Hospital Choices, Hospital Prices and Financial Incentives to Physicians*

Kate Ho and Ariel Pakes

August 2013

Abstract

We estimate a preference function which rationalizes hospital referrals for privately-insured birth episodes in California. The function varies across insurers and is additively separable in: a hospital price paid by the insurer, the distance traveled, and plan and severity-specific hospital fixed effects (capturing various dimensions of hospital quality). We use an inequality estimator that allows for errors in price and detailed hospital-severity interactions and obtain markedly different results than those from a logit. The inequality estimator indicates that insurers with more capitated physicians are more responsive to hospital prices. Capitated plans are willing to send patients further to utilize similar-quality lower-priced hospitals; but the trade-off between quality and costs does not vary with capitation rates.

1 Introduction

The Patient Protection and Affordable Care Act of 2010 includes provisions to establish Accountable Care Organizations (ACOs) in the Medicare program. These are groups of health care providers that offer services to a large number of patients and are eligible to share in any cost savings they achieve for the Medicare program. Private sector ACOs are forming in parallel to this initiative, often with very similar structures and payment arrangements. The proportion of savings ACOs keep is linked to their performance on quality standards. The goal of the initiative is to reduce health care costs in both Medicare and private insurance and to improve coordination of care. The tying of cost savings to quality standards is designed to mitigate the fear that an increase in incentives to control costs might lead to a decrease in the quality of care.

The health policy literature has noted that similar cost control incentives are currently utilized by some health maintenance organizations (HMOs) in California and elsewhere.[1] In this paper we investigate the impact of these incentives on the cost and quality of care provided to patients, focusing in particular on the choice of hospital for privately insured patients giving birth. We use

*We thank Mark Shepard, Lucia Tian Tian and Zach Brown for excellent research assistance. Thanks to participants at numerous seminars and to four helpful referees for their comments and suggestions. All remaining errors are our own.

[1] See, for example, Hammelman et al (2009).

hospital discharge data for managed care enrollees from California in 2003 to estimate a hospital referral model which posits that referrals are a plan specific function of the prices insurers pay to the hospitals, distance from home to hospital, and a severity-specific measure of "hospital quality". The results are used first to study the relationship between referrals to hospitals, hospital prices, and differences in the cost-control incentives used by different insurers. We then refine our estimates in a way that allows us to analyze how the implicit trade-offs between price, our severity-specific measure of hospital quality, and distance differ with these incentives.

The question of whether patients enrolled in insurers that give physicians an incentive to control hospital costs are admitted to low-priced hospitals is important for several reasons. Hospital costs make up more than 30% of national health care spending so our findings are potentially important for the overall impact of ACOs on costs. Moreover any analysis of the effect of hospital mergers on hospital prices will require an assumption regarding the effect of price increases on referrals. Assuming away this effect will likely result in an over-estimate of the price increases that will result from a merger. Similarly a merger analysis that does not take into account price effects will likely over-estimate the merged entity's incentives to invest in new high-cost technologies and under-estimate its incentives to invest in cost-reducing technologies.

The process by which a patient chooses a hospital involves multiple players. Decisions are made by referring physicians in consultation with their patients. Insurers often attempt to influence physician choices through direct financial incentives. In California in particular, they often remunerate the physician group through fixed (capitated) monthly payments per patient to cover the cost of patient services. The most common type of capitation involves payments covering only services provided by the physician group but when this is the case "shared risk arrangements" almost always apply, under which a target is set for total (including hospital) spending and cost savings or overruns relative to the target are shared between the physician group and the ACO. Less commonly, global capitation payments cover the cost of all services received by patients, including inpatient hospital stays. Both types of capitation contracts generate incentives for physician groups to refer their patients to low-cost hospitals. Physician groups often pass the financial incentives directly to their member physicians; they may also make physicians' promotion on the pay scale contingent (formally or informally) on their management of costs. Thus, if we assume that higher quality hospitals negotiate higher prices, physicians in groups receiving capitation payments face a trade-off between incentives to reduce costs and other issues patients care about such as quality of care and convenience. Moreover the extent to which patients and/or doctors care about these factors is likely to vary with the severity of illness.

The incentives for ACOs and for the California insurers in our data are similar in that the provider group either bears the financial risk for hospital payments or benefits from hospital savings relative to a benchmark. Also in both cases the incentives are based on the costs incurred by the group (rather than by individual physicians), with no formal guidance on how these incentives are passed down to individual physicians or patients. Approximately 430 ACOs had been set up by January 2013 (Muhlestein (2013)). The details of their payment arrangements are evolving over

time, but they all involve some form of shared savings when costs are less than a benchmark[2]. Their structure also varies: in 2013, 189 were integrated with a hospital system, while most of the remainder were sponsored by physician groups which contracted with hospitals outside the organization. We come back to some of the likely implications of our results for these distinctions below.

Our analysis uses hospital discharge data to estimate an insurer-specific preference function which rationalizes hospital referrals for women giving birth in California. Unfortunately our dataset does not identify the physician referring each patient to her hospital; we therefore cannot directly observe physician behavior. However we do observe each patient's insurance carrier and the extent to which each carrier uses capitation payments. In 2003 73% of payments made to primary physicians by the six largest carriers in our data were capitation payments; the proportions varied substantially across carriers from 97% for Pacificare to 38% for Blue Cross (more detail on the data underlying these numbers is given below). We ask whether the observed referrals for patients whose insurers have different capitation rates indicate different trade-offs between price, quality and patient convenience factors. We view the allocation of a patient to a hospital as the outcome of a multi-stage process that includes, for example, the patient's choice of obstetrician and the obstetrician's allocation to affiliated hospitals as well as the obstetrician's choice of hospital for the particular patient. We do not model the protocol that leads to these choices; rather we estimate the relative weights on different factors that emanate from that protocol and the extent to which these weights are related to capitation rates.

The analysis builds on the previous literature on hospital demand. Previous papers consider the factors affecting patients' hospital choices in some detail but almost exclusively make the simplifying assumption that the hospital is chosen without regard for the price paid by the insurer. To include the price variable one has to address three problems. First, the observed price is a "list price" for the relevant hospital discharge. Prices actually paid by the insurer are list prices multiplied by a proprietary discount. To address this issue we import data on the average hospital discount from hospital financial reports. The true discount could vary across insurers and we treat that in two complementary ways: we allow for errors in our price variable and present results which use additional data to estimate the variation in discounts across insurers. Second the expected price that generates hospital choices is inherently unobservable. We assume that expectations are on average correct, and construct a price variable which is the average realized price for patients admitted with the same diagnosis and similar co-morbidities at that hospital. Those predictions will have estimation error, but the estimation procedure we develop averages those errors out.[3] The third problem relates to price endogeneity: the expected price for a patient with a particular diagnosis is likely to be correlated with the unobserved hospital quality for that diagnosis. We

[2] See the Department of Health and Human Services (November 2012). Initial arrangements for Medicare ACOs were largely based on shared savings with no requirement to share in any losses relative to the benchmark, but as time passes they increasingly involve physician groups bearing at least some financial risk.

[3] Our baseline results assume no measurement error in distance. Later we consider a specification which is robust to errors generated by the distance between the centroids of our geographic areas and true locations, and find no major difference in results.

control for these unobservables by developing an estimation procedure which allows for hospital fixed effects that vary freely with severity of diagnosis.

We begin with a standard logit model for hospital choice. This does not allow for errors in the price variable and has a limited ability to allow for hospital fixed effects that vary with severity of diagnosis. We expect the omission of hospital severity interactions to bias the price coefficient upwards, and measurement error in the price measure to bias the price coefficient towards zero. When we pool all delivery and birth discharges we obtain a positive and significant coefficient on price. However, when we narrow the sample to the least risky pregnancies (a more homogenous diagnosis group) the price coefficient becomes negative. We then allow the price coefficient to vary by insurance carrier and find that the carriers with the highest proportion of payments to physicians made through capitation contracts have negative significant price coefficients while other carriers with a higher proportion of fee-for-service contracts have insignificant coefficients on price. Since neither endogeneity nor errors in variables are fully addressed by this technique we doubt that the estimates obtained here accurately measure the true responses to price.

So we develop a methodology that addresses these problems. It is based on revealed preference: we assume that the hospital chosen for each patient is preferred to any of the other hospitals in her choice set. The preference function is assumed to be additively separable in the price paid by the insurer, distance, and an insurer and severity specific hospital quality. We identify pairs of patients who have the same severity and are enrollees in the same insurer but who chose different hospitals. By defining the alternative of each patient as the chosen hospital of the other and summing the two patients' inequalities, we difference out the severity-specific hospital quality terms from the utility equation. By averaging the resulting inequalities over patients and hospitals we eliminate the effects of errors in price measurement. The result is a relatively straightforward estimator of bounds on the (normalized) price coefficient.[4]

The estimates indicate that the price coefficients are far more negative than in the logit analysis and are ordered with respect to the plans' capitation rates. That is, the price paid by insurers to hospitals does impact referrals, and the price response is more elastic for insurers whose physician groups are more highly capitated. We show that these results are robust to a number of perturbations to the specification used in the estimation. We then use the price coefficients to back out bounds on the plan, hospital, and severity specific quality terms and find them to be highly correlated across plans. We therefore add structure and estimate a model where the quality terms for the different plans are affine transforms of one another. This allows us to represent preferences as a linear function of price, quality, and distance which differs across plans only in the coefficients of these variables. We can then examine how the trade-offs between price, quality and distance vary with capitation rates. Though in absolute value the price coefficient varies directly with the

[4] The analysis is similar in spirit to previous papers that match treatment to control groups based on observable data and assume that unobserved information does not affect response to treatment. The propensity score literature, and difference-in-differences analyses more generally, fall into this category. See Rosenbaum and Ruben (1983) for propensity score estimators and, for example, Card and Krueger (1994) for difference-in-difference estimators. The framework we use for the analysis of the inequalities is described in more detail in Pakes, Porter, Ho and Ishii (2011) and Pakes (2010).

capitation rate so does the quality coefficient. Consequently the ratio, or trade-off, between price and quality is evaluated in much the same way across plans. In contrast the trade-off between distance and price is evaluated differently. That is highly capitated more price sensitive plans tend to send their patients longer distances to obtain the same quality service at a lower price (but do not trade-off costs against quality differently). To the extent that "hospital quality" is related to health outcomes, we can examine the relationship of quality to capitation rates directly by importing data on measures of health outcomes for the mothers and infants into our data. Conditional on patient severity, there is no evidence that capitation rates are related to any of our health outcome measures.

The remainder of the paper is structured as follows. In Section 2 we discuss the relevant previous literature. Section 3 describes important features of the market and Section 4 describes the data. Section 5 sets out the full model we wish to analyze, Sections 6 and 7 summarize the restrictions required for the logit and inequalities methods and set out their results, Section 8 considers the trade-off between price, quality and distance, and Section 9 concludes.

2 Previous Literature

Two sets of previous papers are relevant for our analysis. The first, summarized by Glied (2000), considers HMO gatekeeping and cost controls. Glied's summary suggests that HMOs have lower inpatient admissions and costs than other insurers; however these papers do not analyze the relationship between hospital price and referrals.[5] There are a few more recent studies that consider similar questions. For example, Cutler et al (2000) compare the treatment of heart disease in HMOs and traditional insurance plans and find that HMOs have 30% to 40% lower expenditures. Virtually all the difference comes from lower unit prices rather than differences in actual treatments. However they do not investigate whether price reductions are due to lower negotiated prices within a hospital or to referring patients to cheaper hospitals (the focus of our study). Gaynor, Rebitzer and Taylor (2004) look in more detail at how HMOs achieve cost savings. They analyze physician responses to group-based financial incentive contracts within a single HMO. They find that spending on medical utilization, particularly for outpatient services, increases with the size of the physician group receiving group-based incentives. That is, spending is negatively correlated with the power of incentives to limit these expenditures.[6]

There are also some papers evaluating recent initiatives that implement cost-control incentives like those that may be used for Accountable Care Organizations. For example the Alternative Quality Contract (AQC) was adopted by Blue Cross Blue Shield of Massachusetts in 2009. It introduced physician incentives similar to those created by the global capitation contracts in our data. Physician groups entered into five year global budget contracts under which they received a budget per enrolled patient and were accountable for costs of all services provided to those patients,

[5] More recent reviews by Chandra, Cutler and Song (2012) and McClellan (2011) come to similar conclusions.

[6] Other recent papers considering the responsiveness of health care providers to financial incentives include Ketcham, Leger and Lucarelli (2012), Limbrock (2011) and Bajari, Hong, Park and Town (2012).

including inpatient care. Song, Zafran et al (2011) use a difference-in-difference analysis (where the first difference is across time and the second is between intervention and control groups) to analyze the impact of this initiative. They find that, in the first year, it was associated with reduced growth in spending on outpatient services and improved quality of care and that most of the savings came from referring patients to lower-priced providers. We also conducted a preliminary analysis of the response of the hospital referrals of physicians to capitation payments in Ho and Pakes (2011). There we regressed a severity-adjusted price measure on the proportion of the insurer's payments to primary physicians that were capitated and market fixed effects[7]. The coefficient on capitation was negative and statistically significant; consistent with the hypothesis that insurer capitation payments influence physician referrals. However, simple regressions like these cannot provide more than suggestive evidence since they do not account for the trade-offs made between price, other hospital characteristics, and convenience factors in the hospital choice equation.[8]

The second relevant literature estimates discrete choice models of hospital demand: see Gaynor and Vogt (2000) for a survey.[9] Almost all of these papers exclude the price paid by the insurer to the hospital from the utility equation. One exception is Gaynor and Vogt (2003) which uses assumptions to define a single price index for each hospital that is included in the utility equation. However, that paper assumes away interactions between patient and hospital characteristics in determining procedures and therefore prices. It also does not consider the impact of physician incentives on the price coefficient.

3 Background on the Market

The analysis in this paper focuses on enrollees of health maintenance organizations (HMOs).[10] The seven largest HMOs had 87% of the California HMO market at the end of 2002. Our analysis focuses on six of these seven: we exclude Kaiser Permanente because the prices paid by this vertically integrated insurer to its hospitals are not observed in our data.[11]

Each HMO contracts with a network of providers (physicians and hospitals); enrollees seek care within that network. Each pregnant woman chooses an obstetrician from within the network and is referred to one of the small number of network hospitals with which the obstetrician is affiliated. Our analysis controls for any variation in networks across insurers by conditioning on the observed provider network when estimating the choice equation. While HMOs could, in theory, influence

[7] We adjusted for severity by constructing the following price ratio measure: $p_i^{ratio} = \frac{p_i}{\bar{p}_{s_i}}$ where p_i was the hospital price for patient i and $\bar{p}_{s_i}$ was the average of that variable for same-severity patients across all hospitals in the sample.

[8] Duggan (2000) considers hospital referrals for Medicaid patients. He finds that private hospitals in California responded to new financial incentives to treat Medicaid patients, by cream-skimming the most profitable Medicaid patients from publicly-owned hospitals. The reallocation was especially pronounced for pregnant women.

[9] Examples include Luft et al (1990), Burns and Wholey (1992), Town and Vistnes (2001), Capps, Dranove and Satterthwaite (2003)and Ho (2006), all of which either omit price entirely or include only the list price.

[10] The 2003 California medical care market is described in detail in Baumgarten (2004). Several previous papers describe the contractual arrangements between health plans and physicians in California, including Rosenthal et al (2001) and Grumbach et al (1998a. and b.).

[11] The insurers we do consider are Blue Cross, Blue Shield, Health Net, Pacificare, Aetna and Cigna. Blue Cross of California is independent of other Blue plans, including Blue Shield of California.

hospital referrals for their enrollees by either defining narrow hospital networks, or requiring patient cost sharing, in practice this is not usually the case.[12] There are two exceptions. One is the use of out-of-pocket payment schemes that require the consumer to share in hospital costs. Most out-of-pocket payments are a fee (or co-pay) which is fixed across hospitals and much smaller than hospital costs. However a minority of patients pay a co-insurance rate; i.e. a fixed percentage of hospital costs. Our data does not contain information on co-insurance rates, but we know from industry reports that they were used by only a small fraction of enrollees in 2003, and to the extent they were used, they were used disproportionately by the low capitation insurers in our data set, and therefore do not explain our findings on the differences in price sensitivities across insurers[13]. The second exception is the possible existence of tiered networks. These are arrangements where insurers group hospitals into categories based on costs and patients pay more for services from higher-cost hospitals. They were first introduced in California in 2002-3. We do not observe the details of any tiered networks and so cannot condition on them. However the use of these products in 2003 was not correlated with capitation so it does not explain our findings.[14]

HMOs in California do not generally use hospital payment mechanisms that provide incentives either to control costs or improve quality. Most hospitals in California are paid by the insurance carrier on a per service or per diem basis.[15] Payment arrangements for physicians, in contrast, are often structured to generate cost-control incentives. Most HMOs contract on a non-exclusive basis with large physician groups,[16] making capitated (fixed) monthly payments to the group for every enrollee who uses it as his or her primary care clinic. The most common alternative is a fee-for-service payment arrangement. The extent of financial risk passed to the medical group varies across capitated contracts. In around 20% of cases the monthly ("global capitation") payment covers all services needed by the physician group's patients including inpatient hospital stays. These physician groups have a clear incentive to refer their patients to lower-cost hospitals. The remaining 80% of capitation contracts involve payments that cover only the cost of services provided by physicians within the group, perhaps with the addition of ancillary services like outpatient medical tests. The HMO makes separate payments to hospitals for providing secondary care. Physician groups again

[12] Ho (2006) finds that on average 83% of hospitals were included in each HMO's network in a sample of 43 large markets (including seven in California) in 2003. Capps, Dranove and Satterthwaite (2003) report similar evidence.

[13] The Kaiser Family Foundation Employer Health Benefits Survey 2003 reports that 5% of covered workers in an HMO paid a co-insurance rate for hospital admissions while 49% paid a deductible or copay; the remaining enrollees paid neither. In contrast 14% of PPO enrollees paid a co-insurance rate, 26% paid a deductible or copay and the remainder paid neither. Our data contains only HMO enrollees for four out of six insurers. However we note below that for Blue Shield and Blue Cross only we observe admissions from PPO as well as HMO plans, since both are included in the Knox Keene plan for these two insurers. Blue Shield (a not-for-profit) and Blue Cross have the lowest capitation rates in our data. This use of patient cost-sharing is therefore negatively correlated with capitation, so it is likely to bias our estimates towards finding no difference in price coefficients between high- and low-capitation insurers. See Section 4 and Appendix 1 for details.

[14] Pacificare, Health Net and Blue Shield were the earliest adopters (in 2002) and only Blue Shield made the tiered product mandatory for employers. See Yegian (2003) and Robinson (2003) for details. We show below that Blue Shield is a low capitation insurer. Moreover it is the only not-for-profit insurer in our data set, and after reporting our estimates of its price coefficient, we omit it from the rest of the analysis.

[15] Hospital capitation payments existed but were uncommon in 2003. See Appendix 1 for more details.

[16] There are two types of physician groups: medical groups and Independent Practice Associations (IPAs). See Appendix 1 for details.

have incentives to control hospital costs because "shared risk arrangements" almost always apply, under which a spending or utilization target is set and cost savings or overruns relative to the target are shared between the physician group and the HMO.[17] These arrangements are very similar to the "shared savings" arrangements that are often instituted for Accountable Care Organizations.

Our dataset does not distinguish between global and non-global capitation arrangements. We investigate the extent to which referrals from physician groups with more highly capitated contracts of any kind are more (or less) sensitive to price. We also do not observe the physician or physician group referring each patient to a hospital. We do see the name of each patient's insurer and the percent of each insurer's primary services and other medical professional services that are capitated. In the analysis below we compare the importance of price in determining the hospital choice for patients enrolled in high-capitation insurers to its importance for those in low-capitation insurers.

Our finding that referrals from high capitation insurers tend to be to less costly hospitals could be due to individual physicians referring capitated patients to cheaper hospitals and non-capitated patients to others (which is consistent with Melichar, 2009), or to more cost conscious physicians being associated with physician groups that are more highly capitated. Physicians in more highly capitated physician groups could be more cost conscious either because they respond to incentives they face in those groups, or because inherently cost conscious doctors gravitate towards high capitation groups. Our data does not allow us to distinguish between these alternative mechanisms.

4 The Dataset

4.1 Data Sources and the Price Measure

We use five datasets. The first is hospital discharge data covering all patient discharges from hospitals in California in the year 2003 from the state's Office of Statewide Planning and Development (OSHPD). This provides information on each patient's zip code, demographic characteristics, health insurer, the hospital chosen and patient diagnosis details: both the "principal" diagnosis recorded as the major cause of admission and a list of up to 24 other diagnoses for each patient. We link this to OSHPD hospital financial data, to the OSHPD Birth Cohort file which contains outcome variables for both mother and infant, and to hospital characteristics data from the American Hospital Association for 2003. Finally we have access to the State of California Department of Managed Health Care Annual Financial Reporting Forms for 2003. These include balance sheets, income statements and some information on enrollment, utilization and types of payment to providers for all Knox Keene plans (essentially the same as HMOs) in California. We consider only birth and delivery-related admissions records and only private Knox Keene enrollees.[18] Our analysis covers

[17] Rosenthal et al (2001) note that 85-90% of non-global capitation revenues were generated from contracts with shared hospital risk. Robinson and Casalino (2001) report similar findings. According to this literature, fee-for-service contracts at the time of our data did not generally involve shared hospital risk arrangements.

