqrt{n}(\vartheta_n - \vartheta)) = (\Gamma'_3 \Gamma_3)^{-1} \Gamma'_3 V \Gamma_3 (\Gamma'_3 \Gamma_3)^{-1}$$

where

$$V = avar \frac{1}{\sqrt{n}} \left[\sum_{j,m} g_{jm}(\vartheta, h(\beta, \alpha)) + \frac{\partial g}{\partial h} \frac{\partial h}{\partial \beta} (\beta_n - \beta) + \frac{\partial g}{\partial h} \frac{\partial h}{\partial \alpha} (\alpha_n - \alpha) \right]$$

that is, the asymptotic variance of the first three terms of the Taylor series expansion for $\sqrt{n}G_n$. If we define

$$M_1 = \frac{1}{n} \sum_{j,m} \frac{\partial g}{\partial h} \frac{\partial h}{\partial \beta}$$

and M_2 as the analogous expression for Step 2, we can write

$$V = avar \left[\frac{1}{\sqrt{n}} \sum_{j,m} g_{jm} + \sqrt{n}M_1(\beta_n - \beta) + \sqrt{n}M_2(\alpha_n - \alpha) \right]$$

The fact that α_n and β_n are GMM estimators implies that

$$(\beta_n - \beta) = \frac{L_1}{N_1} \sum_{i,h,l} f_{ihl}(\beta) + o_p(1)$$

and similarly for $(\alpha_n - \alpha)$. This results in the following formula for V :

$$V = E(gg') + \frac{n}{N_1} M_1 v c_1 M_1' + \frac{n}{N_2} M_2 v c_2 M_2' + \sqrt{\frac{n}{N_1}} E(gf' L_1' M_1) + \sqrt{\frac{n}{N_1}} E(M_1' L_1 f g')$$

$$\begin{aligned}
& + \sqrt{\frac{n}{N_2}} (Egm'L'_2M_2) + \sqrt{\frac{n}{N_2}} E(M'_2L_2mg') + \frac{n}{\sqrt{N_1N_2}} E (M'_1L_1fm'L'_2M_2) \\
& + \frac{n}{\sqrt{N_1N_2}} E (M'_2L_2mf'L'_1M_1)
\end{aligned}$$

where vc_1 and vc_2 are the variance–covariance matrices for the first and second steps, respectively. The first three terms in this expression represent the impact of the variance–covariance matrices from each of the three steps on the final standard errors; those for the first two steps are increasing in the derivative of the third-step moment with respect to the first- and second-step estimators respectively. The last six terms measure covariance effects, allowing for correlated errors across the three steps of estimation. The required standard errors are estimated by replacing expectations with sample averages in the above expression.

ACKNOWLEDGEMENTS

This paper is a revised version of the first chapter of my dissertation. I am grateful to David Cutler and Ariel Pakes for their invaluable support and advice. I also thank John Asker, Nancy Dean Beaulieu, Joy Ishii, Julie Mortimer, Michael Ostrovsky and seminar participants at Harvard University for helpful comments and suggestions. All remaining errors are my own.

