

Pricing and Welfare in Health Plan Choice[†]

By M. KATE BUNDORF, JONATHAN LEVIN, AND NEALE MAHONEY*

Premiums in health insurance markets frequently do not reflect individual differences in costs, either because consumers have private information or because prices are not risk rated. This creates inefficiencies when consumers self-select into plans. We develop a simple econometric model to study this problem and estimate it using data on small employers. We find a welfare loss of 2–11 percent of coverage costs compared to what is feasible with risk rating. Only about one-quarter of this is due to inefficiently chosen uniform contribution levels. We also investigate the reclassification risk created by risk rating individual incremental premiums, finding only a modest welfare cost. (JEL G22, I13, I18)

Whether competition in health insurance markets leads to efficient outcomes is a central question for health policy. Markets are effective when prices direct consumers and firms to behave efficiently. But in health insurance markets, prices often do not reflect the different costs of coverage for different enrollees. This generates two concerns. If insurers receive premiums that do not reflect enrollee risk, they have an incentive to engage in risk selection through plan design (Rothschild and Stiglitz 1976; Newhouse 1996). Similarly, if consumers face prices that do not reflect cost differences across plans, they may select coverage inefficiently (Akerlof 1970; Feldman and Dowd 1982). While it is widely recognized that these problems may impair the efficiency of competitive health insurance markets, evidence on their quantitative importance for social welfare is limited.

In the US private market, employers often contract with insurers to create a menu of plans from which employees select coverage. The government or a quasi-public organization plays a similar role in the US Medicare program and the national systems of Germany and the Netherlands. To address incentive problems in plan design, these intermediaries have begun to “risk adjust” payments to plans (van de Ven and Ellis 2000). Consumer prices, however, are typically not adjusted for individual risk. Certain aspects of risk may be private information, and in the United States, regulations prohibit employers and public programs from charging enrollees

*Bundorf: Stanford University School of Medicine, 259 Campus Drive, Stanford, CA 94305-5405 (e-mail: bundorf@stanford.edu); Levin: Department of Economics, Stanford University, 579 Serra Mall, Stanford, CA 94305 (e-mail: jdlevin@stanford.edu); Mahoney: RWJ Scholars in Health Policy Program, Harvard University, 1730 Cambridge Street, Cambridge, MA 02138 (e-mail: neale.mahoney@gmail.com). We thank Liran Einav, Amy Finkelstein, Will Manning, and Robert Town for helpful suggestions. Levin gratefully acknowledges support from the National Science Foundation and the Center for Advanced Study in the Behavioral Sciences.

[†] To view additional materials, visit the article page at <http://dx.doi.org/10.1257/aer.102.7.3214>.

different amounts based on nearly all observable health-related factors.¹ Moreover, even within institutional limitations, contributions set by employers or in regulated markets may not be welfare maximizing given the complexities of self-selection.

In this article, we analyze the effect of plan pricing on allocative efficiency. We begin by making a basic theoretical point regarding plan prices and efficient matching. Existing work suggests that while poorly chosen contribution policies may lead to inefficient outcomes, the problem can be solved by choosing an optimal uniform contribution even in the presence of substantial asymmetric information (e.g., Feldman and Dowd 1982; Cutler and Reber 1998; Pauly and Herring 2000; Cutler and Zeckhauser 2000). These analyses, however, assume perfect correlation between enrollee risk and preferences for coverage and make strong assumptions about the relationship between preferences and plan costs.² We show that if these assumptions are violated, a uniform contribution policy (i.e., a policy under which individuals face the same prices for the plans) cannot induce efficient consumer choices. In principle, however, risk-adjusted contributions can correct or mitigate the distortion.

The main part of the paper builds on this point and looks empirically at the welfare costs of self-selection. We develop a simple econometric model of health plan demand and costs, estimate the model on a novel dataset of small employers, and then use the parameter estimates to simulate the welfare implications of alternative pricing policies. In our simulations, observed pricing policies are less efficient than what could be achieved with risk-rated plan contributions. The shortfall is between \$60 and \$325 annually per enrollee, or 2–11 percent of coverage costs, depending on the cost differences across plans for the highest-cost enrollees. Approximately one-quarter of this inefficiency can be attributed to nonoptimal uniform contributions; capturing the remainder would require setting different premiums for people in the same firm. We also account for the possibility that employees choose plans based on private information about their health status. We calculate that asymmetric information between consumers and the relatively sophisticated risk-adjustment system used by insurers in our setting reduces welfare by an additional \$35–\$100 annually per enrollee. Despite these inefficiencies, our estimates still suggest choice is beneficial because of the variation in household preferences.