[18] Knox Keene plans are defined as plans that are overseen by the California Department of Managed Health Care (DMHC) and subject to the Knox Keene Act. They are not precisely the same as HMOs: while most insurers designate only their HMO plans as Knox Keene plans, Blue Shield and Blue Cross PPO products were also included in this category in 2003. We cannot distinguish between PPO and HMO enrollees for these two insurers at the

only the six largest insurers other than Kaiser Permanente: these make up approximately 90% of the non-Kaiser observations in the data. We infer the hospital network of each insurer using the discharge data by assuming that a hospital is in the network if at least 3 patients are admitted from the particular insurer.[19] Consistent with Kessler and McClellan (2000), we assume that patients consider traveling up to 35 miles to visit a general hospital and up to 100 miles to visit a teaching hospital.[20]

We do not observe the price charged to the insurer by the hospital, and, as a result, have to construct our price variable. Our data does include the list price for every discharge. List prices are a standard set of prices listed by hospitals in each year for all their services. All patients are quoted the same list price for the same service. However, only uninsured patients and some patients using an out-of-network provider are actually asked to pay the list price, and even they are frequently offered a discount by the hospital. Each insurance company has a contract with each provider in its network that defines a discount from the list price for its enrollees. We also observe the average negotiated discount at the hospital level, calculated as the total contractual adjustments from private managed care payors divided by the total charges (the sum of list prices for all inpatient and outpatient episodes) for the relevant hospital-year. Both variables are recorded in the hospital's financial statements.[21]

The price we need to construct is the price that the decision-maker expects to pay for a patient entering the hospital in a given condition. It is defined below by a combination of diagnosis, age, and comorbidity information known at the time the patient is admitted to the hospital. We assume that expected prices are on average correct, and construct a baseline price variable as the average realized list price for a given condition in a particular hospital multiplied by 1 minus the average hospital discount. Since the estimation methodology we rely on for our conclusion averages over agents, any remaining expectational error should average out when we sum across severities.[22]

individual discharge level. Capitation rates are also reported for the full Knox Keene plan. This likely generates some of the cross-insurer variation in capitation rates in the data. To account for the differences between HMO and PPO plan types we drop hospitals to which very few patients are admitted for these two insurers (thereby removing from the data the hospitals that are most likely to be out-of-network). To the extent that we have not dropped all those PPO patients who pay a coinsurance rate rather than a fixed copay, we bias our estimates against finding a positive relationship between price-sensitivity and capitation rates. This because Blue Shield and Blue Cross are the lowest capitation insurers in our dataset. See discussion in Appendix 1 for details.

[19] We check the implied network definitions against hand-collected data (described in detail in Ho (2006)) from seven California markets in 2003. The definition is conservative: that is, the networks implied by our methodology contain fewer hospitals than the networks in the hand-collected data and if an implied network contains a particular hospital it is also included in the hand-collected data in the vast majority of cases.

[20] We repeated the analysis using a 20 mile radius for low-risk pregnancies in Los Angeles, since a 35 mile radius could be too large in this urban area. The results were very similar to those reported below. Still we note that our market definitions may mask some remaining variation in the data.

[21] The fee-for-service arrangement is not the only way hospitals are reimbursed. Per diem and DRG or case-based payments are also possible: see Appendix 1 for details. In principle we have the right average discount since contractual adjustments are defined as the sum of charges for the hospital's private managed care patients less its total revenue for those patients. However since the discount applies to all such patients and we only analyze Knox Keene inpatient labor and delivery admissions, there may be some error in it. We average over patients in different hospitals, which under assumptions to be specified later addresses the measurement error issue.

[22] We take averages over patients who enter the hospital with a given severity level. Our definitions of severities differ across our model specifications and are detailed below. Gaynor and Vogt (2003) use a similar methodology,

We demonstrate below that there is meaningful variation in this price measure both across patients admitted with different conditions and across hospitals for a given condition. However, it is clearly subject to measurement problems. There is a trade-off between aggregation error, if our groups of similar patients for the expected list price calculation are defined too broadly, and measurement error if they are too narrow implying small sample problems. We return to this issue below. There may also be specification error since we observe the discount at the hospital rather than the hospital-insurer level.[23] We examine the robustness of our results to specification error in the price variable in Section 7 and Appendix 2. There we use additional data on the share of each hospital's total inpatient revenues coming from each insurer to estimate a model of the discount as a function of hospital, insurer and market characteristics. We then repeat our inequalities analysis using price measures derived from the estimated discount, and find only minor differences in our results.

4.2 Summary Statistics

Table 1 sets out summary data on the six insurers included in the analysis; data for Kaiser are also included for comparison. These data give a broader picture of the insurers we consider than can be provided by our specific dataset. Since the effect of capitation payments on the price coefficient will be identified from variation across these six insurers, our goal here is to summarize the differences between them on other relevant dimensions. The first three columns provide enrollment data, showing that of the insurers we consider, Blue Cross and Blue Shield have the largest commercial plan enrollment while Aetna and Cigna have the smallest. Every insurer in our dataset has over 70% of its enrollment in commercial plans. The fourth column lists the number of delivery discharges included in our analysis for each plan; the breakdown is approximately proportionate to the commercial enrollment numbers. The next column lists the percent of each Knox Keene insurer's primary services that are capitated.[24] There is considerable dispersion across insurers, from Pacificare with 97% capitated payments to Blue Cross with 38%. The rest of the table demonstrates that insurers with a high percent of capitated payments are not obviously different from other insurers on observable dimensions. In particular, premiums per member per month, inpatient utilization, and prescription drug costs are not highly correlated with capitation rates. That is to the extent that these variables reflect which types of patients choose which insurers, this suggests that patient selection into insurers is not related to capitation rates.

defining price as the observed list price multiplied by 1 minus the average discount.

[23] Specification error may also be generated because the observed discount is an average for both inpatient and outpatient services and for all managed care payors (including Point of Service plans) rather than just for Knox Keene inpatient events. An examination of the robustness of our results to this source of error is given in Section 7 and Appendix 2.

[24] Capitation payments for primary professional services are defined in the Knox Keene insurers' Annual Financial Statements as "capitation costs incurred by the reporting entity to primary care physicians, dentists and other professionals for the delivery of medical services". They include capitation payments to obstetricians. The statements also record capitation payments for other medical professional services, defined as including, for example, services provided by optometrists, nurses, ambulance drivers and technicians.

Blue Shield and Blue Cross, which have the lowest proportion of capitated payments, were historically different from other insurers. They were 501(c)(4) tax exempt as social welfare plans, acting as administrators of Medicare and providing coverage to state and federal government employees. By 2003, however, Blue Cross and Blue Shield companies were franchisees, independent of the association and each other. They were no longer tax exempt and could be for-profit corporations. In California Blue Cross was an investor-owned for-profit organization with a lower medical loss ratio (defined as medical and hospital expenses divided by premium revenues for the whole insurer) and similar inpatient utilization to other insurers in the market. Blue Shield was still quite different from the other insurers we consider. It was the only not-for-profit company we analyzed and had relatively high inpatient utilization figures (although its premiums and medical loss ratio were quite low). As a result Blue Shield's administrators and physicians may have been less receptive to financial incentives than those of other insurers, an issue we return to below.

Table 2 provides summary statistics on the discharges in the dataset. There are 88,157 patients and 195 hospitals.[25] There are 38 hospitals in the average patient's choice set. 27% of discharges are from teaching hospitals. The average price paid (approximated as list price*(1-average discount)) is $4,317 for birth admissions. The average length of stay is 2.5 days. The importance of the distance between the patient's home and her hospital is clear from the raw data. The average distance between a patient and a hospital in her choice set is 24.6 miles; the average distance to the chosen hospital is 6.7 miles. Distance will be an important variable in the utility equation estimated below.

The table also records means for three potential measures of outcomes: death while in hospital, transfer to an acute care setting (at this hospital or a different hospital) and transfer to a skilled nursing facility (again at either this or a different hospital). These are useful inputs to an initial investigation of the patterns in the data although we will not use them in our full model. The average probability of each event is low for delivery admissions: 0.01% for death, 0.3% for acute care transfer and 1.5% for transfer to a skilled nursing facility.

Table 3 demonstrates that the variation in price and in outcomes across patient ages and comorbidities is intuitive. Here we add both data on infant outcome variables and data that follows both mother and baby over time which enables us to calculate the probability of readmission within a 12 month period[26]. We also aggregate the probabilities of death, acute care transfer and skilled nursing facility transfer into a single probability of discharge to a location other than home.

Comparing older to younger women who give birth, we find that the birthing episodes of women who are aged over 40 are significantly more expensive and that these older women have higher

[25] This is the sample used for the logit analysis. We follow the previous literature by accounting for all hospitals in the choice set for that analysis. Average discount data is missing for some hospitals; we fill it in using regression analysis. We exclude these hospitals from the inequalities analysis because only pairwise comparisons between hospitals on which we have complete price information are required for the inequality estimation procedure. For similar reasons we also exclude the small number of hospitals reporting more than 5% capitated revenues. We are left with 64,691 patients and 157 hospitals. Using this smaller sample for the logits has no qualitative effect on the estimates.

[26] The data are taken from the OSHPD Birth Cohort File for 2003. All summary statistics are very similar to those of our main dataset.

probabilities of readmission within 12 months and of discharge to "other than home" than younger women. Also the infants of the older woman incur significantly larger hospital expenses and are significantly more likely to be discharged to "other than home". Interestingly infant readmission probabilities are not significantly different across these two groups[27].

We use the Charlson score (Charlson et al, 1987) as a measure of patient severity: this assigns integer-valued weights (from 0 to 6) to comorbidities other than principal diagnosis where higher weights indicate higher severity. The weights are summed to generate a single integer-valued index. For example, patients with comorbidities indicating that they have diabetes or mild liver disease would receive a Charlson score of 1; those with renal disease or any malignancy would have a Charlson score of 2; those with a metastatic solid tumor or AIDS would have a Charlson score of 6. A patient with both diabetes and renal disease would have a score of 3. The index was developed by physicians and is widely used to measure severity based on diagnoses listed in patient records. Table 3 indicates that women with higher Charlson scores in our data, and their infants, had more costly deliveries and higher probabilities of adverse outcomes than women with lower Charlson scores. All of these differences are significant at p=0.05. Our main analysis will allow the Charlson score, interacted with other severity measures such as age and principal diagnosis, to affect preferences for different hospitals.

5 The Model

As noted the hospital chosen is a result of a complex decision process. The woman first chooses an obstetrician, typically with knowledge of which hospitals the obstetrician can admit patients to, and then the obstetrician, in consultation with the patient, chooses among the hospitals where he (or she) has admitting privileges. We assume this process generates an ordering of the hospitals in the insurer's network which is derived from the patient's and doctor's preferences. The patient's preferences are affected by the distance from her home to the hospital, her assessment of the severity of her condition, and by the hospital's characteristics (some of which we, the researchers, may not have measures of). The physician's choice is influenced by the patient's preferences, their assessment of the severity of the patient's condition and the quality of the hospital services for that severity, and the price charged by the hospital to the insurer. The potentially observable part of the preference function whose maximum determines the hospital (h) that patient i of insurer π is allocated to, is assumed to take the additively separable form

$$W_{i,\pi,h} = \theta_{p,\pi} p(c_i, h, \pi) + g_\pi(q_h(s), s_i) + f_\pi(d(l_i, l_h)) \tag{1}$$

[27] Our price sensitivity analysis includes only the expenses charged to the mother. Infant health, and therefore expected expenses, are generally not known when the hospital is chosen and therefore are not relevant for this part of the analysis. However hospital quality for the infant as well as the mother should be an input to the hospital choice. So we consider both infant and maternal outcomes in Section 8 where we ask whether health outcomes differ by insurer (conditional on severity).

where

- $p(c_i, h, \pi)$ is the price insurer π is expected to pay at hospital h for a patient who enters in condition (which includes diagnosis) c_i,
- s_i is a measure of the severity of the patient's condition,
- $q_h(s)$ is a vector of perceived qualities of hospital h, one for each different severity,
- $g_\pi(q_h(s), s_i)$ is the plan and severity specific function which determines the impact of the quality of a hospital for the given severity on the choice of hospitals, and
- l_i is patient i's location, l_h is hospital h's location, $d(\cdot)$ provides the distance between the two locations, and $f_\pi(\cdot)$ is an increasing function of that distance which may differ by plan.

for $i = 1, \ldots, n_{\pi,s}$, $h = 1, \ldots, H$, and $\pi = 1, \ldots, \Pi$. The condition (c_i) of the patient on admission will sometimes be referred to as the diagnosis and the severity (s_i) groups are aggregates of the admission conditions (so there is variance in price conditional on severity). The precise definitions of both these terms differ somewhat across estimation strategies and will be defined below. Also when implementing these alternative estimation strategies we will add an unobserved (or disturbance) term whose properties will be assumed to differ across those strategies.

In its most general form, which will be our preferred form, the function $g_\pi(\cdot)$ is allowed to differ arbitrarily: across plans, among sickness levels for a given hospital, and across hospitals. Also particular hospitals can have higher quality for some diagnoses and/or sickness levels than for others. So different aspects of a hospital's characteristics can be differentially important for hospital choice for different severities[28]. It also permits physicians to differ in their intensity of preferences for quality relative to price and distance when considering patients of different sickness levels, and allows different insurers to differ in both their quality assessments of hospitals and the weight they give to quality relative to price and distance. Finally keep in mind that the ordering implicit in equation (1) is an outcome of a decision making protocol that involves both the physicians' and the patients' preferences, so the analysis allows for patient (as well as physician) differerences across plans.

For some of the specifications we will have to constrain $g_\pi(\cdot)$ to be a parametric function of patient and hospital characteristics. To the extent that the parametric assumption does not capture all the variance in $g_\pi(\cdot)$ the residual variance will create an additional unobservable that may bias the other parameters of interest. In particular if the "unobserved quality" represented by this residual is correlated with price we would expect it to cause a positive bias in the price coefficient. This is the reason that our preferred mode of analysis starts by not constraining $g_\pi(\cdot)$ in any way. Only after estimating the price coefficients for each plan do we come back to asking whether there are constraints that the hospital-severity-plan specific quality estimates seem to satisfy.

[28] For example, Goldman and Romley (2008) find evidence that amenities such as food quality and staff attentiveness play an important role in hospital demand, and their relative importance in the hospital choice equation may differ by severity of illness.

6 Logit Analysis

We begin with a multinomial logit model of hospital choice, as it provides a familiar starting point for investigating the patterns in the data. The logit model makes the following assumptions.

$$p(c_i, h, \pi) = \delta_h^o lp^o(c_i, h) \tag{2}$$

where $lp^o(c_i, h)$ is the average list price of patients who enter hospital h with condition c_i and δ_h^o is one minus the average discount rate hospital h gives managed care providers,

$$g_\pi(q_h(s), s_i) = q_h + \beta z_h x(s_i) \tag{3}$$

where q_h are hospital fixed effects, $x(s_i)$ are functions of the sickness level of the patient and z_h are hospital characteristics, both of which are specified below, and

$$f_\pi(d(l_i, l_h)) = \theta_{d1} d(l_i, l_h) + \theta_{d2} d(l_i, l_h)^2. \tag{4}$$

Adding the disturbance $\varepsilon_{i,\pi,h}$, and substituting into equation (1) the logit model becomes

$$W_{i,\pi,h}^l = \theta_{p,\pi} p(c_i, h, \pi) + g_\pi(q_h(s), s_i) + \theta_{d1} d(l_i, l_h) + \theta_{d2} d(l_i, l_h)^2 + \varepsilon_{i,\pi,h}. \tag{5}$$

The properties of the model are completed by assuming our composite agent knows $\varepsilon_{i,\pi,h}$ at the time the hospital choice is made, and the vector of disturbances has a distribution, conditional on the other right hand side variables, which is i.i.d. Type 1 extreme value. Notice that since we have not indexed an "outside" option, the logit analysis conditions on women who only consider giving birth at a hospital. We estimate this model using maximum likelihood.

We consider three different assumptions regarding the price coefficient $\theta_{p,\pi}$, in the estimation

$$\begin{aligned} (a)\ \theta_{p,\pi} &= \theta_p; \\ (b)\ \theta_{p,\pi} &= \theta_{p,\pi}; \\ (c)\ \theta_{p,\pi} &= \theta_0 + \theta_1 . pcap_\pi \end{aligned} \tag{6}$$

where $pcap_\pi$ is the insurer's capitation rate.

Equation (2) states that the price is exactly equal to our measure of the expected list price for the patient's diagnosis multiplied by one minus the observed average discount. I.e. this specification assumes no measurement or expectational error in price. We define the expected list price to be the average list price for the particular hospital over patients in diagnosis categories defined by unique combinations of: age (categories 11-19, 20-39, 40-49 and 50-64), principal diagnosis (21 categories for women giving birth including, for example, "normal delivery", "previous Cesarean Section" and "early labor"), Charlson score and diagnosis generating the Charlson score. Both principal diagnosis and Charlson score are based only on diagnoses known on admission. We are constrained to using

these fairly broad definitions of similar patients because we encounter small sample problems when we define narrower groups. To the extent that either the aggregation generates measurement error in our price measure, and/or the small cell sizes generate estimation error in our estimate of expected price, we expect an attenuation bias in the estimated price coefficient.

Equation (3) restricts the $g_\pi(\cdot)$ term in a way consistent with the previous literature: we assume it is determined by a hospital fixed effect plus interactions between hospital characteristics and patient characteristics that are known on admission and expected to be correlated with severity. In the inequalities analysis below we define over 100 patient severity groups and allow these to freely interact with hospital fixed effects. We can not do this in the logit analysis because it would imply estimating almost 20,000 coefficients and a similar number of expected price terms (without error), putting us in a range of values where an incidental parameter problem, similar to that described in Neyman and Scott (1948), would make coefficient estimates very unreliable. So we assume the interaction terms are determined by linear interactions between hospital and consumer diagnostic characteristics. Included in z_h are the number of nurses per bed and indicators for teaching hospitals, for-profit hospitals and hospitals that offer transplant services (a proxy for high-tech hospitals). We also include a measure of the quality of delivery and birth services taken from Ho (2006): hospitals were rated on a scale from 0 to 1, where 0 indicated that no delivery/birth services were provided and a higher rating indicated that a less common (assumed to be higher-tech) service was offered. The patient characteristics in x_i are the expected probabilities of death in hospital and of transfer to acute care setting or skilled nursing facility given the patient's age group, principal diagnosis and Charlson score.

While these interactions, like those used in the previous literature, are sensible given the constraints imposed by the methodology, we do not expect them to be sufficient to fully address the price endogeneity issue noted above. To the extent they do not there will be an error in the approximation in equation (3) so that equation becomes

$$g_\pi(q_h(s), s_i) = q_h + \beta z_h x(s_i) + e_\pi(q_h(s), s_i).$$

The logits assume $e_\pi(q_h(s), s_i) \equiv 0$. If this is not the case and hospital quality is both regarded as more important for more severely ill women and is positively correlated with hospital price, we expect this error to bias the price coefficient upwards. Note, however, that if we only consider the data on women with similar severities we get close to fixing s_i, i.e. then $s_i \approx s^*$. Among these women the error is essentially constant for each hospital and is therefore captured by the hospital fixed effect. So a separate analysis for women with similar severities should mitigate this source of omitted variable bias.

The logit analysis assumes that both the distance coefficient and the quality-severity interaction terms do not vary across insurers. We begin by also assuming a common price coefficient across insurers. After presenting these results, we provide another set of results which partially controls for the omitted variation in patient-severity hospital-quality interactions by restricting our attention to the least sick patients in our data. For comparison we also estimate separately on the rest of the

sample, a group with a more diverse set of severity conditions. Finally we let the price coefficient differ across insurers and investigate whether there is a significant relationship between the percent of the insurer's payments to primary physicians that are capitated and the price coefficient.

6.1 Logit Results

A summary of the results is reported in Table 4. The price coefficients, price interaction terms and distance coefficients are reported, together with the sample size, for each specification. In each case the distance coefficient is negative and highly significant, with a magnitude that is consistent with estimates from the previous literature. [29]

The price coefficient from the full sample of delivery/birth discharges is positive and significant with a t-value of approximately 5. Recall that we would expect a positive bias in that coefficient if high priced hospitals were high quality hospitals, and the severity-hospital interactions we included were not sufficient to control for hospital quality conditional on the severities that determined hospital choice. To see if this might be the source of a problem, the next column provides the results from the same specification when we restrict the sample to the least-sick women. We identified these women with the help of obstetrical experts at Columbia Presbyterian hospital. They are the subset of patients who are aged 20-39, have a Charlson score of 0, and whose principal diagnosis and comorbidities were defined by the obstetricians to be "routine". Our sample contains 43,742 of these patients. Since they have similar severities we would expect there to be less variance across patients in the importance of hospital quality differences, and this should mitigate the omitted variable bias.