REFERENCES

- Beaulieu N. 2002. Quality information and consumer health plan choices. *Journal of Health Economics* **21**(1): 43–63.
- Ben-Akiva M. 1973. Structure of passenger travel demand models. PhD thesis, Department of Civil Engineering, MIT, Cambridge, MA.
- Berry S, Levinsohn J, Pakes A. 1995. Automobile prices in market equilibrium. *Econometrica* **60**(4): 889–917.
- Buchmueller T. 2005. Price and the health plan choices of retirees. NBER Working Paper 11395.
- Buchmueller T, Feldstein P. 1997. The effect of price on switching among health plans. *Journal of Health Economics* **16**(2): 231–247.
- Burns L, Pauly M. 2002. Integrated delivery networks: a detour on the road to integrated health care? *Health Affairs* **21**(4): 128–143.
- Burns L, Thorpe D. 2001. Why provider-sponsored health plans don't work. *Healthcare Financial Management*, Resource Guide: 12–16.
- Capps C, Dranove D, Greenstein S, Satterthwaite M. 2001. The silent majority fallacy of the Elzinga–Hogarty criteria: a critique and new approach to analyzing hospital mergers. NBER Working Paper 8216.
- Capps C, Dranove D, Satterthwaite M. 2003. Competition and market power in option demand markets. *RAND Journal of Economics* **34**(4): 737–763.
- Chernew M, Gowrisankaran G, McLaughlin C, Gibson T. 2004a. Quality, price, and employers' choice of health plan. *Journal of Health Economics* **23**: 471–492.
- Chernew M, Gowrisankaran G, Scanlon D. 2004b. Learning and the value of information: the case of health plan report cards. Mimeo, Washington University in St. Louis.
- Cutler D, McClellan M, Newhouse J. 2000. How does managed care do it? *RAND Journal of Economics* **31**(3): 526–548.
- Cutler D, Reber S. 1996. Paying for health insurance: the tradeoff between competition and adverse selection. *Quarterly Journal of Economics* **113**(2): 433–466.
- Feldstein M. 1973. The welfare loss of excess health insurance. *Journal of Political Economy* **81**(2): 251–280.
- Gaynor M, Vogt W. 2003. Competition among hospitals. *RAND Journal of Economics* **34**(4): 764–785.

- Hanushek E. 1974. Efficient estimators for regressing regression coefficients. *American Statistician* **28**(1): 66–67.
- Hayashi F. 2000. *Econometrics*. Princeton University Press: Princeton, NJ.
- Hicks J. 1939. *Value and Capital*. Clarendon Press: Oxford.
- Ho K. 2005. Insurer–provider networks in the medical care market. NBER Working Paper 11822.
- Lewis J. 2000. Estimating regression models in which the dependent variable is based on estimates with application to testing key’s racial threat hypothesis. Mimeo, Princeton University.
- McFadden D. 1973. Conditional logit analysis of qualitative choice behavior. In *Frontiers in Econometrics*, Zarembka P (ed.). Academic Press: New York; 105–142.
- McFadden D. 1981. Econometric models of probabilistic choice. In *Structural Analysis of Discrete Data*, Manski C, McFadden D (eds). MIT Press: Cambridge, MA; 198–272.
- Miller R, Luft H. 1997. Does managed care lead to better or worse quality of care? *Health Affairs* **16**(5): 7–25.
- National Center for Health Statistics. 2004. *Health, United States, 2004*. US Department of Health and Human Services, Centers for Disease Control and Prevention.
- Newhouse J. 2002. *Pricing the Priceless: A Health Care Conundrum*. MIT Press: Cambridge, MA.
- Nevo A. 2000. Mergers with differentiated products: the case of the ready-to-eat cereal industry. *RAND Journal of Economics* **31**(3): 395–421.
- Nevo A. 2001a. Measuring market power in the ready-to-eat cereal industry. *Econometrica* **69**(2): 307–342.
- Nevo A. 2001b. New products, quality changes and welfare measures computed from estimated demand systems. NBER Working Paper 8425.
- Ohio Chamber of Commerce Task Force on Provider Access. 2000. Final Report, April.
- Pakes A. 1997. Lecture notes for advanced applied econometrics. Mimeo, Harvard University.
- Pauly M. 1986. Taxation, health insurance, and market failure in medical care. *Journal of Economic Literature* **24**(2): 629–675.
- Petrin A. 2001. Quantifying the benefits of new products: the case of the Minivan. NBER Working Paper 8227.
- Royalty A, Solomon N. 1998. Health plan choice: price elasticities in a managed competition setting. *Journal of Human Resources* **34**(1): 1–41.
- Strombom B, Buchmueller T, Feldstein P. 2002. Switching costs, price sensitivity and health plan choice. *Journal of Health Economics* **21**(1): 89–116.
- Tay A. 2003. Assessing competition in hospital care markets: the importance of accounting for quality differentiation. *RAND Journal of Economics* **34**(4): 786–814.
- Town R, Vistnes G. 2001. Hospital competition in HMO networks. *Journal of Health Economics* **20**: 733–753.
- White H. 1980. A heteroscedasticity-consistent covariance matrix estimator and a direct test for the heteroscedasticity. *Econometrica* **48**(4): 817–838.