When we use the sample of less sick woman, the price coefficient becomes negative (magnitude -0.017) and marginally significant (standard error 0.009). The same specification on the subsample with the sickest patients, the group of patients where we think the quality severity interactions are likely to be both more variant and more important in determining hospital choice, yields a positive price coefficient again (of .012), and this time with a t-value of 6. We conclude that we need a better way to control for hospital quality/patient severity interactions.

Next we look for interactions between price and insurer fixed effects. Insurers in the table are sorted by declining proportion of capitated payments to primary physicians. When we use the sample with the least sick patients, Blue Cross and Blue Shield, the plans which have the lowest proportions of capitated payments, have small, positive and insignificant price coefficients. All four of the remaining insurers have price coefficients less than 0. The negative price coefficients are significant for Pacificare and Health Net, two of the three carriers that favor capitation the most (97% of payments for Pacificare and 80% for Health Net). The remaining carriers, Aetna and Cigna, have relatively small sample sizes (6291 and 8097 birth discharges respectively, compared to 15,479 for Pacificare and 16,950 for Health Net), which helps explain the larger standard errors on their price coefficients. When we remove the price-insurer interactions and instead include an interaction between price and the percent capitation in the insurer, the price coefficient is positive

[29] See, for example, Gaynor and Vogt (2003) and Ho (2006).

and the interaction term negative with almost twice the magnitude of the price coefficient. Both are highly significant; the t-value of the capitation interaction is 7.7.

When we do the same exercise with the subsample of sicker patients, the price-insurer interaction term is still negative for Pacificare, although insignificant at p=0.05 and smaller in magnitude than for the healthier population. All other insurers' price coefficients are positive and three out of five are statistically significant, again pointing to the need for a better way to control for hospital quality/patient severity interactions. The third specification, including a price-percent capitation interaction, again generates a positive price coefficient and a negative interaction term (implying that insurers that favor capitated payments generate physician referrals that are more price-based than those of other physicians). However, the magnitudes are much smaller than for the least sick sample and the implied overall price coefficient is positive even for insurers with 100% capitated payments to primary physicians.

We now go back to the sample with the least sick patients and look for the implications of the logit analysis for the relative magnitudes of the distance and price effects. Consider first the distance coefficient. We calculate the impact of a one mile increase in distance for hospital h, holding all else fixed, on the probability that a particular patient i visits that hospital. We then take the average over patients and a weighted average over hospitals. The average effect of the one mile distance increase is a 13.7% reduction in the probability that the hospital is chosen. [30] Next we conduct a similar exercise to evaluate the magnitude of the price effect. Consider Pacificare's referrals for its least sick patients; the insurer with the most negative estimated price coefficient. The implied average effect of a $1000 increase in a hospital's price, holding all other prices constant, is a 5.2% reduction in the probability that the hospital is chosen. So the price increase we would require in order to compensate for a one mile increase in distance would be approximately $2,600. This is more than two thirds the average price for the less-sick patients (which is $3380 and has a standard deviation of $1870.) All the other insurers' price coefficients are considerably less negative, implying a considerably larger price distance tradeoff. These numbers accentuate our worry that omitted variables and errors in observed variables may be causing important biases in the logit estimates.

7 Inequalities

As noted we are worried that the logit analysis does not fully control for variation in quality conditional on severity and that this might cause a positive bias in the price coefficient. In addition that analysis compels us to use average prices within quite broadly-defined patient groups because narrower groups would increase the variance in our estimated price accentuating the impact of measurement error in price. We do not know of a way to address both the absence of adequate controls for quality by severity and measurement errors in price in the context of a multinomial

[30] The average distance to the chosen hospital for the less-sick patients included in the sample is 6.45 miles; the standard deviation is 10.11 miles. The weighted average probability that a particular hospital is chosen is 2.7%, where the weight is the number of discharges.

choice model[31], so we now provide an alternative estimation technique.

The method is based on a revealed preference inequality: it is assumed that the chosen hospital is preferred to feasible alternative hospitals. Patients are assigned to detailed severity groups based on information in the discharge data. We consider all couples of same-insurer, same-severity patients whose chosen hospitals differ but both of whose choices were feasible for both agents. Within each couple we sum the inequalities obtained from the fact that each patient preferred her choice to the choice made by the other. Since the severity-hospital interactions from the two inequalities are equal but opposite in sign, when we sum the inequalities the interaction terms difference out. Revealed preference implies that this sum is positive, and this constrains the remaining parameters.

More formally let $S_\pi(h, h', s)$ be the set of patients from plan π with severity s who chose hospital h but had hospital h' in their choice set. For notational simplicity let $\Delta x(i, h, h') = x_{i,h} - x_{i,h'}$ for $x = W(\cdot)$ or price, and let $\Delta f_\pi(l_i, l_h, l_{h'}) = f_\pi(d(l_i, l_h)) - f_\pi(d(l_i, l_{h'}))$ provide the difference between the distance from patient $i's$ home to hospitals h and h'. Then the observable part of our inequalities is formed by taking couples of patients $i \in S_\pi(h, h', s)$ and $i' \in S_\pi(h', h, s)$ and using equation (1) to form

$$\Delta W(i, h, h') + \Delta W(i', h', h) \quad = \tag{7}$$

$$\theta_{p,\pi}\Big[\Delta p(c_i, h, h') + \Delta p(c_{i'}, h', h)\Big] + \Big[\Delta f_\pi(l_i, l_h, l_{h'}) + \Delta f_\pi(l_{i'}, l_{h'}, l_h)]\Big].$$

Revealed preference implies that this equation is expected to be greater than zero when we evaluate it at the true value of θ and the expected prices. We then average over all couples $i \in S_\pi(h, h', s)$ and $i' \in S_\pi(h', h, s)$ for all $h' \neq h$, and this averages out the expectational and measurement errors in price.

Since we have removed the quality/severity interaction terms and no longer need to estimate their coefficients we can define our severity groups at a more detailed level than was possible in the logit analysis. Moreover since we average over all such couples in all severity groups we eliminate the effect of estimation error in price so we can define the price terms in as narrow a set of price groupings as we like. Note, however, that this procedure does rely on the expected price varying

[31] We did experiment with two techniques that addressed the absence of quality controls. One was to estimate separate logit models with separate hospital fixed effects for different severity groups (similar to those estimated for sicker and less-sick patients, but using the more detailed severity definitions utilized in the inequalities analysis below). The problem here was that there were too few patients per hospital in almost all of our severities. There are well over a hundred severity groups in our data, and the six largest severity groups together account for only 45% of our sample. Moreover in the sixth largest severity group over half of the hospitals have less than twenty admissions, 45% have less than ten, and almost 20% have two or less; and the numbers are notably worse for smaller severities. More formally for consistency we require the number of observations for each severity group/hospital combination to grow large, and this is not a reasonable approximation for samples based on a single severity group. In addition, since there have to be even more price groups than severity groups, use of the separate logit fixed effect analysis for each severity group requires us to use more detailed price groups than in the logits reported above and therefore accentuates the price measurement problem (whereas the inequalities average the differences in prices between price groups within a severity over severity groups). We also tried Chamberlain (1980)'s conditional likelihood estimator. At least under particular distributional assumptions this addresses the absence of quality controls, though not the measurement error in price. However the conditional likelihood estimator is not suitable for our problem because its computational burden grows as the combinatorial formula for the number of ways the total patients in a severity group can be divided among the hospitals conditional on the given number of patients (in the severity group) and hospitals. The number of patients and hospitals in our study is just too large to make this feasible.

within a hospital across patients who have the same severity of illness; otherwise the price terms will be differenced out along with the interaction terms.

7.1 Severity and Price Groups

Our severity groups are assumed to be defined in sufficient detail that the severity-hospital interactions absorb all unobserved variation, other than price and distance, that affects choices and may be correlated with price. That is though we require price variation across patients in different price groups within a given severity, that variance cannot affect choices except through the price variable itself. We now provide details of our severity and price definitions and consider whether these requirements are satisfied. Our definitions are chosen following the advice of the obstetricians we consulted at Columbia Presbyterian Hospital. As one input to the definitions, these experts assessed the list of principal diagnoses and comorbidities in our data, assigning each a rank from 1 to 3. A "1" indicated a routine diagnosis (such as normal birth or immunization of the newborn) and "3" indicated something more serious; see Appendix 3 for a complete list.

We use narrower definitions for severity and price than were used in the logit analysis. Severity groups are now defined by a unique combination of age, principal diagnosis, Charlson score, diagnosis generating the Charlson score and the rank of the most serious comorbidity, other than principal diagnosis, that is listed in the discharge record. [32] Our expected list price, on the other hand, is now defined as the average list price for the particular hospital across women with the same severity (as just defined) who also have the same number of most seriously-ranked comorbidities.[33] That list price is interacted with 1 minus the average hospital discount to calculate our baseline price variable.[34] These definitions generate many more groups than those used in the logit analysis. For example, for the first insurer in our data, there are 63 populated groups defining prices using the logit-based categories (recall that the logits did not allow severity groups to interact with the hospital fixed effects). There are 106 severities and 272 price groups under the more detailed definitions.

The obstetrical experts we interviewed advised us that these detailed price groups, conditional on severity, were unlikely to be important in terms of hospital choice. The price groupings are more detailed than those used for severity only in that they break out patients by the number of comorbidities of the highest rank as well as the identity of that rank. The number of comorbidities

[32] For example a woman aged 25 with a normal delivery, Charlson score of 1 caused by diabetes and a maximum rank of 2 would be in a different severity group from a woman with the same age, principal diagnosis and maximum rank but whose Charlson score of 1 was caused by mild liver disease. A woman aged 25 with a normal delivery, Charlson score of 0 and maximum rank of 1 would have a different severity group from a similar woman whose maximum rank was 2 (but where that co-morbidity was not severe enough to trigger a Charlson score above 0).

[33] For example if two women have the same age and principal diagnosis and a zero Charlson score but one has a migraine (a rank 1 comorbidity) and one has a viral infection (rank 2), the women have different severities and different prices. If neither woman had a migraine but one had a viral infection and the other had a viral infection and also a thyroid disorder (both rank 2 comorbidities), they would have the same severity but different prices.

[34] Provided expectations are unbiased the average of actual prices will converge to the average of expected prices, so we could have used actual and not a measure of expected prices in our inequalities. However the expected price variable we do use has the advantage that it uses the information from all same plan same price group patients in comparisons across couples of hospitals, not just those for whom there was a feasible switch.

of a given rank, conditional on severity, was considered unlikely to affect the hospital choice. While a physician might refer a pregnant woman with a normal delivery but a comorbidity of rank 2 (such as a viral infection or a thyroid disorder) to a different hospital from a similar patient with only rank-1 comorbidities, this would be a hospital well-equipped to deal with high-risk pregnancies rather than the specific comorbidity, and the presence of two rather than one rank-2 comorbidities would not affect the referral decision. In contrast, our experts agreed that the number of comorbidities of a particular rank would be likely to affect the tests performed and drugs prescribed and therefore the price.

The price variation used in the inequality analysis is a difference in price differences. More precisely it is a difference in expected price differences, where an expected price is measured by average prices associated with a given admission condition (defined by a combination of diagnosis, age, and comorbidities) at a hospital. We never compare prices between hospitals for different admission conditions.

More concretely, the price variable that enters the inequalities is the sum of the price differences for two patients who went to different hospitals but who could have chosen each others' hospitals. For each patient we find the price difference for that patient between the two different hospitals (i.e. we are comparing the price for the same price groups at different hospitals). Both patients are in the same severity group (this is what differences out the severity specific fixed effect) but in different price groups. Because they are in different price groups the difference in prices for the two patients at the two different hospitals is different. It is the variance in this "second difference", or in this difference in difference, that is driving the results.

The associated variances are illustrated in Table 5. The data used in the table are from patients who have a Charlson score of 0 and a given maximum rank but different numbers of most seriously-ranked comorbidities. 63450 out of 64691 patients in our inequalities sample have a Charlson score of 0[35]. Cells in the table correspond to a particular maximum rank and number of diagnoses of that rank, but average over four age groups and 21 diagnoses which are disaggregated in the actual analysis[36]. The difference across hospitals for a patient who is admitted in a given price group is quantified in the columns labelled SD which give the variance for that price group across hospitals (averaged over age and diagnosis). There clearly are differences across hospitals in a given price group and these differences increase as we move down the rows of the table (i.e. as we increase the number of comorbidities of maximum rank). The second difference is across price groups within a severity group (defined by diagnosis, age, and the maximum rank of the comorbidities). This is illustrated by the differences across rows in the columns labelled Price ($) in Table 5 (again averaged over age and diagnosis). Clearly the mean prices are ordered as expected, and the differences are usually highly significant (the bracketed numbers provide the standard error of the mean).

[35] We exclude a few patients who have no comorbidities known on admission besides the principal diagnosis and therefore have 0 diagnoses of maximum rank.

[36] We use aggregated groups because for many of the hospitals there are no patients in some of our actual severity groups and these hospital severity pairs would not be used in the inequality analysis which follows.

Finally we note that when we go to the actual disaggregated groups (disaggregated further by age and diagnosis), an analysis of variance indicates that moving from severity to price groups explains an additional 12% of the variance in price (from 50% to 62% of the total variance), ensuring there is meaningful variance in price after we fully condition on our severity groups. Of course in addition to requiring that there is variance in price conditional on severity we are also assuming that hospital "quality" does not differ across price groups within a severity. There may be many dimensions of quality but expected health outcomes are clearly among them. Online Appendix 4 provides the analogue of Table 5 when the four mother and infant outcome measures used in Table 3 are substituted for the "Price ($)" columns in Table 5[37]. There we see that as we move from max rank 1 to 2 and then max rank 2 to 3 in a given row (i.e. conditional on a given number of max rank comorbidities) 19 out of 20 of the differences are of expected sign (the single difference with the wrong sign is not significant) and most are statistically significant (17 out of 19)[38]. When we consider differences in outcomes across entries that correspond to different numbers of comorbidities conditional on the same maximum rank, 6 out of 28 are decreases when we expect increases (though none are significant decreases), and of the remaining 22 only 8 are significant. Though the latter differences (those across rows within a given column) are small relative to those across columns, they are large enough to induce us to use the more detailed severity groups actually used in the analysis. When we further disaggregate by diagnosis and age (i.e. to the severity groups we actually use) and perform χ^2 tests for the outcome differences across price groups conditional on severity groups and hospitals (to correspond to our severity specific hospital effects) we find no significant differences.

7.2 Inequality Analysis

The inequality model makes the following assumptions:

$$p(c_i, h, \pi) = \delta_h^o lp^o(c_i, h) - \epsilon_{p(c_i),\pi,h} \equiv p^o(c_i, h, \pi) - \epsilon_{p(c_i),\pi,h}. \tag{8}$$

$\delta_h^o lp^o(c_i, h)$ was assumed equal to expected price in the logit analysis, so the difference between this specification and that used in the logit analysis is that the inequality analysis allows for measurement error in price and the logit analysis did not. We assume this error is mean zero conditional on the patient's plan and chosen hospital. Also

$$g_\pi(q_h(s), s_i) = g_\pi^o(q_h(s), s_i) - \epsilon_{s_i,\pi,h}, \tag{9}$$

i.e., the inequality analysis places no restrictions on the quality severity interactions and allows for classification error in those interactions which is assumed to be mean zero conditional on the

[37] As for Table 3 we use a slightly different dataset that includes infant outcome variables. We exclude a small number of patients with no comorbidities known on admission other than the principal diagnosis.

[38] These numbers only consider differences where both entries are based on more than 400 patients. We thank Jesse Shapiro for suggesting this table.

patient's plan and chosen hospital. Finally

$$f_\pi(d(l_i, l_h)) = \theta_{d,\pi} d(l_i, l_h), \tag{10}$$

which differs from the specification in the logit analysis in that the squared term in distance has been eliminated because it did not affect any of the parameters of interest and would complicate the algebra below.

Substituting into equation (7) for a same-plan same-severity couple who could have chosen each other's hospital (an $i \in S_\pi(h, h', s)$ and $i' \in S_\pi(h', h, s = s_i)$) our revealed preference inequality becomes

$$0 \;\leq\; \Delta W^I(i, h, h') + \Delta W^I(i', h', h) \;\; = \tag{11}$$

$$\theta_{p,\pi}\Big[\Delta p^o(c_i, h, h') + \Delta p^o(c_{i'}, h', h)\Big] + \theta_{d,\pi}\Big[\Delta d(l_i, l_h, l_{h'}) + \Delta d(l_{i'}, l_{h'}, l_h)\Big] - \Delta\epsilon_{i,h,h'} - \Delta\epsilon_{i',h',h}$$

where $\Delta\epsilon_{i,h,h'} \equiv \theta_{p,\pi}\Delta\epsilon_{p(c_i),\pi,h,h'} + \Delta\epsilon_{s_{i'},\pi,h'.h}$. Note that all coefficients are plan specific.

Our inequalities for hospital h and insurer π are simply averages of equation (11) across switches between patients who chose hospital h and those who chose another hospital but could have chosen h and had the same severity and insurer as the patient who chose h . Formally, let $N_{h,h',s}$ be the number of patients in the set $S(h, h', s)$ and for any $x(\cdot)$ define

$$\Delta\overline{x}(h, h', s) \equiv \frac{1}{N_{h,h',s}} \sum_{i \in S(h,h',s)} \Delta x(i, h, h'). \tag{12}$$

Then averaging equation (11) over $i \in S(h, h', s)$ and $i' \in S(h', h, s)$ we get

$$\theta_{p,\pi}\Big(\Delta\overline{p}(h, h', s) + \Delta\overline{p}(h', h, s)\Big) + \theta_{d,\pi}\Big(\Delta\overline{d}(h, h', s) + \Delta\overline{d}(h', h, s)\Big) - \Delta\overline{\epsilon}(h, h', s) - \Delta\overline{\epsilon}(h', h, s) \geq 0.$$

The moment inequalities we use in estimation are weighted averages of these inequalities where the weights are given by the fraction of comparisons that each contributes, or

$$w(h, h', s) \equiv \frac{N_{h,h',s} N_{h',h,s}}{\sum_s \sum_{h'>h} N_{h,h',s} N_{h',h,s}}.$$

If we let $\to_P$ read converges in probability, and note that our assumptions imply

$$\sum_{s,h'>h} w(h, h', s)\Big(- \Delta\overline{\epsilon}(h, h', s) - \Delta\overline{\epsilon}(h', h, s)\Big) \to_p 0,$$

then our model implies that

$$\sum_{s,h'>h} w(h, h', s)\left[\theta_{p,\pi}\Big(\Delta\overline{p}(h, h', s) + \Delta\overline{p}(h', h, s)\Big) + \theta_{d,\pi}\Big(\Delta\overline{d}(h, h', s) + \Delta\overline{d}(h', h, s)\Big)\right] \to_P \kappa \geq 0. \tag{13}$$

The inequality in equation (13) is in terms of observables and the parameters of interest. Since it is an inequality, if a particular $[\theta_{p,\pi}, \theta_{d,\pi}]$ satisfies (13), then so will $[\kappa\theta_{p,\pi}, \kappa\theta_{d,\pi}]$ for any $\kappa > 0$. This implies that there is a free normalization, so we set $\theta_{d,\pi} = -1$. Thus we will only be able to estimate the ratio $\theta_{p,\pi}/||\theta_{d,\pi}||$, which at the risk of some notational confusion, we will henceforth simply call $\theta_{p,\pi}$ [39]

The inequality in (13) bounds $\theta_{p,\pi}$. We can generate additional inequalities, and therefore bounds, by multiplying each inequality in equation (11) with an "instrument", say z, whose sign is the same for all observations. The additional moments will generate lower bounds if the expected value of $[\Delta\overline{p}(h, h', s) + \Delta\overline{p}(h', h, s)] * z$ is positive and upper bounds otherwise**.** For a variable to be an instrument it must be known by the agents when their decisions are made and mean independent of the measurement errors. Our instruments are the positive and negative parts, respectively, of the distance difference terms defined above: $\Delta d(l_i, l_h, l_{h'})_+ \equiv max\{d(l_i, l_h, l_{h'}), 0\}$, $\Delta d(l_i, l_h, l_{h'})_- \equiv min\{d(l_i, l_h, l_{h'}), 0\}$, and analgously, $\Delta d(l_{i'}, l_{h'}, l_h)_+$, $\Delta d(l_{i'}, l_{h'}, l_h)_-$.

Details. We conduct the initial analyses separately by insurer. This allows all hospital-quality/ patient-severity interactions as well as the price coefficient to differ by plan. The left hand side of equation (13) is computed separately for each hospital. We then weight each of these terms by its estimated standard error and choose as our (set) estimator of $\theta_{p,\pi}$ those values that minimize the sum of squares of the negative parts of the resulting moments. For small hospitals we are concerned that the average error either in the inequality, or more likely in the estimate of its standard error, may not be close to zero. As a result we develop a separate inequality for each hospital that has more than 1000 patient switches but average over all hospitals with less than 1000 patient switches. Overall we have between 78 and 285 moments per insurer: one for each combination of an instrument and a major hospital and an additional moment per instrument that includes hospitals with fewer patients.[40] 95% confidence intervals for the estimates are generated using the method developed in Pakes, Porter, Ho and Ishii (2011). [41]

[39] The inequality in equation (13) is uniform over the (bounded) parameter space. Also, though the model allows for detailed hospital quality/patient severity interactions and errors in the price variable, it does rule out determinants of choice that are patient-hospital specific and are both: (i) not controlled for by the severity/quality interactions, price, or distance, and (ii) not differenced out by adding the difference in preference for hospital h over h' for one patient to the difference in preference for hospital h' over h for the other. The logit model does not allow for patient severity/hospital quality interactions or errors in the measure of price, but does allow for unobservables that are patient and hospital specific (though, in the version estimated, they have to be independently and identically distributed across both patients and hospitals). To the extent that the inequality model does not account for all the patient/hospital specific determinants of choice there will be a selection problem in the resulting estimates which would be expected to narrow the estimated bounds. If the selection term is important we should expect a (non-random) fraction of our inequalities to switch signs, and if selection is important enough we will reject the null that there are values of θ which satisfy all the inequalities. In fact we accept below. Moreover there is no evidence of a disproportionate number of negative inequalities (see below).

[40] We exclude from the analysis hospitals that have fewer than 150 switches with all other hospitals in the sample combined (when instruments are included, each pair of hospitals is required to have at least 150 switches whose value of the instrument is non-zero). We also tried estimating the coefficients keeping the smaller-hospital moments separate. The estimated coefficients were almost always smaller in magnitude than our baseline results, consistent with small hospitals introducing measurement error, but in qualitative terms the story did not change.

[41] This requires development of the variance covariance of the inequalities across moments, and the formula for

7.3 Allowing for Within Hospital Variation in Discounts Across Insurers

Our model allows for error in our hospital price measure but assumes that error is mean zero conditional on hospital and plan. This would not be the case if there were important differences in discounts within a hospital across insurers, so we now consider an extension that allows us to check for such differences.

Our data include the average negotiated discount at each hospital, our d_h. Our procedure for testing whether our results are robust to within-hospital plan-specific discounts begins with estimates of a model that explains d_h as a function of the share of a hospital's revenues that comes from each insurer interacted with hospital, plan, and market characteristics. The additional data on the insurer-specific share of hospital revenues needed for this analysis comes from the OSHPD hospital discharge and financial records for 2003. The estimated model is then used to generate a prediction for $d_{\pi,h}$. There are two possibilities. First we can use the model's prediction directly; we denote this $\hat{d}^1_{\pi,h}$. Alternatively we can subtract the predicted discounts of other insurers (appropriately weighted) from the observed d_h to generate a second prediction $\hat{d}^2_{\pi,h}$. The second procedure uses more of the hospital specific information on discounts, as it includes the hospital specific error in the discount equation, though the first is the traditional way of constructing model estimates. Finally the predictions are used to define price measures $p^1(.) = (1 - \hat{d}^1_{\pi,h})lp^o(c_i, h)$ and $p^2(.) = (1 - \hat{d}^2_{\pi,h})lp^o(c_i, h)$ which are substituted for our estimates of price in equation (8) in the inequality analysis. The errors from this prediction now combine with the other sources of measurement error to determine the $\varepsilon_{i_h,\pi,h}$ in that equation. So this analysis requires the additional assumption that the prediction error is uncorrelated with the list price and our instruments conditional on plan and hospital.

We specify a logistic functional form for $d_{\pi,h}$. The explanatory variables that are interacted with plan revenue shares at the hospital include indicators for for-profit hospitals and hospitals that are members of systems (groups of providers that bargain jointly with insurers), indicators for teaching hospitals, and both insurer and market fixed effects. However we also estimate other specifications that replace the fixed effects with market and insurer characteristics to check that our results are consistent with previous papers analyzing the impact of those characteristics on hospital prices. Details on both the models estimated and their coefficient estimates are provided in Appendix 2. Though the fits are not extraordinary (with R^2s just under 0.5), the results are intuitive and accord with the prior literature (for example discounts increase with the number of hospitals per population and decrease with the number of insurers per population).[42]

that matrix is available upon request. We do not implement other methods to generate confidence intervals, such as those in Chernozhukov, Hong and Tamer (2007) or Andrews and Soares (2009), because these require recomputing the variance of the moments at every value of the parameter vector evaluated. This is infeasible in terms of computational time in our application.

[42] Dranove and Satterthwaite (2000) and Gaynor and Vogt (2000) provide good reviews of the literature on discounts. Several other specifications were investigated. For example we replaced the insurer fixed effects with the plan percent capitation. This coefficient was positive and the other coefficient estimates were qualitatively unaffected by this change, foreshadowing our results that accounting for variation in discounts across insurers does not change the major results of our inequalities analysis. We also investigated whether the proportion of the insurer's patients sent to a particular hospital was correlated with the discount by including an interaction of this proportion with

7.4 Allowing for Variation in Discounts Across Diagnoses

Our baseline discount analysis makes the assumption that discounts are fixed across diagnoses within a hospital-insurer pair. In reality discounts may vary across services within a hospital. To ensure that ignoring this variation was not biasing our results we repeated the discount regressions allowing the discount for deliveries to differ from that for other diagnoses. We also ran specifications allowing the labor/birth discount to differ across different types of hospitals: for example hospitals with high-tech delivery services could have a particularly good reputation for obstetrics and therefore negotiate low discounts. Augmenting our baseline analysis to allow for a separate discount for delivery episodes generated a significant coefficient which implied that births had a 6 higher discount than the average for other diagnoses. However we did not find significant differences in this discount across hospitals. When we substituted the delivery-specific discounts into the inequality analysis the coefficients differed very little from the baseline specification. We also tried estimating a different discount for Cesarean sections but the estimated coefficient was not statistically significant. See Appendix 2 for further information.

7.5 Allowing for Errors in Our Distance Measure

Our measure of the distance between patients and hospitals is the distance between the centroid of the patient's home zip code and the zip code of the hospital. It therefore contains measurement error.[43] Assuming the error is mean zero implies that it does not affect the properties of our original inequality (equation 13) as that simply averages out the estimation error. However when we take the positive and negative parts of the distance and use them as instruments, those instruments will, in general, contain an error which is correlated with the error in distance in the original equation, and this can generate biases in our estimate of $\theta_{p,\pi}$.

To ensure that this was not having a major impact on our results we modified our distance instruments ($\Delta d(l_i, l_h, l_{h'})_+$, $\Delta d(l_i, l_h, l_{h'})_-$, $\Delta d(l_{i'}, l_{h'}, l_h)_+$, $\Delta d(l_{i'}, l_{h'}, l_h)_-$) as follows. Instead of using $\Delta d(l_i, l_h, l_{h'})_+$ we used $\Delta \tilde{d}(l_i, l_h, l_{h'})_+$ where

$$\Delta \tilde{d}(l_i, l_h, l_{h'})_+ = 1 \text{ if } \Delta d(l_i, l_h, l_{h'})_+ \geq 3, \quad \text{and } \Delta \tilde{d}(l_i, l_h, l_{h'})_+ = 0 \text{ otherwise.}$$

We did the analogous transform to the other distance instruments. This addresses the problem if we assume that the error in the distance is not greater than three miles, so that we know the incremental distance between hospitals is positive (negative) if the observed differences are greater

insurer fixed effects in the model. This relationship between the "channeling" of patients to a particular provider and the prices negotiated with that provider is analyzed in Sorensen (2003). When we excluded market fixed effects we estimated a significant positive relationship between patient channeling and discounts (a negative relationship between channeling and prices) for just one insurer, Blue Shield. The coefficient became insignificant when we added market fixed effects. We repeated the inequalities analysis for Blue Shield using this discount specification and the results changed very little.

[43] Latitudes and longitudes of the hospital's and patient's zip codes are taken from the Bureau of Census 1999 zip code file. Additional measurement error is caused by the fact that, if the zip code could not be located in the Bureau's internal database, the county internal point was assigned to the zip code. Multiple zip codes therefore have the same recorded latitude and longitude in some cases.

than three miles.

7.6 Inequality Results

Table 6 reports results from the inequalities analysis. The first column reports our main results. These assume that the price measure obtained by multiplying hospital specific discounts by the expected list price is correct up to an error which is mean zero conditional on the plan and the choice of hospital. The next two columns use $p^1(.)$ and $p^2(.)$ respectively. Their validity requires additional assumptions on the prediction errors generated when forming these variables, but they allow us to investigate whether our results are robust to allowing for plan-specific discounts within hospitals. The final column uses the modified distance measure described above to check for the possible impacts of errors in the distance measure.

There is no value for $\theta_{p,\pi}$ that satisfies all the inequality constraints in any specification except one (Blue Shield in column 4 of the table). When this occurs the estimation algorithm produces a point estimate: the value of $\theta_{p,\pi}$ that minimizes the sum of squares of the negative part of the (standardized) moments. Given the number of inequalities we have for each of our plans we are not surprised to find point estimates. In finite samples when each moment is evaluated at the true value of the parameter vector it generates a variable which distributes approximately normally. Consequently the greatest of the values from the moments which provide lower bounds has a positive bias. Similarly the least upper bound has a negative bias. So depending on the magnitude of the biases the bounds can easily cross producing a point estimate. The expected magnitude of these biases increases with the number of moments.

There is a standard statistical test for whether sampling errors of this form exist and the results always indicated that we could accept the null that there were values of θ that satisfied all the inequalities. Table 7 makes it clear why we accept the null. It provides estimates of the t-statistics obtained when we evaluate all moments used at the estimated value of $\theta_{p.\pi}$ for each plan. The model predicts that the expectation of all moments are non-negative. Our results indicate that of the 977 moments evaluated only 60, or about 6%, are less than zero, and only 7 out of 977, or 0.7% of the moments, have t-values less than -2. In four of the six plans studied none of the moments are significantly negative at the traditional p-value of .05. Health Net has 2 out of 182 moments with t-value less than -2 and Blue Cross has 5 out of 285 with t-values less than -2. For these two insurers we re-estimate $\theta_{p,\pi}$ after dropping the moments with t-statistics less than -2. The results are reported in the rows of Table 6 labeled "Drop $t < -2$".

In the first column of Table 6 the price coefficients for all insurers other than Blue Shield are negative and statistically significant at p=0.05. That for Blue Shield is small, negative and statistically insignificant. As is traditional for set estimators, we focus on confidence intervals for $\theta_{p,\pi}$. These are illustrated for each insurer in Figure 1. The coefficients for all insurers except Blue Shield are ordered by decreasing percent capitation (that is, the upper bound of the confidence interval for one insurer is below the lower bound for the insurer with the next-highest percent capitation). The picture is less clear for Blue Shield. Its confidence interval is above that for Blue

Cross and crosses zero.

The results from substituting $p^1(.)$ and $p^2(\cdot)$ for the $p^o(\cdot)$ in the inequality analysis are provided in columns (2) and (3) of Table 6, respectively. They are similar to the results in our main specification. The two major differences occur when using $p^2(\cdot)$; then the Health Net coefficient estimated when we drop the two negative moments is larger in absolute value, and the Blue Shield coefficient is positive with a confidence interval that crosses zero (instead of being negative with a confidence interval crossing zero). We conducted a number of other robustness tests that involved the price variable, but none had anything but the expected effect on the results [44].

Column 4 provides the results that use the modified distance instrument that takes account of the possibility of error in our distance measure (it uses the $\tilde{d}(\cdot)$ instruments defined in the prior subsection). The results are very similar to those above. All coefficients but those for Blue Shield, and to a lesser extent Health Net are similar to those in column 1. The confidence interval for Blue Shield indicates that the data are not informative about the Blue Shield price coefficient, and the Health Net coefficient, though always significantly negative, varies in magnitude with whether or not we keep the two inequalities that are significantly negative.

With the possible exception of Blue Shield, these results indicate that the allocation of patients to hospitals responds to the prices the insurer pays those hospitals. Moreover insurers with more capitated payments to physicians have hospital referral processes that place a more negative weight on prices than other insurers. We can not say much about Blue Shield. The bounds for its coefficient are large and vary quite a bit across the specifications we tried. Recall that Blue Shield is the only not-for-profit insurer in our data, and so might be expected to differ. As a result we disregard Blue Shield in the analysis that follows.

The difference between the inequality results and those from the logit analysis are striking. To get an idea of the importance of this difference Table 8 compares the elasticity of price with respect to distance computed using the logit estimates for the least sick patients (table 4), to those same elasticities computed using the price coefficient estimated from the inequalities (column 1 of table 6). That is, we measure on average how much further the consumer would have to drive (in percentage terms) to just offset a one percent price increase.

Consider first the comparison of elasticities derived from the logit price coefficients to those derived from the inequality estimates for the plans where the logits estimated a negative price coefficient. All the elasticities obtained from the inequality estimates are more than an order of magnitude larger than those obtained from the logit estimates, and some are more than two orders of magnitude larger. In addition two of the elasticities obtained from the logit estimates have the wrong sign. Notice also that the inequality estimates indicate that the average elasticity increases by almost a factor of four when we move from the least capitated for profit insurer (Blue Cross) with a capitation rate of 38% to Pacificare whose capitation rate is 97%.

[44] We repeated the inequalities analysis using just the list price (rather than its interaction with the discount). The pattern of results was unchanged in that high-capitation insurers had more negative price coefficients in general than other insurers. However all price coefficients were closer to zero than those in Table 6, consistent with our expectation that measurement error should affect these results.

8 Cost, Quality, and Distance Trade-Offs

This section is divided into two parts. In the first we derive insurer and severity specific estimates of quality differences across hospitals. The estimated qualities are quite similar across insurers. We illustrate this by estimating a model which requires the within-severity ratings of different plans to be linear transforms of each other, and then providing a plot of the restricted on the unrestricted quality estimates. The next subsection examines implications of these results. When the within-severity orderings are linear transforms of each other each plan's preference ordering over hospitals (equation 1) is a different linear function of price, distance, and a common quality index. This allows us to investigate how the quality-price-distance trade-off varies across insurers. We conclude this subsection by comparing our results on these trade-offs to the implications of the data on health outcomes (the data that underlies table 3). This reinforces our results and provides some external validity for our quality controls.

8.1 Plan and Severity Specific Hospital Quality Terms

The revealed preference inequality in equation (11) implies that any given value of $\theta_{p,\pi}$ generates a set of bounds for differences in the quality terms across hospitals. We evaluate these differences at the estimates of $\theta_{p,\pi}$ given in column 1 of Table 6.[45] Since we can only recover differences in quality and we can only compare hospitals within a market, we estimate quality coefficients that are: plan, severity, and market specific. To ease notation we will omit the plan and market indices below.

Recall that $\Delta\overline{x}(h,h',s)$ is the average of $\Delta x(i,h,h')$ among the patients with severity s who chose hospital h but could have chosen h'. Then since every one of those patients chose h over h' revealed preference implies

$$g(q_h,s) - g(q_{h'},s) \geq -\theta_p \Delta\overline{p}(h,h',s) + \Delta\overline{d}(h,h',s) + \Delta\overline{\epsilon}_{h,h',s} \equiv \hat{q}(h,h',s) + \Delta\overline{\epsilon}_{h,h',s},$$

with $\hat{q}(h,h',s)$ observable. Moreover since the $\overline{\epsilon}_{h,h',s}$ are mean zero conditional on the hospital choice

$$\hat{q}(h,h',s) \;\sim>\; \mathcal{N}\Big(\; \underline{q}(h,h',s), \sigma^2(h,h',s)/N_{h,h',s}\Big), \quad \text{where } \underline{q}(h,h',s) \leq g(q_h,s) - g(q_{h'},s), \quad (14)$$

"$\sim>$" reads converges in distribution to, $\mathcal{N}(\cdot,\cdot)$ is the normal distribution, $\sigma^2(h,h',s)$ is the variance of $(-\theta_p \Delta p(c_i,h,h') + \Delta d(l_i,l_h,l_{h'}))$ across observations in $S(h,h',s)$, and $N_{h,h',s}$ is the cardinality of that set.

Each couple of hospitals generates two quality bounds of this form: one from the patients who chose h but could have chosen h' and one from those that chose h' and could have chosen h. The former provides a lower bound and the latter an upper bound to the difference in quality between

[45] The correlation of the quality terms across different values within the confidence intervals of $\theta_{p,\pi}$ is nearly one, so it makes little difference which of the values in the confidence intervals reported there we use.

h and h'. So if there are H hospitals in a market, there are $H(H-1)$ estimates of quality bounds for each severity.

Estimating the quality bounds. Recall that we can only bound *differences* in hospital quality. So we set one hospital's quality to be zero (the same hospital for each insurer for a given severity and market). Indexing that hospital by H one can show that the inequalities that relate to hospital h are given by

$$\overline{q}(h,s) \equiv \min_{h' \neq h} E\Big[-\hat{q}(H,h',s) - \hat{q}(h',h,s)\Big] \geq g(q_h,s) \geq \max_{h' \neq h} E\Big[\hat{q}(h',H,s) + \hat{q}(h,h',s)\Big] \equiv \underline{q}(h,s). \tag{15}$$

We stack these inequalities for each hospital, weight each by its estimated standard error, and then find the (set) estimator that minimizes the squared inequality violations.

Recall that we used very detailed severity groups for the estimation of price coefficients so as to ensure we eliminated biases that might be caused by unobserved quality terms. The sample sizes associated with those groups are quite small and to obtain the quality estimates we do not average over severity groups as we did to obtain the price coefficient estimates**.** Moreover we are not interested in orderings of hospitals at that fine a level of severity. We therefore use the severity classifications given to us by the obstetricians we consulted to aggregate into five "super-severity" groups. These consist of four groups all of whose patients have identical principal diagnosis and comorbidity rankings, and a fifth group which contains all the remaining patients[46]. Finally the actual estimates of the quality bounds depend on the prior estimates of $\theta_{p,\pi}$. The results presented below use the point estimates from the first column of Table 6, but the implied quality bounds varied very little when we considered other points within their respective confidence intervals.

In computing the quality estimates we included moments for patients who went to hospital h but could have chosen hospital h' for a given severity if there were five or more patients who were admitted to hospital h and had hospital h' in their choice set. Over our 12 markets and five severity groups, we obtained 1176 quality estimates[47]. Almost all of these estimates, 1078 of them, come from our five largest markets (Los Angeles, Orange County, Inland Empire, the Bay Area, and San Diego), so we confine the remainder of the analysis to these five markets.

Does the implied order make logical sense? For the ordering across hospitals to make logical sense it must obey transitivity. There are at least two ways we can check this, one of which does

[46] The super-severity groups are: Group I contains 55% of patients who have a rank 1 (routine) principal diagnosis, rank 1 comorbidities and are young; Group II has 11% of patients who have rank 2 principal diagnosis, rank 1 comorbidities and young; Group III has 15% of patients who have rank 1 principal diagnosis, maximum rank 2 comorbidities and are young; Group IV has 12% of patients and they have a rank 2 principal diagnosis, maximum rank 2 comorbidities and young; and Group V has 6% of patients that are not included in the other groups.

[47] 477 of these were sets and 699 were points. We tested whether the points satisfied the appropriate vector of moment inequality constraints (the sets necessarily do). Slightly more than half did. However when we go to switches between individual hospitals for a given severity and plan there is a limited amount of data per moment. So the asymptotic approximation inherent in the moment inequality test statistic is questionnable. Moreover, as we show below the actual estimates satisfy most of the properties our priors might associate with them.

not rely on our estimates of the price coefficient and one of which does.

Temporarily ignore estimation error. Then if both $\Delta\overline{p}(h,h',s)$ and $\Delta\overline{d}(h,h',s)$ are positive the perceived quality of h for severity s must be higher than that of h'. This because patients chose hospital h over h' despite the fact that h was both more distant and had higher prices. This fact generates a partial ordering across hospitals that *does not require* either estimates of the price coefficients or estimates of the quality terms. Alternatively we could use our estimate of $\theta_{p,\pi}$ to ask whether the partial order obtained from the sign of $\Sigma(\theta_{p,\pi}\Delta\overline{p}(h,h',s) - \Delta\overline{d}(h,h',s))$ for each pair (h,h') obeys transitivity.

The estimation procedure we use does not guarantee that the order we obtain from using either of these sets of inequalities satisfies the logical condition of transitivity. I.e. there could be cycles of the form

$$h_1 \succ h_2, \;\; h_2 \succ h_3, \;\; \text{but} \;\; h_3 \succ h_1,$$

or even more simply we could find that

$$h_1 \succ h_2 \;\; \text{but} \;\; h_2 \succ h_1.$$

To check this we compute all possible cycles from the both the "non-parameteric" bounds and the bounds that use our estimates of θ_p for each of our five insurers, in each of our five markets for each of our five severities. The non-parametric procedure yields only 543 possible orderings, and none violate transitivity. When we use our estimates of θ_p and the estimation algorithm described above there are 10,526 possible orderings, and of these 1069, or about 11% actually cycle. However almost all of these are associated with bounds that are estimated imprecisely. Only 3 or .03% of the possible cycles are significantly negative at the 5% level. We take this as evidence that the data generates a hospital ordering that satisfies rationality constraints. We now consider that ordering in more detail.

Similarity of the implied orders across plans 1078 estimates is still too many to examine individually, and our primary interest is not in the quality estimates per se but in the implied trade-off between price, quality, and distance. Moreover the similarity in the estimated rankings of hospitals across insurers within our severity groups is striking, and this implies that some aggregation across plans is warranted. Note that since patients are assigned to a unique insurer, there is no statistical relationship between the moments used for the different insurers (except the relationship due to our using the same price measure across insurers). The similarity in ranks is a result of the similarity in (almost) statistically independent quality estimates generated by the referral processes of the different plans.

Figure 2 illustrates this similarity. This figure plots the estimates obtained from imposing the constraints that the plan specific orderings for our five severity groups in our five largest markets are linear transforms of one another. That is, reintroducing the plan (π) and market (m) indices,

we substitute

$$g_\pi(q_h, s) = \alpha^0_{\pi,m,s} + \alpha_{\pi,m,s} q_{h,s} \tag{16}$$

for the quality terms into equation (15), and re-estimate. When we do this the α^0 coefficients can not be separated from the quality of the reference hospital in each market and, since we can only compare quality estimates across plans, the $\alpha_{\pi,m,s}$ can only be analyzed proportionately to those of a base plan. Consequently in what follows we set the $\alpha_{\pi,m,s}$ coefficient for Blue Cross equal to one in each market and severity.

When we impose the constraints in equation (16) we estimate 452 coefficients. Figure 2 plots the constrained against the unconstrained estimates[48]. The fitted line captures 98.2% of the variance of the unconstrained estimates. We then imposed the futher constraint that

$$\alpha_{\pi,m,s} = \alpha_\pi. \tag{17}$$

This reduced the number of parameters estimated to 380. Figure 3 plots the constrained against the unconstrained estimates after imposing the additonal constraint. The fitted line now captures 95.7% of the unconstrained variance. Though the difference in fit between figures 2 and 3 is noticeable, it is rather small; we lose about three hundredths of one percent of the fit per additional constraint. Moreover if we impose the constraint in equation (17) there is a straightforward way to compare the way different plans trade-off costs, quality and distance.

8.2 Trade-Offs

We now accept the constraints in equations (16) and (17) and substitute the results into the equation which determines hospital choice (equation 1). To get directly at the price-quality and distance-quality trade-offs we divide the resulting equation by α_π so

$$W_{i,\pi,h} \propto \left(\frac{\theta_{p,\pi}}{\alpha_\pi}\right) p(c_i, h, \pi) - \left(\frac{1}{\alpha_\pi}\right) d(l_i, l_h) + q_{h,s_i} + \left(\frac{1}{\alpha_\pi}\right) \epsilon_{i,\pi,h}. \tag{18}$$

Table 9 provides the plan-specific estimates of the coefficients in equation (18).

The first two rows of the table reproduce the capitation rates and price coefficients from prior tables. As noted the (absolute value of the) price coefficients are ordered by the capitation rates. The third row shows that the quality coefficients are ordered in exactly the same way. As a result the ratio of the price coefficient to the quality coefficient is virtually constant across plans. The fourth row shows that this ratio lies between -0.29 and -0.30 for all five plans. In addition this ratio is estimated quite precisely. If we take upper and lower bounds to that ratio obtained by dividing the upper (lower) limit of the confidence interval for $\theta_{p,\pi}$ by the lower (upper) limit of the

[48] When the unconstrained quality estimate was a set the error in the fit of the point was set to zero if the line went through the set, and was set to the distance between the set's bound and the line when it did not. 62 of the 452 constrained coefficients were sets. When the constrained estimate was a set and the line went through the set, we placed the point on the line in the figures, and when the line did not go through the set we chose the closest value to the line from the set.

confidence interval for $1/\alpha_\pi$, we find that the lower limit only varies between -0.31 and -0.40 while, with the exception of Health Net, the upper bounds vary only between -0.22 and -0.25 (Health Net has an upper bound of -0.15, and as noted earlier our estimates of its values are somewhat sensitive to the precise specification of the price and distance terms).

Table 10 provides the estimates of $\theta_{p,\pi}/\alpha_\pi$ before we impose the constraint in equation (17); i.e. for each market and severity separately. There we see that Table 9 does hide some variance in the estimates of the parameter determining the cost-quality trade-off across markets and severities. However the difference between these numbers and those for $\theta_{p,\pi}/\alpha_\pi$ in Table 9 is largely in the smaller markets and severities. As a result when we impose the constraint in (17), the LA, Bay Area, and (to a lesser extent) Orange County moments for the first three severities dominate, and they do not differ much across either plans or severities.

The ratio of the price to the quality coefficient represents the trade-off between costs and quality. What the estimates are telling us is that the cost-quality trade-off is, as far as we can tell, independent of the capitation rate. This despite the fact that the higher the capitation rate the more sensitive hospital referrals are to price. Apparently though the high capitation plans are willing to send their patients to further away hospitals to save on hospital costs, they are not willing to sacrifice quality for cost savings. The trade-off between cost, quality and distance only differs between plans in the trade-off between patient convenience and cost, not between quality and cost.

Of course our "quality" measure is simply whatever is implicit in the referral process: it captures everything that makes the hospital attractive after accounting for price and distance. The results we report above indicate that the quality rankings are similar across insurance plans. Online Appendix 5 shows that there is fairly substantial variation in our measure of quality. There we convert the quality variable to \$000 by calculating $\alpha_\pi q_{h,s}/\theta_{p,\pi}$, and then we calculate its value for each patient-hospital pair. We take the standard deviation across hospitals for each patient and then average over patients in each super-severity group and each market. The average variation in this quality measure is of the same order of magnitude as the variation in price within each market and super-severity group. Also, perhaps not surprisingly, the estimates indicate that the variation in both price and the quality variable increases with the severity of illness [49].

This variance in our quality measure could be attributable to many things: patient preferences for hospital amenities, physician perceptions of clinical quality, and any other factors (other than price and distance) that affect referrals. We have separate work in progress to investigate which observable characteristics are most related to it. We can however make one connection which throws further light on our focus here: the relationship between capitation rates and the tradeoff between price, convenience and quality. Table 3 showed that older and sicker women had significantly higher probabilities of readmission within 12 months and of discharge to "other than home", and their

[49] Online Appendix 5 summarizes the variation in quality and in price faced by the typical patient across hospitals. The table records, for each super-severity group and each market where quality terms were estimated, the cross-patient average of the standard deviation in expected price (in \$000) and in quality (measured in the units described above) across hospitals in the choice set. The variance in both variables increases with severity of illness and the two are comparable in magnitude (for example for the least-sick super-severity the average standard deviation in quality in Los Angeles is \$1,136 while the average price standard deviation is \$1,278).

infants were also more likely to be discharged to "other than home". We calculated χ^2 test statistics for differences across plans in the probability of each adverse outcome conditional on each of our five super-severity groups. Not one of the forty test statistics (ten pairwise comparisons of insurers, for four outcomes each) was significant at the traditional 5% level.

9 Conclusions

The results of this paper indicate that the prices paid by insurers to hospitals for obstetric care: (i) affect allocations of patients across the hospitals in the network, and (ii) have an impact which is greater the more highly capitated the insurer. Our second major finding relates to the trade-offs made between price, quality and patient convenience. We find that, in the more highly-capitated insurers, price reductions are achieved by sending patients to relatively far-off hospitals. There is no evidence that quality of care, or health outcomes, suffer as a result of this behavior: the trade-off between quality and price is constant across plans while that between convenience and price is not. Our estimates summarize the preferences generated by a complicated decision-making process which involves both physician and patient choices. The data do not allow us to investigate the extent to which particular mechanisms drive the estimates, so we have to leave that question to future research.

At least in the context of obstetrics, these findings have obvious implications for the impact of the ongoing institutional changes in the health care sector. Most importantly the use of capitation payments in Accountable Care Organizations is likely to reduce costs and is unlikely to result in a reduction in quality of care. There are also several other possible implications. For example, as noted earlier, currently just under half of Accountable Care Organizations include a member hospital. There may be a benefit to such vertical integration through improvements in coordination of care. However, if providers favor within-ACO hospitals, vertical integration is likely to limit the cost reductions that would otherwise result from capitation. This is one of several tradeoffs that merit further investigation. Also the capitation incentives used to reimburse ACOs are supposed to condition on quality, and the right measure of quality is unclear. Our methods result in a measure of hospital quality that reflects patient and physician preferences. There is a question worth exploring about whether our quality measure would be helpful in this context.

References

1. Bajari P, Hong H, Park M and RJ Town. 2012. "Regression Discontinuity Designs with an Endogenous Forcing Variable and an Application to Contracting in Health Care." *Working paper.*

2. Baumgarten, A. 2004. "California Health Care Market Report 2004", prepared for the California HealthCare Foundation, http://www.chcf.org/topics/view.cfm?itemID=114640

3. Burns LR. and Wholey DR. 1992. "The Impact of Physician Characteristics in Conditional

Choice Models for Hospital Care". *Journal of Health Economics*, 11: 43-62.

4. Capps C., Dranove D. and Satterthwaite M. 2003. "Competition and Market Power in Option Demand Markets." *RAND Journal of Economics*. 34(4): 737-763.

5. Card D. and AG Krueger. 1994. "Minimum Wages and Employment: A Case Study of the Fast Food Industry in New Jersey and Pennsylvania." *American Economic Review.* 84(4): 774-5.

6. Chamberlain, G. 1980. "Analysis of Covariance with Qualitative Data." *Review of Economic Studies.* XLVII, 225-238.

7. Chandra A, Cutler DM and Z Song. 2012. "Who Ordered That? The Economics of Treatment Choices in Medical Care." In *Handbook of Health Economics Volume 2.* Eds M. Pauly, T. McGuire and Pedro P. Barros, North-Holland, 397-432.

8. Charlson ME, Pompei P, Ales KL, MacKenzie CR. "A new method of classifying prognostic comorbidity in longitudinal studies: development and validation." *J Chronic Dis.* 1987;40:373–83.

9. Cutler DM, McClellan M and JP Newhouse. 2000. "How Does Managed Care Do It?" *RAND Journal of Economics*, 31(3): 526-548.

10. Department of Health and Human Services. November 2012. "Accountable Care Organizations: What Providers Need to Know." *Medicare Learning Network, ICN 907406.*

11. Duggan M. 2000. "Hospital Ownership and Public Medical Spending". *Quarterly Journal of Economics*, 1343-1374.

12. Dranove D. and MA Satterthwaite. 2000. "The Industrial Organization of Health Care Markets". In: Schmalensee R., Willig RD. (eds.), *Handbook of Industrial Organization*, edition 3, volume 2, chapter 20, Elsevier, Amsterdam.

13. Gaynor M., Rebitzer JB. and Taylor LJ. 2001. "Physician Incentives in Health Maintenance Organizations." *Journal of Political Economy,* 112(4): 915-931.

14. Gaynor M. and Vogt WB. 2003. "Competition among Hospitals." *RAND Journal of Economics.* 34(4): 764-85.

15. Gaynor M. and Vogt WB. 2000. "Antitrust and competition in health care markets." in: A. J. Culyer & J. P. Newhouse (ed.), *Handbook of Health Economics*, edition 1, 1(27): 1405-1487 Elsevier.

16. Glied, S. 2000. "Managed Care". *Handbook of Health Economics*, in: A. J. Culyer & J. P. Newhouse (ed.), *Handbook of Health Economics*, edition 1, volume 1, chapter 13, pages 707-753 Elsevier.

17. Goldman D. and Romley JA. 2008. "Hospitals as Hotels: The Role of Patient Amenities in Hospital Demand." *NBER Working Paper Number 14619.*

18. Grumbach K., Coffman J., Vranizan K., Blick N. and E. O'Neil. 1998. "Independent Practice Association Physician Groups in California." *Health Affairs*, 17(3): 227-237.

19. Grumbach K., Osmond D., Vranizan K., Jaffe D. and A. Bindman. 1998. "Primary Care Physicians' Experience of Financial Incentives in Managed-Care Systems." *The New England Journal of Medicine*, 339(21): 1516-1521.

20. Ho, K. 2006. "The Welfare Effects of Restricted Hospital Choice in the U.S. Medical Care Market", *Journal of Applied Econometrics* 21(7): 1039-1079.

21. Ho K. and A. Pakes. 2011. "Do Physician Incentives Affect Hospital Choice? A Progress Report". Forthcoming in *International Journal of Industrial Organization.*

22. Kessler DP. and McClellan MB. 2000. "Is Hospital Competition Socially Wasteful?" *Quarterly Journal of Economics*, 115: 577-615.

23. Ketcham J, Leger PT and C Lucarelli. 2012. "Standardization Under Group Incentives." *Working paper.*

24. Limbrock F. 2011. "Pecuniary and Non-Pecuniary Incentives in Prescription Pharmaceuticals: The Case of Statins." *The B.E. Journal of Economic Analysis and Policy,* 1(2): 1.

25. Luft HS., Garnick DW., Mark DH., Peltzman DJ., Phibbs CS., Lichtenberg E. and McPhee SJ. 1990. "Does Quality Influence Choice of Hospital?" *The Journal of the American Medical Association*, 263(21): 2899-2906.

26. McClellan M. 2011. "Reforming Payments to Health Care Providers: The Key to Slowing Healthcare Cost Growth While Improving Quality?" *Journal of Economic Perspectives* 25(2): 69-92.

27. Melichar, l. 2009. "The effect of reimbursement on medical decision making: Do physicians alter treatment in response to a managed care incentive?" *The Journal of Health Economics*, 28: 902-907.

28. Muhlestein D. 2013. "Continued Growth of Public and Private Accountable Care Organizations." *Health Affairs Blog.* Available at http://healthaffairs.org/blog/2013/02/19/ continued_growth_of_public_and _private_accountable_care_organizations/.

29. Neyman J. and E.L. Scott. 1948. "Consistent Estimation from Partially Consistent Observations", *Econometrica* 16, 1-32.

30. Pakes A, 2010, "Alternative Models for Moment Inequalities¡. *Econometrica*, 78, 1783-1822.

31. Pakes A., Porter J., Ho K. and J. Ishii. 2011. "Moment Inequalities and Their Estimation". Harvard University working paper.

32. Robinson J. 2003. "Hospital Tiers in Health Insurance: Balancing Consumer Choice with Financial Incentives." *Health Affairs Web Exclusive* W3: 135-146.

33. Robinson J. and L. Casalino. 2001. "Reevaluation of Capitation Contracting in New York and California." *Health Affairs Web Exclusive*, W11-W19.

34. Rosenbaum P. and DB Rubin. 1983. "The Central Role of the Propensity Score in Observational Studies for Causal Effects." *Biometrika.* 70: 41-55.

35. Rosenthal M., Frank R., Buchanan J. and A. Epstein. 2001. "Scale and Structure of Capitated Physician Organizations in California." *Health Affairs*, 20(4): 109-119.

36. Rosenthal M., Frank R., Buchanan J. and A. Epstein. 2002. "Transmission of Financial Incentives to Physicians by Intermediary Organizations in California." *Health Affairs*, 21(4).

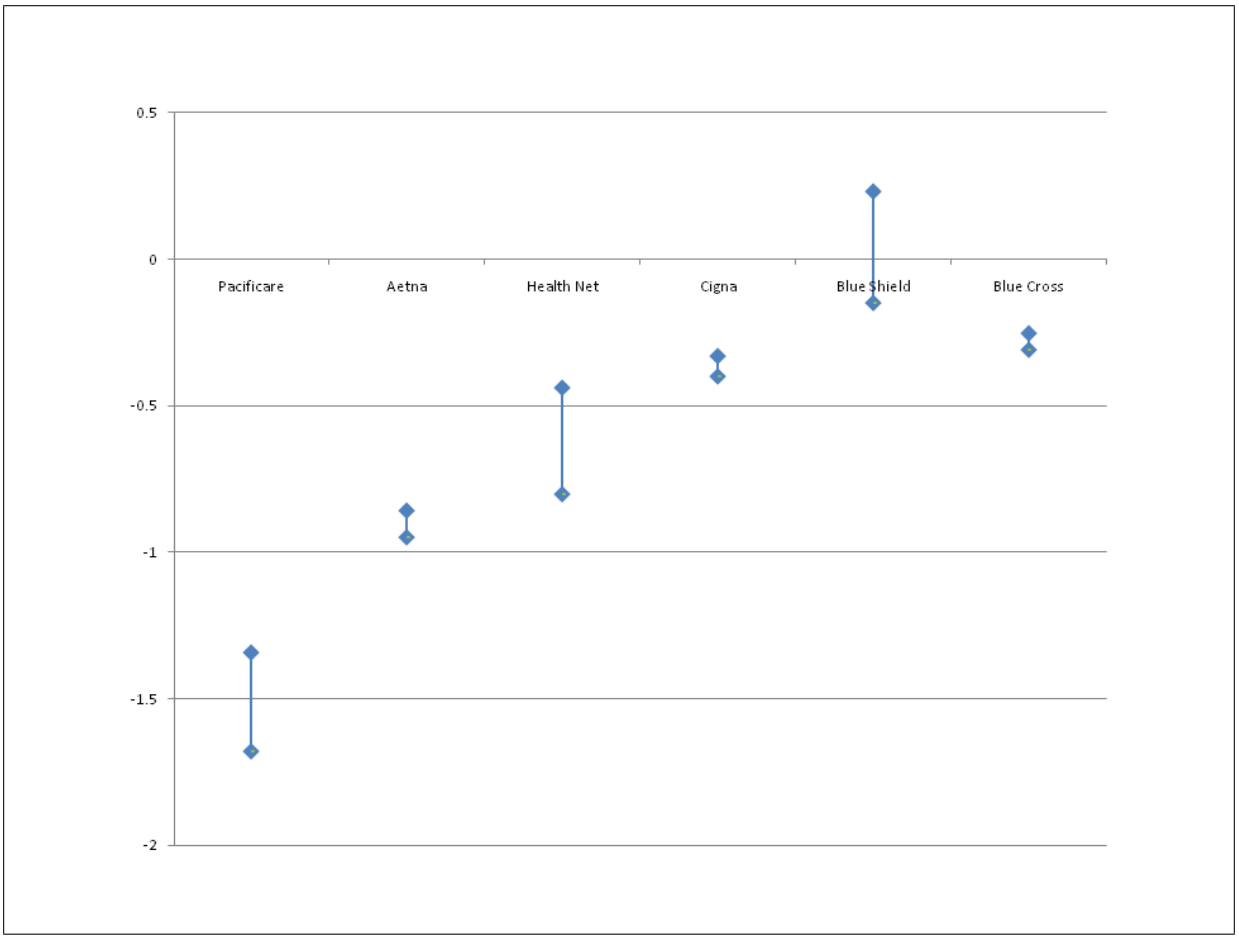
37. Song ZS, Safran DG, Landon BE, He Y, Ellis RP, Mechanic RE, Day MP and ME Chernew. 2011. "Health Care Spending and Quality in Year 1 of the Alternative Quality Contract." *The New England Journal of Medicine* , 365(10): 909-918.

38. Sorensen A. 2003. "Insurer-Hospital Bargaining: Negotiated Discounts in Post-Deregulation Connecticut." *Journal of Industrial Economics*, LI(4): 469-490.

39. Town R. and Vistnes G. 2001. "Hospital Competition in HMOs." *Journal of Health Economics* 20: 733-753.

40. Yegian J. 2003. "Tiered Hospital Networks: Reflections from the California HealthCare Foundation." *Health Affairs Web Exclusive* W3: 147-153.

Figure 1: Correlation of Estimated Price Coefficient with Insurer's Percent Capitation Payments



Notes: Graph to illustrate confidence intervals for insurer price coefficients, reported in Table 6. Estimates are from model where $p(.) = (1 - d_h)lp(c_i, h)$.

Figure 2: Graph of Constrained against Unconstrained Quality Estimates

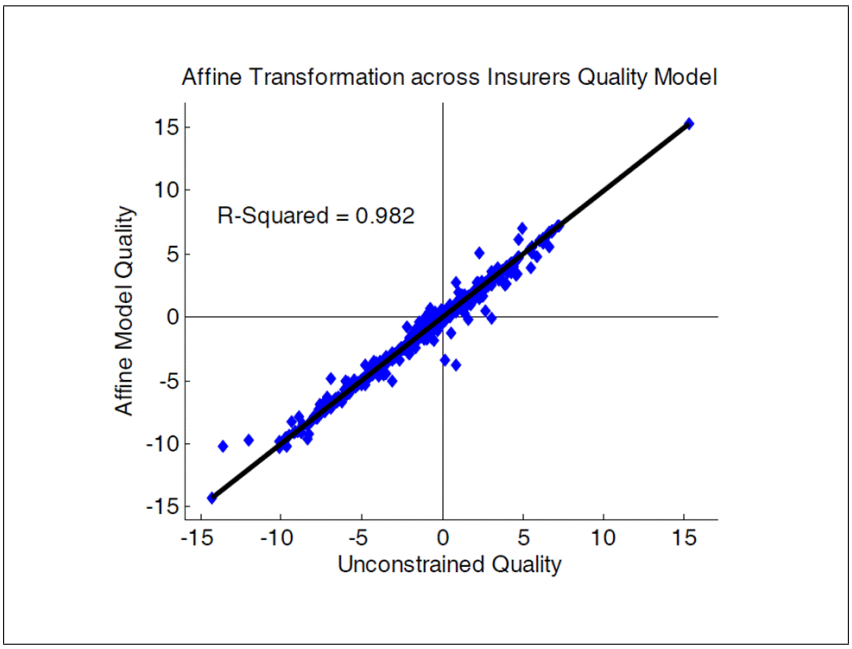
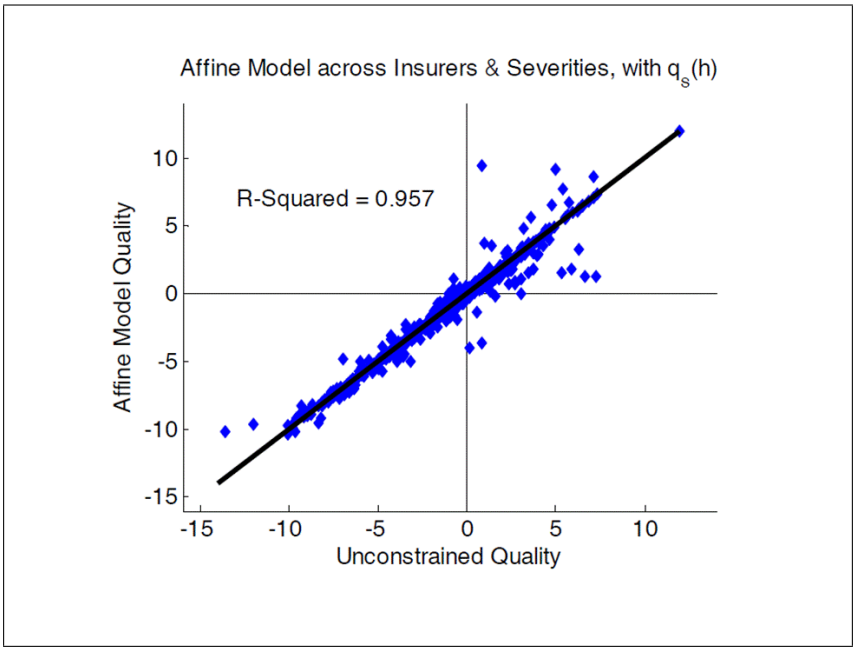


Figure 3: Adding a Constraint to the Quality Estimates



Notes: Figure 2 plots constrained quality estimates against unconstrained estimates, where unconstrained are $g_\pi(q_{h,}s)$ and constrained are defined as $\alpha^0_{\pi,m,s} + \alpha_{\pi,m,s} q_{h,s}$. Figure 3 repeats the exercise but defines constrained estimates as $\alpha^0_{\pi,m,s} + \alpha_\pi q_{h,s}$. See Section 8.1 for details.

Table 1: Summary Statistics by Insurer

	2002 enrollment			Birth	% Prim	Tax	Premium	Admin	Medical	Inpatient utilizn		Prescrip
	Commerc	Medicare	Medi-Cal	discharg	Capitn	Status	pmpm	expense	loss ratio	discha	days	drugs
Aetna	485,787	37,312	0	6,291	0.91	FP	152.42	19.33	86.2%	38.4	139.8	23.15
Blue Cross	3,486,358	251,299	1,099,044	25,038	0.38	FP	186.86	21.22	78.9%	38.4	142.4	20.92
Blue Shield	2,231,350	67,049	0	16,302	0.57	NFP	146.33	22.72	83.5%	50.3	176.4	20.51
Cigna	634,568	0	0	8,097	0.75	FP	-*	27.07	84.6%	39.8	137.1	15.63
Health Net	1,665,221	101,317	349,826	16,950	0.80	FP	184.92	18.60	86.3%	39.0	137.8	21.08
Pacificare	1,543,000	386,076	0	15,479	0.97	FP	149.92	24.51	88.4%	44.5	156.5	20.48
Kaiser	5,790,348	671,858	104,844	0		NFP	163.44	5.23	97.7%	49.1	158.1	0.44

Notes: Data on the six insurers included in our analysis and on Kaiser Permanente; the latter is excluded from our later analysis because the prices paid to hospitals are not reported. Source for all fields except Birth discharges and % primary capitation: Baumgarten (2004). 2002 enrollment provided separately for commercial plans, Medicare plans and Medi-Cal/Healthy Families plans. "Birth discharg" is the number of discharges in the data sample used in our analyses. "% Prim Capitn" is the percent of payments to primary providers made on a capitated basis in 2003 (source: State of California Department of Managed Health Care Annual Financial Reporting Forms, 2003). "Premium pmpm" is HMO commercial premium revenue per member per month (* information for Cigna was not included in the source data). "Admin expense" is per member per month administrative expenses for entire insurer in 2002, "Medical loss ratio" is medical and hospital expenses divided by premium revenues for entire insurer in 2002. Inpatient utilization and prescription drug data are for commercial plan only in 2002: "discha" is dicharges per 1000 members, "days" is acute days per 1000 members and "Prescrip drugs" is outpatient prescription drug expenses per member per month.

Table 2: Summary Statistics by Discharge

	Birth only	
	Mean	Std. Devn.
Number of patients	88,157	
Number of hospitals	195	
Number of insurers	6	
Hospitals per patient choice set	38	
Teaching hospital	0.27	
Distance to all hospitals (miles)	24.6	25.6
Distance to chosen hospital	6.7	10.3
List price	$13,312	$13,213
Discounted price	$4,317	$4,596
Length of stay	2.54	2.39
Died	0.01%	0.004%
Acute transfer	0.3%	0.02%
Skilled Nursing Transfer	1.5%	0.04%

Notes: Summary statistics for dataset comprising private enrollees of the six largest HMOs excluding Kaiser who are admitted for delivery-related diagnoses. "Discounted price" is list price*(1-discount). "Died" is the probability of death while in hospital, "Acute Transfer" the probability of transfer to an acute care setting (in this or a different hospital) and "Skilled Nursing Transfer" the probability of transfer to a skilled nursing facility (again at this or a different hospital). "Std Devn" for "Died", "Acute transfer" and "Skilled Nursing Transfer" are calculated under the assumption that the 0/1 variable is binomially distributed.

Table 3: Prices and Outcomes by Patient Type

		Mother			Infant		
	N	Price*(1-disc)	Readmission	Not Home	Price*(1-disc)	Readmission	Not Home
Overall	73118	4291 (4373)	2.39% (0.06%)	1.62% (0.05%)	2675 (18324)	9.42% (0.1%)	6.60% (0.1%)
Age							
< 40	71074	4259 (4329)	2.36% (0.1%)	1.60% (0.1%)	2627 (18103)	9.41% (0.1%)	6.50% (0.1%)
> 40	2044	5420 (5571)	3.52% (0.4%)	2.10% (0.3%)	4337 (24787)	9.64% (0.6%)	9.88% (0.7%)
Signif diff		0.000	0.000	0.038	0.000	0.365	0.000
Charlson							
0	71804	4256 (4265)	2.33% (0.1%)	1.58% (0.1%)	2619 (18101)	9.36% (0.1%)	6.53% (0.1%)
>0	1314	6227 (8135)	5.78% (0.6%)	3.42% (0.5%)	5735 (27745)	12.3% (0.9%)	10.5% (0.9%)
Signif diff		0.000	0.000	0.000	0.000	0000	0.000

Notes: Data taken from OSHPD Birth Cohort 2003 (a slightly different dataset that includes infant outcome variables). "Readmission" is percent of patients readmitted to hospital within 12 months of birth episode. "Not Home" is percent of patients discharged somewhere other than home; this includes transfer to acute care setting, transfer to skilled nursing facility, discharge against medical advice and death. Standard deviations in parentheses; for Readmission and Not Home we report standard errors which are calculated assuming that the 0/1 variables are binomially distributed. Charlson scores assign weights to comorbidities (known on admission to hospital) other than principal diagnosis where higher weight indicates higher severity. Value 0-6 are observed in the data. "Signif diff" states significance level at which we cannot reject the hypothesis that the means in the two samples are the same; these are the results of a t-test for price*(1-discount) and a z-test assuming two binomial distributions for Readmission and Not Home.

Table 4: Logit Analysis Results

	All birth patients	Least sick patients			Sickest patients		
Price	0.010** (0.002)	-0.017* (0.009)		0.069** (0.014)	0.012** (0.002)		0.028** (0.006)
Price interactions:							
% Capitated				-0.127** (0.016)			-0.025** (0.008)
Pacificare			-0.077** (0.012)			-0.006 (0.006)	
Aetna			-0.011 (0.016)			0.021** (0.008)	
Health Net			-0.038** (0.011)			0.007 (0.005)	
Cigna			-0.021 (0.014)			0.004 (0.007)	
Blue Shield			0.018 (0.011)			0.024** (0.004)	
Blue Cross			0.008 (0.011)			0.014** (0.003)	
Distance	-0.215** (0.001)	-0.215** (0.002)	-0.215** (0.002)	-0.215** (0.002)	-0.217** (0.002)	-0.216** (0.002)	-0.216** (0.002)
Distance squared	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)	0.001** (0.000)
$z_h x_i$ controls (15 coeffts)	Y	Y	Y	Y	Y	Y	Y
Hospital F.E.s (194 coeffts)	Y	Y	Y	Y	Y	Y	Y
N	88,157	43,742	43,742	43,742	44,059	44,059	44,059

Notes: N = # of patients. Least sick: aged 20-39 with zero Charlson scores and "routine" principal diagnoses and comorbidities (see Appendix 1). Sickest: all other patients. $z_h x_i$: interactions between hospital characteristics (teaching hospitals, for profit hospitals and hospitals that offer transplants, number of nurses per bed, a variable summarizing quality of delivery services provided) and patient characteristics (probabilities of death while in hospital, transfer to an acute care facility and transfer to a skilled nursing facility conditional on principal diagnosis, age category and Charlson score). Sample sizes for sickest and least sick populations do not sum to total N because a small number of hospitals with very few admissions were dropped when estimating on each subsample; patients choosing these hospitals were also dropped.

Table 5: Price Variance Across Aggregated Price and Severity Groups.

Number diags	Max rank 1			Max rank 2			Max rank 3		
of max rank	Pats	Price ($)	SD	Pats	Price ($)	SD	Pats	Price ($)	SD
1	23029	3431 (15)	1612	13128	4968 (42)	2476	1273	7448 (356)	4256
2	11757	4145 (28)	2180	4196	6019 (88)	2785	64	11536 (2337)	20370
3	4077	4682 (60)	2356	1274	7428 (212)	3609	8	12733 (4009)	11338
4	1179	5505 (149)	2590	380	8602 (462)	5283	1	25573 (-)	-
5	331	6189 (254)	3123	110	10186 (1002)	6084	0	-	
≥ 5	95	7663 (936)	4896	55	13365 (1596)	8880	0	-	
Total	40468	3857 (15)		19143	5488 (40)		1346	7687 (13065)	

Notes: Distribution of patients from inequalities sample who have a Charlson score of 0 across comorbidity ranks. "Pats" shows the number of patients in each "max rank" group and each "number diags of max rank" group. Here "Max rank j" means the maximum rank of a comorbidity for this patient, as defined by obstetrical experts at Columbia Presbyterian Hospital, is j. "Number diags of max rank" groups patients according to the number of comorbidities in their discharge record with the relevant max rank. Patients in different rows of a particular column of the table will have different price groups. "Price ($)" is the average observed price*(1-discount) for patients in this group; standard errors in parentheses. "SD" is the cross-hospital standard deviation of the mean observed price*(1-discount) in this hospital for the patients in this group.

Table 6: Results of Inequalities Analysis

			Column 1			Column 2			Column 3			Column 4		
		percent capitated	$p(.) = (1 - d_h)lp(c_i, h)$			$p(.) = (1 - \hat{d}^1_{\pi,h})lp(c_i, h)$			$p(.) = (1 - \hat{d}^2_{\pi,h})lp(c_i, h)$			$\tilde{d}(i_h, h, h')_+$		
			$\hat{\theta}$	$[CI_{LB},$	$CI_{UB}]$	$\hat{\theta}$	$[CI_{LB},$	$CI_{UB}]$	$\hat{\theta}$	$[CI_{LB},$	$CI_{UB}]$	$\hat{\theta}$	$[CI_{LB},$	$CI_{UB}]$
Pacificare		0.97	-1.50**	[-1.68,	-1.34]	-1.07**	[-1.52,	-0.62]	-1.47**	[-1.64,	-1.32]	-1.42**	[-1.66,	-0.94]
Aetna		0.91	-0.92**	[-0.95,	-0.86]	-0.68**	[-0.72,	-0.62]	-0.77**	[-0.81,	-0.71]	-0.73**	[-0.77,	-0.66]
Health Net	Baseline	0.80	-0.17**	[-0.27,	-0.13]	-0.11**	[-0.23,	-0.07]	-0.20**	[-0.30,	-0.17]	-0.60**	[-0.71,	-0.51]
	Drop t $\leq$ -2		-0.78**	[-0.80,	-0.44]	-0.41	[-0.43,	0.94]	-1.87**	[-1.89,	-1.33]	-2.16**	[-2.26,	-1.13]
Cigna		0.75	-0.35**	[-0.40,	-0.33]	-0.35**	[-0.39,	-0.33]	-0.32**	[-0.36,	-0.30]	-0.66**	[-0.71,	-0.58]
Blue Shield		0.57	-0.06	[-0.15,	0.23]	0.18	[-0.16,	0.79]	0.004	[-0.28,	0.70]	set	[-0.95,	0.65]
Blue Cross	Baseline	0.38	-0.10**	[-0.24,	-0.01]	-0.03	[-0.18,	0.39]	-0.09**	[-0.22,	-0.01]	-0.27**	[-0.31,	-0.25]
	Drop t $\leq$ -2		-0.29**	[-0.31,	-0.25]	-0.12**	[-0.14,	-0.05]	-0.18**	[-0.21,	-0.14]	-0.31**	[-0.34,	-0.27]

Notes: Results of inequalities analysis. We include 157 hospitals in total. Estimated coefficient is the ratio of the price coefficient to the distance coefficient in the utility equation, where prices are measured in $000 and distance in tens of miles. Three price measures are used; they are calculated using the observed average hospital discount, and the two estimated hospital-insurer level discounts discussed in Section 7.3, respectively. Specification includes four distance-based instruments (positive and negative parts of $d(i_h, h) - d(i_h, h')$ for each patient) plus a constant in the instrument set. In Column 4 these instruments are replaced with indicators for distance differences being greater than 3 miles (and used the price measure from Column 1). The rows labeled "drop t $\leq$ -2" report results when we dropped moments whose t-statistic values were less than -2 (2 out of 182 for Health Net; 5 out of 285 for Blue Cross) and repeated the estimation process. "set" for Blue Shield in Column 4 indicates a range of values.

Table 7: Summary of t-statistics from Inequalities Analysis

	Pacificare	Aetna	Health Net	Cigna	Blue Shield	Blue Cross
Summary of t-statistics						
Number positive	152	75	173	93	170	254
Ave value of positive	12.7	22.5	17.1	19.5	19.1	21.5
Number negative	11	3	9	2	4	31
Number t < -2	0	0	2	0	0	5

Notes: Summary of estimated t-statistics of the moments used in inequalities analysis. T-statistic = value of the moment at the estimated $\theta_{\pi,p}$ (for specification where $p(.) = (1 - d_h)lp(c_i, h)$). Under the model all moments should be non-negative.

Table 8: Magnitudes of Logit and Inequality Results

	percent capitated	Logits (less-sick patients) average η_i	Inequalities (all patients) average η_i
Pacificare	0.97	0.33	11.08
Aetna	0.91	0.10	11.47
Health Net	0.80	0.15	6.52
Cigna	0.75	0.10	2.49
Blue Shield	0.57	-0.08	0.51
Blue Cross	0.38	-0.03	3.24

Notes: Estimated cross-patient average value of $\eta_i = \frac{\partial d_i}{\partial p_i}\frac{p_i}{d_i}$ for each insurer implied by logit and inequality analyses. Logit model uses less-sick population as defined in notes to Table 4. Inequality model uses price defined using discount δ_h (Column 1 of Table 6).

Table 9: Trade-offs Aggregated Over Markets and Severities.

Insurer	P-care	Aetna	HNet	Cigna	BC
% cap	0.97	0.91	0.80	0.75	0.38
$\theta_{p,\pi}$	-1.50	-0.92	-0.78	-0.35	-0.29
α_π	5.13	3.12	2.63	1.20	1.00
$\theta_{p,\pi}/\alpha_\pi$	-0.293	-0.295	-0.297	-0.291	-0.290
$1/\alpha_\pi$	0.20	0.32	0.38	0.83	1.00
Upper and Lower Bounds on C.I. $\theta_{p,\pi}/\alpha^*_\pi$					
Lower	-0.38	-0.36	-0.35	-0.40	-0.31
Upper	-0.23	-0.23	-0.15	-0.22	-0.25

*Calculated as lower bound (upper bound) θ_p divided by upper bound (lower bound) α_π.

Table 10: Cost-Quality Trade-offs By Market and Severity*

Insurer	P-care	Aetna	HNet	Cigna	BC
LA S1	-0.29	-0.29	-0.29	-0.29	-0.29
LA S2	-0.33	-0.31	-0.31	-0.39	-0.29
LA S3	-0.28	n/a	-0.30	-0.29	-0.29
LA S4	-0.31	-0.30	-0.32	-0.34	-0.29
LA S5	-0.29	n/a	n/a	n/a	-0.29
Bay S1	-0.34	-0.32	-0.32	-0.30	-0.29
Bay S2	-0.44	-0.89	-0.48	-0.43	-0.29
Bay S3	-0.37	-0.35	-0.39	-0.33	-0.29
Bay S4	-0.38	-0.31	-0.29	-0.33	-0.29
Bay S5	n/a	n/a	-0.19	-0.35	-0.29
Ora S1	n/a	-0.28	n/a	-0.28	-0.290
Ora S2	n/a	-0.21	n/a	-0.31	-0.290
Ora S3	n/a	-0.24	n/a	-0.30	-0.290
Ora S4	n/a	-0.41	n/a	-0.40	-0.290
Ora S5	n/a	n/a	n/a	-0.69	-0.290
SD S1	-0.42	-0.55	-0.41	-0.26	-0.290
SD S2	-2.07	-0.88	-0.50	-107.9	-0.290
SD S3	-0.40	-0.43	-0.51	-0.31	-0.290
SD S4	-0.28	-0.23	-0.28	-0.73	-0.290
SD S5	-1.31	n/a	-56.5	n/a	-0.290
IE S1	-2.02	-0.68	-0.81	-0.46	-0.290
IE S2	-1.37	n/a	-0.78	n/a	-0.290
IE S3	-1.02	n/a	-0.50	n/a	-0.290
IE S4	-0.52	n/a	-0.67	n/a	-0.290
IE S5	n/a	n/a	n/a	n/a	n/a

* LA=Los Angeles, Bay=Bay Area, Ora=Orange County, SD=San Diego,IE=Inland Empire.

Appendices For Online Publication

Appendix 1: Details on the Market and the Data

This paper focuses on hospital referrals for pregnant women who are enrolled in private HMO plans in California. The referring physician is an obstetrician who is often a member of a large physician group. There are two types of physician groups: medical groups and Independent Practice Associations (IPAs). On average they each cover 50,000 lives and contain between 200-300 physicians per group. Approximately two-thirds of patients covered by non-Kaiser physician organizations are in IPAs and one-third are in medical groups (see Rosenthal et al (2001)). Physicians in medical groups are either employees or partners of the group. IPAs are administrative organizations that contract with independent physicians or clinics and sign network contracts with health plans on behalf of their physicians. They exist primarily to negotiate and manage capitation contracts for their member physicians. As discussed in the paper, capitation contracts generate incentives at the physician group level to utilize low-cost hospitals.

If capitation arrangements are to influence hospital referral choices, however, cost-control incentives must be passed from the physician group to the individual physician. The connection is clear when the physician is a partner in a medical group since his or her own income is directly linked to the group's profitability but less clear for other physicians. Rosenthal et al (2002) consider this issue, tracking the flow of financial incentives from physician organizations to physicians in California. They find that the majority of physician groups receiving capitation payments pass financial risk on to individual physicians, in the form of either capitation-based compensation, cost-of-care bonuses or profit sharing.[50]

Our model assumes that hospitals are reimbursed on a fee-for-service basis. In reality different insurers may use different payment mechanisms to reimburse different hospitals in their networks. The major possibility, in addition to fee-for-service payments, is a per-diem payment arrangement under which the hospital receives a fixed number of dollars per day of inpatient stay. We have some information at the hospital and insurer level on the payment mechanisms used but this information is not provided at the discharge level.[51] The weighted average percent of payments that are made on a per-diem basis (where the weight is the number of enrollees in the plan) is fairly low at 21%. Two of the six carriers in our data, Aetna and Health Net, report no per-diem payments in 2003. Still, there is clearly some variation in the data in terms of payment mechanisms which will generate measurement error in our price variable.

[50] Grumbach et al (1998a) survey California IPAs and have similar findings. They also note that IPAs that are paid on a fee-for-service basis make fee-for-service payments to their member physicians.

[51] Case-based or D.R.G. payments are also possible: our data do not distinguish between them and fee-for-service payments but we expect case-based payments to be less common since they are predominately used by Medicare rather than private payors. Capitation payments to hospitals are possible but uncommon. Only 24 of the 195 hospitals in the full dataset (e.g. in Table 2) have over 5% of revenues from capitation. 104 report zero capitation payments. Our logit analysis includes all hospitals, including those that receive capitation payments. In a robustness test we redefine price to be price*(1-percent of revenues received on a capitated basis). The results are very similar to those from the baseline logit analysis. The inequalities analysis excludes the hospitals reporting that more than 5% of their revenues are paid on a capitation basis.

We note in Section 4 that our dataset does not precisely identify HMO enrollees for every insurer. Instead it groups together all Knox Keene enrollees for a particular insurer, defined as enrollees in plans that are overseen by the California Department of Managed Health Care (DMHC) and subject to the Knox Keene Act. All California HMOs are Knox Keene plans. In addition, Blue Cross and Blue Shield PPO products were Knox Keene plans in 2003, the year of our data. 63% of Blue Shield's Knox Keene enrollees, and 72% of those for Blue Cross, were in the PPO rather than HMO product. We cannot distinguish between PPO and HMO enrollees for these two insurers at the individual discharge level. Capitation rates are also reported for the full Knox Keene plan. This likely generates some of the cross-insurer variation in capitation rates in the data: PPOs usually pay their physicians on a fee-for-service basis, unlike HMOs, consistent with Blue Shield and Blue Cross having the lowest capitation payment rates in our data.[52] Provided we control for other differences between HMO and PPO plan types this is not a problem: in fact it provides helpful variation to assist us in identifying the effect of capitation on physician behavior. We note that PPOs use the same mechanism for hospital referrals as HMOs except that patients have more discretion: by paying a relatively high out-of-pocket price they can choose to visit an out-of-network hospital or physician. Pricing policies can also be different. While an HMO enrollee probably pays the same small copay whatever hospital she chooses, approximately 15% of PPO enrollees pay a coinsurance rate (a fixed percentage of the total price) that is lower if they choose an in-network hospital than if they go outside.[53] We drop hospitals to which very few patients are admitted for these two insurers, expecting thereby to remove out-of-network hospitals from the data.[54] Any remaining difference in pricing strategies for PPO plans biases our estimates towards finding no difference in price coefficients between high- and low-capitation insurers, since patients presumably have a higher sensitivity to price than do physicians and our model conflates the price coefficients of patients and physicians for Blue Cross and Blue Shield.

We make several assumptions to define hospital prices for the logit analysis that are not needed for the inequality analysis. If discount information is missing we fill it in for the logit analysis using regression analysis. (These observations are excluded from the inequalities analysis.) For approximately 5% of the hospitals in the sample we do not observe the discount for the calendar year but do observe discount data for both relevant fiscal years (from the annual financial statements; fiscal years vary across hospitals). We fill in the missing calendar year information using the predictions from a regression of calendar year discounts on fiscal year discounts and hospital characteristics (fixed effects for hospital systems, service type, control type, Hospital Referral Region, teaching

[52] However this is not the only reason for variation in the percent capitation variable across insurers. Interviews with officials at the DMHC indicate that not all PPO plans are exclusively fee-for-service and not all HMOs in California are exclusively capitated.

[53] The Kaiser Family Foundation Employer Health Benefits Survey 2003 shows that the difference in pricing strategies was not large in that year. 14% of covered workers in a PPO plan paid a coinsurance rate, 26% paid a dollar copay and 59% paid neither. In contrast 5% of HMO enrollees paid a coinsurance rate and 49% paid a copay.

[54] The inequalities analysis drops hospitals with fewer than 150 switches with other hospitals in the data. This implies dropping approximately 10% of hospitals for each of Blue Shield and Blue Cross. Given that on average 83% of the hospitals in the market are included in each insurer's network (Ho (2006)), this is likely to be sufficient to exclude out-of-network hospitals.

hospitals and particular services provided and lagged numbers of doctors and beds, all as reported in the American Hospital Association data for 2003). The R^2 of the regression is 0.61. A few other hospitals have missing discount data for the relevant fiscal years and the calendar year; in this case we use the predictions of a regression of calendar year discounts on hospital characteristics which has a R^2 of 0.49.

In addition, for the logit analysis, if the set of patients to be used to determine a patient's price in a particular hospital is empty, we expand the group of "similar" patients to include women in the same age category and with the same Charlson score and principal diagnosis. If this is also empty we expand it to include all same-age category same-principal diagnosis patients, then all same-principal diagnosis women. If this group is also empty we take the mean of the non-missing prices already calculated for the particular patient. (This is not an issue for the inequality analysis: we only compare hospitals where prices can be calculated for both switching patients.)

Appendix 2: Estimation of the Discount Variation Across Insurers

This appendix provides details of the method discussed in Section 7.3 that was used to estimate the variation in discounts across insurers. We begin with the average negotiated discount at the hospital level, d_h.[55] This is a weighted average of the discounts for both inpatient and outpatient services to both Knox Keene and Point of Service (POS) insurers. We assume for the moment that the discount at the hospital-insurer level, $d_{\pi,h}$, does not differ across diagnoses for a given (π, h) pair; we relax this assumption in the following section. We use data from the OSHPD hospital discharge and financial records for 2003 that are not used in the main analysis. First, we have discharge data covering all Knox Keene inpatient events in the year 2003, including diagnoses other than delivery and births. We observe a list price for every discharge. Second, the hospital financial reports include data on hospital h's total charges (sum of list prices) for managed care (Knox Keene and POS) inpatient services and separately for managed care outpatient services.

If $s_{\pi,h}$ ($s^o_{\pi,h}$) is the share of Knox Keene π's inpatient (outpatient plus POS inpatient) charges in hospital h we know that:

$$d_h = \sum_\pi s_{\pi,h} d_{\pi,h} + \sum_\pi s^o_{\pi,h} d^o_{\pi,h} \tag{19}$$

where $\sum_\pi (s_{\pi,h} + s^o_{\pi,h}) = 1$. We are constrained by lack of data on $s^o_{\pi,h}$. We therefore assume that $d^o_{\pi,h} = d_{\pi,h}$. We can always write $s^o_{\pi,h} = s_h s_{\pi,h} + e_{\pi,h}$ where $s_h \equiv \sum_\pi s^o_{\pi,h} / \sum_\pi s_{\pi,h}$, and can be calculated from the observed data, and $\sum_\pi e_{\pi,h} = 0$. Substituting we have:

$$d_h = \sum_\pi (1 + s_h)\, s_{\pi,h} d_{\pi,h} + \tilde{e}_h \tag{20}$$

where $\tilde{e}_h = \sum_\pi e_{\pi,h} d_{\pi,h}$.

To proceed we need a specification for HMO inpatient discounts at different hospitals. We begin by writing

$$d_{\pi,h} = d_0 + \tilde{d}_h + \tilde{d}_{\pi,h}$$

where $\forall h$, $\sum_\pi \tilde{d}_{\pi,h} = 0$, so that $d_0 + d_h$ is the mean hospital discount, and $\sum_h \tilde{d}_h = 0$ so that d_0 is the mean of the (mean) hospital discount (across hospitals). Our reduced form model for the mean hospital discount is

$$\tilde{d}_h = \left(\frac{\exp(X_{h,m}\beta^h)}{1 + \exp(X_{h,m}\beta^h)} - d_0 \right) + v_h \equiv \Big(f(X_{h,m}, \beta^h) - d_0 \Big) + v_h \tag{21}$$

where $X_{h,m}$ are hospital characteristics or their interactions with market characteristics and v_h is mean independent of $X_{h,m}$. The reduced form model for an insurer's deviation from the mean

[55] We conduct this analysis using the discount d_h rather than one minus the discount, which is defined above as $\delta^o_h = 1 - d_h$.

discount is

$$\tilde{d}_{\pi,h} = \frac{\exp(X_{\pi,h,m}\beta^{\pi}) - \frac{1}{N_{\pi,h}}\sum_{\pi}\exp(X_{\pi,h,m}\beta^{\pi})}{\frac{1}{N_{\pi,h}}\sum_{\pi}\exp(X_{\pi,h,m}\beta^{\pi})} + v_{\pi,h} \equiv f(X_{\pi,h,m}\beta^{\pi}) + v_{\pi,h} \tag{22}$$

where $X_{\pi,h,m}$ are insurer characteristics and their interactions with market and hospital characteristics and $N_{\pi,h}$ is the number of insurers contracting with hospital h and where $v_{\pi,h}$ is mean independent of $X_{\pi,h,m}$ and $\forall h$, $\sum_{\pi} v_{\pi,h} = 0$ (since $\sum_{\pi} \tilde{d}_{\pi,h} = 0$).

Substituting these specifications into equation (20) generates the following equation which can be estimated using nonlinear least squares:

$$d_h = f(X_{h,m}, \beta^h) + \sum_{\pi} (1 + s_h)\, s_{\pi,h} f(X_{\pi,h,m}\beta^{\pi}) + e_h \tag{23}$$

where $e_h = \sum_{\pi} (1 + s_h)\, s_{\pi,h} v_{\pi,h} + v_h + \tilde{e}_h$.

The estimates, set out in Tables 1 and 2 of this Appendix, are intuitive. Table 1 sets out the results when $X_{h,m}$ includes both hospital characteristics and market fixed effects. Model 1 includes insurer fixed effects; in Model 2 we collapse these into a fixed effect for high-capitation insurers (Pacificare together with Aetna, Health Net and Cigna), a fixed effect for Blue Cross and a continuous variable defined as the insurer's share of HMO enrollment in California.[56] In both cases we find that for profit hospitals and hospitals that are members of systems (groups of providers that bargain jointly with insurers) have significantly higher discounts than other hospitals. At first sight this is surprising since a higher discount implies a lower price paid to the hospital. However, this is likely to be explained by the substantial variation in list prices across hospitals. We show in Table 6 of Ho and Pakes (2011) that for profit hospitals have higher prices net of discounts than not-for-profit hospitals. If we add an indicator for hospitals in systems to the regression we find that system hospitals, too, have significantly higher prices than other hospitals.[57] These results indicate that, while discounts are high for system and for profit hospitals, list prices are higher, so that the net price paid conditional on severity is also relatively high for these providers.

Other hospital characteristics such as indicators for teaching hospitals and hospitals that provide transplants (a measure of high-tech hospitals) are not significant in our analyses. The coefficient on a variable measuring the hospital's share of beds in the market, a potential measure of hospital bargaining power, is negative as expected but not significant at p=0.05. The insurer fixed effects in Model 1 are all statistically insignificant and the magnitudes demonstrate no particular correlation between insurer capitation levels and discounts. In Model 2 the coefficient for high-capitation insurers is slightly negative, and that for Blue Cross is somewhat more negative compared to the

[56] We use the share of enrollment at the state level rather than the market level to help avoid endogeneity problems due to insurers with high discounts in a particular market attracting high enrollment in that market.

[57] The analysis controls for patient severity by using as a price measure the price ratio $p_i^{ratio} = \frac{p_i}{\bar{p}_{s_i}}$ where p_i is the price (list price multiplied by δ_h) for patient i and $\bar{p}_{s_i}$ is the average price for same-severity patients across all hospitals in the sample. The results of these regressions are excluded from this paper to conserve space. They are available from the authors on request.

excluded plan (Blue Shield) although neither coefficient is significant at p=0.05. The coefficient on HMO market share is positive (although again insignificant), consistent with a bargaining power story. We use the results in Model 2 to calculate the predicted $\hat{\delta}_{\pi,h}$ that are used in the inequalities analysis since they provide a somewhat smoother prediction of the variation in discounts across insurers than the results in Model 1. The hypothesis that Model 2 fits the data as well as Model 1 cannot be rejected in an F-test of size 0.05.[58]

In Table 2 we replace the market fixed effects with market characteristics. We view this as an exploratory exercise to check that our results are consistent with the previous literature on the impact of hospital and insurer concentration on prices. Our results are similar to those in previous papers: we find that variables likely to be positively correlated with hospital bargaining power are negatively related to hospital discounts, while those positively related to insurer bargaining power are positively correlated with discounts. For example, in Model 3 we find that when market fixed effects are removed the positive coefficient on the insurer market share variable and the negative coefficient on hospital market share both become significant at p=0.05. Models 4-5 demonstrate that discounts are significantly higher in markets with more hospitals per thousand population and lower in markets with more insurers per 1000 population.

The final step is to use these estimates to generate a prediction for $d_{\pi,h}$. There are two possibilities. First, since:

$$d_{\pi,h} \approx f(X_{h,m}, \hat{\beta}^h) + f(X_{\pi,h,m}\hat{\beta}^\pi) + (v_{\pi,h} + v_h)$$

we define

$$\hat{d}^1_{\pi,h} = f(X_{h,m}, \hat{\beta}^h) + f(X_{\pi,h,m}\hat{\beta}^\pi) \tag{24}$$

and incur the error $e^1_{\pi,h} = v_{\pi,h} + v_h$. Second, since

$$d_{\pi,h} \approx d_h - \sum_\pi (1+s_h)\, s_{\pi,h} f(X_{\pi,h,m}\hat{\beta}^\pi) + f(X_{\pi,h,m}\hat{\beta}^\pi) + \left(v_{\pi,h} - \tilde{e}_h - \sum_\pi (1+s_h)\, s_{\pi,h} v_{\pi,h} \right)$$

we define

$$\hat{d}^2_{\pi,h} = d_h - \sum_\pi (1+s_h)\, s_{\pi,h} f(X_{\pi,h,m}\hat{\beta}^\pi) + f(X_{\pi,h,m}\hat{\beta}^\pi) \tag{25}$$

and incur the error $e^2_{\pi,h} = v_{\pi,h} - \tilde{e}_h - \sum_\pi (1+s_h)\, s_{\pi,h} v_{\pi,h}$. We use the predictions to define price measures $p^1(.) = (1-\hat{d}^1_{\pi,h}) lp^o(c_i,h)$ and $p^2(.) = (1-\hat{d}^2_{\pi,h}) lp^o(c_i,h)$ and use these in the inequalities analysis. The errors $(1-e^1_{\pi,h}) lp^o(c_i,h)$ and $(1-e^2_{\pi,h}) lp^o(c_i,h)$, together with estimation error from this step and measurement error from the expected list price calculation, will be inputs into the error term $\varepsilon_{i_h,\pi,h}$ defined in Section 7.2.

[58] We also estimated the inequalities analysis using the discounts predicted by Model 1; the results were very similar to the main analyses reported in Table 6.

While use of $p^1(.)$ and/or $p^2(.)$ as our price variable mitigates the problems that could arise from using a price variable that does not account for insurer-specific discounts, it probably does not eliminate them. To the extent that doctors know $\nu_{\pi,h}$ and select hospitals based on its value there will still be a selection bias in both of these price variables[59] , and if doctors know ν_h and select based on its value there will be an additional source of selection bias in $p^1(\cdot)$[60].

Allowing Discounts to Differ Across Diagnoses

As an additional robustness test we modify the analysis above to allow discounts for deliveries to differ from those for other diagnoses, and to allow the extent to which they differ to depend on hospital characteristics:

$$d_{\pi,h} = (1 - s^{birth}_{\pi,h}) d^{non-birth}_{\pi,h} + s^{birth}_{\pi,h} \gamma_h d^{non-birth}_{\pi,h}.$$

We parametrize $d^{non-birth}_{\pi,h}$ using the same reduced form expressions and explanatory variables as those used for $d_{\pi,h}$ above. We write γ_h as a linear combination of a constant, an indicator for teaching hospitals, the hospital's share of beds in the market and a variable summarizing the quality of delivery services offered by the hospital, . Under the assumptions that (a) there are no births in outpatient units, (b) the share of π's POS inpatient charges in hospital h that are births is the same as its share of Knox Keene inpatient charges that are births, and (c) the share of outpatient plus POS inpatient charges that are outpatient is the same for all insurers, we can derive an estimating equation similar to equation (23). The results are set out in Appendix 2 Table 3. In Model 1 we include just a constant term in γ_h. Its coefficient is statistically significant with a magnitude of 1.06, indicating that on average discounts for deliveries are 6 percent higher than those for other diagnoses. The coefficients on the hospital and insurer characteristics in $d^{non-birth}_{\pi,h}$ differ very little from the baseline specification (Model 2 of Appendix 2 Table 1). Models 2, 3 and 4 include different combinations of hospital characteristics in γ_h; none of these have significant coefficients at p=0.05.

We complete the robustness test by recomputing the prediction for $d_{\pi,h}$ using the estimates in Model 1 of Appendix 2 Table 3 and repeating the inequalities analysis using this prediction to generate the price variable. The results differ very little from those reported in Table 6.

[59] Only the component of $(1 - e_{\pi,h}) lp^o(c_i, h)$ that differs across c_i groups within a hospital-severity pair will be absorbed into the error term rather than into $g_\pi(.)$. However, the interaction with the list price implies that there will be some such variation and if decision-makers observe it this will cause endogeneity bias. We assume that $\tilde{e}_h$ is unrelated to discounts and therefore not problematic here.

[60] We did investigate the magnitude of the errors through a regression analysis. Note from equation (23) that

$$H^{-1} \sum_h e_h^2 \rightarrow_P \sigma^2_{\tilde{e}} + \sigma^2_h + \sum_\pi (1 + s_h)^2 s^2_{\pi,h} \sigma^2_{\pi,h}$$

where $\sigma^2_{\tilde{e}}$ is the variance of $\tilde{e}_h$ and similarly for σ^2_h and $\sigma^2_{\pi,h}$. We regress e^2_h on a constant term and $\sum_\pi (1 + s_h)^2 s^2_{\pi,h}$ and estimate a constant term of 0.0037 (standard error 0.0034) and an estimate of the coefficient on the X variable of 0.0286 (standard error 0.0107). We compare these numbers to the variance in d_h, a lower bound on the unobserved variance in $d_{\pi,h}$, which is 0.022. We conclude that the variance in $v_{\pi,h}$ is likely larger in magnitude that that of v_h.

Appendix 2, Table 1: NLLS Analysis of Discount Variation

	percent capitated	Model 1 Coefft	(S.E.)	Model 2 Coefft	(S.E.)
Hospital Characteristics					
Constant		-0.07	(0.30)	-0.14	(0.29)
Teaching hospital		-0.03	(0.11)	-0.06	(0.11)
Cost per admission		-0.01	(0.01)	-0.01	(0.01)
For profit		0.44**	(0.12)	0.43**	(0.12)
Offers transplants		-0.05	(0.17)	-0.03	(0.17)
System hospital		0.26**	(0.11)	0.26**	(0.12)
Share of beds in mkt		-12.32	(7.83)	-11.46	(8.16)
Insurer Characteristics					
Pcare/Aetna/HN/Cigna				-0.11	(0.07)
Pacificare	0.97	-0.04	(0.13)		
Aetna	0.91	0.09	(0.20)		
Health Net	0.80	0.12	(0.15)		
Cigna	0.75	-0.42	(0.23)		
Blue Shield	0.57	0.11	(0.15)		
Blue Cross	0.38	0.00	(0.12)	-0.36	(0.22)
Share in CA				1.77	(1.32)
Market FEs?		Yes		Yes	
pseudo-R^2		0.46		0.45	
Number hospitals		144		144	

Notes: NLLS analysis of variation in hospital discounts d_h across hospitals, insurers and markets. Equation for estimation is $d_h = f(X_{h,m}, \beta^h) + \sum_\pi (1 + s_h)\, s_{\pi,h} f(X_{\pi,h,m}\beta^\pi) + e_h$ where $f(X_{h,m}, \beta^h) = \frac{\exp(X_{h,m}\beta^h)}{1+\exp(X_{h,m}\beta^h)}$ and $f(X_{\pi,h,m}\beta^\pi) = \frac{\exp(X_{\pi,h,m}\beta^\pi) - \frac{1}{N_{\pi,h}}\sum_\pi \exp(X_{\pi,h,m}\beta^\pi)}{\frac{1}{N_{\pi,h}}\sum_\pi \exp(X_{\pi,h,m}\beta^\pi)}$. "Cost per admission" is average hospital cost per admission in $000. "Share in CA" is insurer's share of HMO enrollment in California. pseudo-R^2 is 1 - (SSR from full model / SSR from model including only a constant). ** = significant at p=0.05; *=significant at p=0.10.

Appendix 2, Table 2: NLLS Analysis of Discount Variation: Market Characteristics

	Model 3		Model 4		Model 5	
	Coefft	(S.E.)	Coefft	(S.E.)	Coefft	(S.E.)
Hospital Characteristics						
Constant	0.54**	(0.20)	0.13	(0.30)	-0.26	(0.32)
Teaching hospital	0.05	(0.09)	0.08	(0.09)	0.01	(0.10)
Cost per admission	-0.03**	(0.01)	-0.02**	(0.01)	-0.02**	(0.01)
For profit	0.50**	(0.12)	0.53**	(0.12)	0.52**	(0.11)
Offers transplants	0.03	(0.15)	0.03	(0.15)	-0.01	(0.15)
System hospital	0.20**	(0.12)	0.20**	(0.12)	0.21**	(0.12)
Share of beds in mkt	-10.17**	(4.55)	-13.87**	(4.69)	-7.56	(6.22)
Market Characteristics						
Hosps per 1000 pop			69.06**	(39.60)	172.39**	(53.61)
Plans per 1000 popln					-81.83**	(36.32)
Insurer Characteristics						
Pcare/Aetna/HN/Cigna	-0.11	(0.07)	-0.07	(0.07)	-0.06	(0.07)
Blue Cross	-0.55**	(0.22)	-0.48**	(0.24)	-0.45	(0.24)
Share in CA	3.48**	(1.45)	3.14**	(1.51)	2.92**	(1.55)
pseudo-R^2	0.33		0.34		0.36	
Number hospitals	144		144		144	

Notes: NLLS analysis of variation in hospital discounts d_h across hospitals, insurers and markets. See notes to Appendix 2 Table 1 for details. ** = significant at p=0.05; *=significant at p=0.10.

Appendix 2, Table 3: NLLS Discount Analysis with Variation Across Diagnoses

	Model 1		Model 2		Model 3		Model 4	
	Coefft	(S.E.)	Coefft	(S.E.)	Coefft	(S.E.)	Coefft	(S.E.)
Hospital Characteristics								
Constant	-0.16	(0.28)	-0.11	(0.28)	-0.15	(0.29)	-0.11	(0.28)
Teaching hospital	-0.05	(0.11)	-0.05	(0.11)	-0.08	(0.23)	-0.08	(0.23)
Cost per admission	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)	-0.01	(0.01)
For profit	0.42**	(0.12)	0.42**	(0.13)	0.42**	(0.13)	0.42**	(0.13)
Offers transplants	-0.04	(0.17)	-0.05	(0.17)	-0.05	(0.18)	-0.05	(0.18)
System hospital	0.26**	(0.12)	0.26**	(0.12)	0.26**	(0.12)	0.26**	(0.12)
Share of beds in mkt	-11.45	(8.06)	-14.86	(8.58)	-11.70	(8.08)	-14.81	(8.52)
Insurer Characteristics								
Pcare/Aetna/HN/Cigna	-0.11	(0.07)	-0.11	(0.07)	-0.11	(0.07)	-0.10	(0.07)
Blue Cross	-0.36	(0.22)	-0.37	(0.22)	-0.36	(0.22)	-0.37	(0.22)
Share in CA	1.76	(1.33)	1.83	(1.32)	1.81	(1.34)	1.84	(1.32)
γ_h parameters								
Constant	1.06**	(0.23)	0.88**	(0.32)	0.96**	(0.32)	0.87**	(0.33)
Delivery services			0.06	(0.26)	0.11	(0.26)	0.06	(0.26)
Share of beds in market			15.27	(21.65)			15.17	(21.64)
Teaching hospital					0.07	(0.58)	0.07	(0.57)
Market FEs?	Yes		Yes		Yes		Yes	
pseudo-R^2	0.45		0.45		0.45		0.45	
Number hospitals	144		144		144		144	

Notes: NLLS analysis of variation in hospital discounts d_h across hospitals, insurers and markets. See notes to Appendix 2 Table 1 for details. γ_h is a linear expression that determines the extent to which discounts for births differ from those for other diagnoses for hospital h. ** = significant at p=0.05; *=significant at p=0.10.

Appendix 4a: Mother's Outcome Variation Across Aggregated Price and Severity Groups

No. diags	Max rank 1			Max rank 2			Max rank 3		
of max rank	Pats	readm	not home	Pats	readm	not home	Pats	readm	not home
1	26721	1.97%	1.23%	14863	2.46%	1.90%	1310	3.89%	1.68%
		(0.08%)	(0.07%)		(0.13%)	(0.11%)		(0.53%)	(0.36%)
2	14187	2.11%	1.43%	4708	3.40%	2.15%	52	7.69%	1.92%
		(0.12%)	(0.10%)		(0.26%)	(0.21%)		(3.73%)	(1.92%)
3	5086	2.38%	1.75%	1431	3.84%	2.66%	5	0.00%	40.0%
		(0.21%)	(0.18%)		(0.51%)	(0.43%)		(0.00%)	(24.5%)
4	1496	1.80%	2.47%	414	3.86%	3.62%	1	0.00%	0.00%
		(0.34%)	(0.40%)		(0.95%)	(0.92%)		-	-
5	407	1.72%	1.72%	119	5.04%	3.36%	1	100%	100%
		(0.65%)	(0.65%)		(2.01%)	(1.66%)		-	-
≥ 5	124	3.23%	1.61%	47	8.51%	0.00%	0	-	-
		(1.59%)	(1.14%)		(4.11%)	(0.00%)			

Notes: Data taken from OSHPD Birth Cohort 2003 (see notes to Table 3 for details). Comparison of maternal outcomes for patients who have a Charlson score of 0 across comorbidity ranks. Standard errors in parentheses (calculated assuming that 0/1 variables are binomially distributed). "Pats" shows the number of patients in each "max rank" group and each "number diags of max rank" group. Here "Max rank j" means the maximum rank of a comorbidity for this patient, as defined by obstetrical experts at Columbia Presbyterian Hospital, is j. "Number diags of max rank" groups patients according to the number of co-morbidities in their discharge record with the relevant max rank. Patients in different rows of a particular column of the table will have different price groups. "Readm" is the percent of patients readmitted to hospital within 12 months of birth episode; "Not home" is percent of patients discharged somewhere other than home (including transfer to acute care setting, transfer to skilled nursing facility, discharge against medical advice and death).

Appendix 4b: Infant Outcome Variation Across Aggregated Price and Severity Groups

No. diags	Max rank 1			Max rank 2			Max rank 3		
of max rank	Pats	readm	not home	Pats	readm	not home	Pats	readm	not home
1	26721	8.44%	4.49%	14863	10.19%	8.05%	1310	12.98%	13.44%
		(0.17%)	(0.13%)		(0.25%)	(0.22%)		(0.93%)	(0.94%)
2	14187	9.16%	5.17%	4708	11.21%	10.94%	52	15.38%	19.23%
		(0.24%)	(0.19%)		(0.46%)	(0.45%)		(5.05%)	(5.52%)
3	5086	8.61%	6.90%	1431	12.86%	13.14%	5	0.00%	40.00%
		(0.39%)	(0.36%)		(0.89%)	(0.89%)		(0.00%)	(24.49%)
4	1496	9.22%	7.62%	414	12.32%	13.04%	1	0.00%	100%
		(0.75%)	(0.69%)		(1.62%)	(1.66%)		-	-
5	407	7.37%	11.06%	119	15.97%	21.01%	1	100%	100%
		(1.30%)	(1.56%)		(3.37%)	(3.75%)		-	-
≥ 5	124	7.26%	8.87%	47	29.79%	29.79%	0	-	-
		(2.34%)	(2.57%)		(6.74%)	(6.74%)			

Notes: Data taken from OSHPD Birth Cohort 2003 (see notes to Table 3 for details). Comparison of infant outcomes for patients who have a Charlson score of 0 across comorbidity ranks. Standard errors in parentheses (calculated assuming that 0/1 variables are binomially distributed). "Pats" shows the number of patients in each "max rank" group and each "number diags of max rank" group. Here "Max rank j" means the maximum rank of a comorbidity for this patient, as defined by obstetrical experts at Columbia Presbyterian Hospital, is j. "Number diags of max rank" groups patients according to the number of co-morbidities in their discharge record with the relevant max rank. Patients in different rows of a particular column of the table will have different price groups. "Readm" is the percent of patients readmitted to hospital within 12 months of birth episode; "Not home" is percent of patients discharged somewhere other than home (including transfer to acute care setting, transfer to skilled nursing facility, discharge against medical advice and death).

Appendix 5: Average Individual Variation in Quality and Expected Price

	Bay Area	San Diego	Los Angeles	Orange Cty	Inland Empire	*All Markets*
Standard Deviation in Quality ($000)						
Severity 1	1.637	0.716	1.136	1.011	3.259	*1.337*
Severity 2	2.237	0.313	1.202	1.122	3.064	*1.467*
Severity 3	2.454	0.943	1.500	1.299	3.851	*1.779*
Severity 4	1.994	0.434	1.482	1.650	4.431	*1.718*
Severity 5	2.674	0.254	2.145	1.135	3.432	*1.823*
Standard Deviation in Expected Price ($000)						
Severity 1	1.867	0.679	1.278	1.158	1.509	*1.311*
Severity 2	2.917	0.958	1.652	1.651	1.927	*1.857*
Severity 3	3.070	1.243	1.865	2.329	2.099	*2.182*
Severity 4	3.101	1.272	2.199	2.513	2.410	*2.396*
Severity 5	4.000	1.189	2.649	2.346	2.749	*2.678*

Notes: Statistics are cross-patient averages of standard deviations across hospitals in the choice set. Expected price and quality are both measured in $000; quality is defined as $\alpha_\pi q_{h,s}/\theta_{p,\pi}$. Data are recorded for the five largest markets (those for which quality estimates were generated) and for the five "super-severities". The super-severities are: Group 1 are patients who have a rank 1 (routine) principal diagnosis, rank 1 comorbidities and are young; Group 2 are patients who have rank 2 principal diagnosis, rank 1 comorbidities and young; Group 3 are patients who have rank 1 principal diagnosis, maximum rank 2 comorbidities and young; Group 4 are patients with rank 2 principal diagnosis, maximum rank 2 comorbidities and young; and Group 5 has all other patients.

Appendix 3: Categorization of Co-Morbidities by Severity Rank

We asked obstetrical experts at Columbia Presbyterian Hospital to assign a rank to each co-morbidity listed in our discharge data covering privately insured patients admitted for a labor/birth episode in California in 2003. Ranks were numbered from 1 to 3, where 1 indicated a routine diagnosis that would not affect patient treatment in any significant way, 2 indicated a more severe diagnosis and 3 indicated the most severe conditions that would have a substantial effect on the patient's treatment during the labor/birth admission. The list of diagnoses and their assigned ranks is given below. The number of patients with each co-morbidity is also provided. (A single patient may have more than one co-morbidity.)

Diagnosis	# patients	% patients	Rank (1-3)
1. Tuberculosis	9	0	3
2. Septicemia (except in labor)	42	0.02	2
3. Bacterial infection; unspecified sit	668	0.32	2
4. Mycoses	28	0.01	2
6. Hepatitis	119	0.06	2
7. Viral infection	643	0.3	2
8. Other infections; including parasiti	70	0.03	2
9. Sexually transmitted infections (not	19	0.01	2
10. Immunizations and screening for inf	12,523	5.93	1
22. Melanomas of skin	10	0	3
23. Other non-epithelial cancer of skin	6	0	3
24. Cancer of breast	18	0.01	3
26. Cancer of cervix	14	0.01	3
28. Cancer of other female genital orga	2	0	3
32. Cancer of bladder	1	0	3
33. Cancer of kidney and renal pelvis	2	0	3
35. Cancer of brain and nervous system	5	0	3
36. Cancer of thyroid	24	0.01	3
37. Hodgkins disease	8	0	3
38. Non-Hodgkins lymphoma	5	0	3
39. Leukemias	3	0	3
41. Cancer; other and unspecified prima	4	0	3
44. Neoplasms of unspecified nature or	14	0.01	3
46. Benign neoplasm of uterus	1,110	0.53	1
47. Other and unspecified benign neopla	275	0.13	1
48. Thyroid disorders	1,266	0.6	2
49. Diabetes mellitus without complicat	9	0	2
50. Diabetes mellitus with complication	35	0.02	3
51. Other endocrine disorders	81	0.04	2
52. Nutritional deficiencies	22	0.01	1
53. Disorders of lipid metabolism	11	0.01	2
55. Fluid and electrolyte disorders	554	0.26	2
56. Cystic fibrosis	1	0	3
57. Immunity disorders	8	0	2
58. Other nutritional; endocrine; and m	703	0.33	2
59. Deficiency and other anemia	1,542	0.73	1
60. Acute posthemorrhagic anemia	215	0.1	2
61. Sickle cell anemia	59	0.03	3
62. Coagulation and hemorrhagic disorde	338	0.16	2
63. Diseases of white blood cells	37	0.02	2
64. Other hematologic conditions	9	0	2
76. Meningitis (except that caused by t	9	0	3
77. Encephalitis (except that caused by	1	0	3

Diagnosis	# patients	% patients	Rank (1-3)
78. Other CNS infection and poliomyelit	3	0	3
79. Parkinsons disease	2	0	3
80. Multiple sclerosis	28	0.01	3
81. Other hereditary and degenerative n	10	0	3
82. Paralysis	8	0	3
83. Epilepsy; convulsions	146	0.07	3
84. Headache; including migraine	174	0.08	1
85. Coma; stupor; and brain damage	6	0	3
87. Retinal detachments; defects; vascu	5	0	2
88. Glaucoma	3	0	2
89. Blindness and vision defects	17	0.01	2
90. Inflammation; infection of eye (exc	10	0	1
91. Other eye disorders	4	0	1
92. Otitis media and related conditions	16	0.01	1
93. Conditions associated with dizzines	27	0.01	1
94. Other ear and sense organ disorders	21	0.01	1
95. Other nervous system disorders	103	0.05	2
96. Heart valve disorders	540	0.26	3
97. Peri-; endo-; and myocarditis; card	19	0.01	3
98. Essential hypertension	581	0.27	2
99. Hypertension with complications and	18	0.01	3
101. Coronary atherosclerosis and other	1	0	3
102. Nonspecific chest pain	21	0.01	2
103. Pulmonary heart disease	7	0	3
104. Other and ill-defined heart diseas	12	0.01	3
105. Conduction disorders	28	0.01	3
106. Cardiac dysrhythmias	193	0.09	3
107. Cardiac arrest and ventricular fib	2	0	3
108. Congestive heart failure; nonhyper	1	0	3
114. Peripheral and visceral atheroscle	3	0	3
117. Other circulatory disease	187	0.09	2
118. Phlebitis; thrombophlebitis and th	74	0.04	2
119. Varicose veins of lower extremity	4	0	1
120. Hemorrhoids	186	0.09	1
121. ther diseases of veins and lymphat	18	0.01	2
122. Pneumonia (except that caused by t	66	0.03	2
123. Influenza	21	0.01	1
125. Acute bronchitis	13	0.01	1
126. Other upper respiratory infections	190	0.09	1
129. Aspiration pneumonitis; food/vomit	6	0	2
130. Pleurisy; pneumothorax; pulmonary	42	0.02	3
131. Respiratory failure; insufficiency	12	0.01	3
133. Other lower respiratory disease	79	0.04	2
134. Other upper respiratory disease	19	0.01	2
135. Intestinal infection	37	0.02	1
136. Disorders of teeth and jaw	5	0	1
138. Esophageal disorders	101	0.05	2
139. Gastroduodenal ulcer (except hemor	1	0	2
140. Gastritis and duodenitis	24	0.01	1
141. Other disorders of stomach and duo	13	0.01	1
142. Appendicitis and other appendiceal	67	0.03	2
143. Abdominal hernia	94	0.04	1

Diagnosis	# patients	% patients	Rank (1-3)
144. Regional enteritis and ulcerative	55	0.03	2
145. Intestinal obstruction without her	41	0.02	2
146. Diverticulosis and diverticulitis	2	0	2
147. Anal and rectal conditions	16	0.01	1
148. Peritonitis and intestinal abscess	8	0	3
149. Biliary tract disease	401	0.19	2
151. Other liver diseases	84	0.04	2
152. Pancreatic disorders (not diabetes	41	0.02	2
153. Gastrointestinal hemorrhage	12	0.01	3
154. Noninfectious gastroenteritis	61	0.03	1
155. Other gastrointestinal disorders	390	0.18	2
156. Nephritis; nephrosis; renal sclero	11	0.01	2
157. Acute and unspecified renal failur	8	0	3
158. Chronic renal failure	2	0	3
159. Urinary tract infections	838	0.4	1
160. Calculus of urinary tract	216	0.1	1
161. Other diseases of kidney and urete	191	0.09	2
162. Other diseases of bladder and uret	15	0.01	2
163. Genitourinary symptoms and ill-def	97	0.05	1
167. Nonmalignant breast conditions	14	0.01	1
168. Inflammatory diseases of female pe	837	0.4	1
169. Endometriosis	94	0.04	1
170. Prolapse of female genital organs	3	0	1
171. Menstrual disorders	5	0	1
172. Ovarian cyst	297	0.14	1
173. Menopausal disorders	3	0	1
174. Female infertility	6	0	1
175. Other female genital disorders	448	0.21	1
176. Contraceptive and procreative mana	5,442	2.58	1
177. Spontaneous abortion	20	0.01	1
178. Induced abortion	9	0	1
179. Postabortion complications	98	0.05	2
180. Ectopic pregnancy	11	0.01	2
181. Other complications of pregnancy	16,871	7.99	2
182. Hemorrhage during pregnancy; abrup	755	0.36	3
183. Hypertension complicating pregnanc	2,388	1.13	2
184. Early or threatened labor	3,223	1.53	2
185. Prolonged pregnancy	5,103	2.42	1
186. Diabetes or abnormal glucose toler	3,501	1.66	2
187. Malposition; malpresentation	3,375	1.6	1
188. Fetopelvic disproportion; obstruct	3,061	1.45	2
189. Previous C-section	2,592	1.23	1
190. Fetal distress and abnormal forces	2,586	1.22	1
191. Polyhydramnios and other problems	5,086	2.41	2
192. Umbilical cord complication	10,393	4.92	1
193. OB-related trauma to perineum and	3,157	1.49	1
194. Forceps delivery	273	0.13	1
195. Other complications of birth; puer	26,576	12.58	1
196. Normal pregnancy and/or delivery	83,408	39.48	1
197. Skin and subcutaneous tissue infec	66	0.03	1
198. Other inflammatory condition of sk	92	0.04	1
200. Other skin disorders	182	0.09	1

Diagnosis	# patients	% patients	Rank (1-3)
201. Infective arthritis and osteomyeli	2	0	2
202. Rheumatoid arthritis and related d	5	0	2
203. Osteoarthritis	2	0	1
204. Other non-traumatic joint disorder	23	0.01	1
205. Spondylosis; intervertebral disc d	212	0.1	1
206. Osteoporosis	3	0	2
208. Acquired foot deformities	3	0	1
209. Other acquired deformities	6	0	1
210. Systemic lupus erythematosus and c	7	0	2
211. Other connective tissue disease	93	0.04	2
212. Other bone disease and musculoskel	35	0.02	2
213. Cardiac and circulatory congenital	42	0.02	2
214. Digestive congenital anomalies	2	0	2
215. Genitourinary congenital anomalies	240	0.11	2
216. Nervous system congenital anomalie	5	0	2
217. Other congenital anomalies	47	0.02	2
218. Liveborn	1	0	1
219. Short gestation; low birth weight;	2	0	2
224. Other perinatal conditions	6	0	2
225. Joint disorders and dislocations;	5	0	2
226. Fracture of neck of femur (hip)	2	0	2
228. Skull and face fractures	3	0	2
229. Fracture of upper limb	9	0	2
230. Fracture of lower limb	8	0	2
231. Other fractures	15	0.01	2
232. Sprains and strains	21	0.01	1
233. Intracranial injury	6	0	3
234. Crushing injury or internal injury	6	0	3
235. Open wounds of head; neck; and tru	5	0	2
236. Open wounds of extremities	3	0	2
237. Complication of device; implant or	21	0.01	2
238. Complications of surgical procedur	138	0.07	2
239. Superficial injury; contusion	55	0.03	1
240. Burns	2	0	2
242. Poisoning by other medications and	5	0	2
244. Other injuries and conditions due	45	0.02	2
245. Syncope	27	0.01	2
246. Fever of unknown origin	58	0.03	2
247. Lymphadenitis	5	0	2
249. Shock	3	0	3
250. Nausea and vomiting	32	0.02	1
251. Abdominal pain	185	0.09	1
252. Malaise and fatigue	15	0.01	1
253. Allergic reactions	194	0.09	2
255. Administrative/social admission	13	0.01	1
256. Medical examination/evaluation	1	0	1
257. Other aftercare	37	0.02	1
259. Residual codes; unclassified	1,537	0.73	1
650. Adjustment disorders	11	0.01	1
651. Anxiety disorders	129	0.06	1
652. Attention-deficit, conduct, and di	3	0	1
654. Developmental disorders	2	0	1

Diagnosis	# patients	% patients	Rank (1-3)
655. Disorders usually diagnosed in inf	1	0	1
657. Mood disorders	397	0.19	2
658. Personality disorders	5	0	2
659. Schizophrenia and other psychotic	8	0	2
660. Alcohol-related disorders	13	0.01	2
661. Substance-related disorders	164	0.08	2
663. Screening and history of mental he	410	0.19	1
670. Miscellaneous disorders	684	0.32	2