The nature of the offered health plans is important for interpreting these results. We study a setting in which employees choose between two insurers. One offers a fairly broad provider network and relies on patient cost sharing and primary care gatekeepers to control utilization. The other has an integrated and closed delivery system and requires little patient cost sharing. We estimate very different cost structures for these plans. Costs appear to be similar for individuals of average health, but the integrated delivery system has significantly lower costs for those with chronic conditions. We find that consumers select into the plans based on both household preferences and health status, but in contrast to some other studies, we do not observe any single plan experiencing serious adverse selection. A possible explanation is

¹Specifically, federal regulation (29 CFR Part 2590.702) states that employers offering group health plans cannot charge employees different contributions on the basis of “health factors” (section (c)(1)(i)), defined to include health status, claims experience, medical history, genetic information or disability (section (a)(1)(i–viii)).

²Cutler, Finkelstein, and McGarry (2008) stress that a broad view of heterogeneity in preferences is important for understanding many aspects of insurance markets.

that the plans are not ordered clearly by coverage level: instead, consumers face a choice between different physicians and provider organizations, as well as differences in cost sharing.

The “horizontal” differentiation of health plans in our setting seems particularly salient given changes in the health insurance market. In 1987, approximately three-quarters of people with employer-sponsored health insurance had conventional coverage, under which plans differed primarily in their cost sharing. By 2007, in contrast, the market was dominated by managed care plans which use different mixes of supply-side and demand-side utilization management (Kaiser Family Foundation 2007), so that plans vary not just on financial characteristics such as copayments and deductibles, but on physician access and the scope of provider networks. This evolution suggests that classic insights based on purely risk-based selection may not adequately capture the dynamics of today’s market.

In our analysis, two forces play a key role: heterogeneity in household preferences and the cost advantage of the integrated system for individuals in worse health. We estimate that a large fraction of high-cost households would choose the integrated delivery system if they faced premiums that reflected the relevant cost differential. With a uniform contribution policy, however, this would mean charging all households a steep premium for the broad network insurer, creating a welfare cost for lower-risk households who value its more conventional offering. While the exact magnitudes are, of course, specific to our setting, the basic point is not: uniform pricing makes it difficult to pass on targeted cost savings.³

A possible counterpoint is that uniform contribution policies also provide intertemporal insurance. In an employer-provided insurance setting, employees who develop chronic health problems continue to face the same prices as other employees. In contrast, risk-rated pricing can create reclassification risk. Of course, the type of risk-based pricing we consider to correct static distortions involves adjusting only incremental prices for the plans, suggesting it may be possible to provide considerable intertemporal insurance through the base or average plan price. We address this in the article’s final section by combining our model with data on risk-score transitions. We find that for plausible risk attitudes, the welfare cost of reclassification risk under risk-rated incremental prices is less than 10 percent of the static benefits from improved allocation.

Our analysis ties in to past work on health plan choice and the efficiency of health insurance markets. We draw on this work on health plan choice, which is summarized by Glied (2000) and Cutler and Zeckhauser (2000), in modeling how employee demand varies with observed and unobserved risk and preference characteristics. Our article is more directly related to recent work that uses econometric methods to quantify the efficiency implications of adverse selection in health insurance markets (Cutler and Reber 1998; Cardon and Hendel 2001; Carlin and Town 2008; Einav, Finkelstein, and Cullen 2010; Einav, Finkelstein, and Levin 2010; Handel 2012). Our paper points out that uniform pricing, as is commonly observed, may lead to inefficiency when enrollees of similar risk have different preferences for coverage.

³ Here we emphasize general cost savings for households with higher risk scores, but a similar point would apply if certain insurers were able to manage particular chronic conditions more cost effectively but found themselves unable to target these households with attractive premiums.

These other papers, in contrast, analyze alternative institutional features of health insurance markets that contribute to adverse selection. We relate both our empirical approach and our findings to these papers in Section IVD.

Our results may also shed light on two puzzles in the health insurance literature. One is why employers have not systematically adopted contribution policies that pass the full premium increment of choosing higher cost plans on to employees. In our data, only a small fraction of the firms use such a policy, but our results suggest that the efficiency gains from moving in this direction would be relatively modest. The second puzzle is why the integrated model of health care delivery has struggled to catch on widely. We find that the integrated insurer achieves substantial savings for people in poor health, but that current pricing makes it difficult to target these households where it has a comparative advantage.

We emphasize that our analysis has some important limitations. First, it is based on a particular, and only moderately sized, sample of workers and firms. To address this, we perform a variety of sensitivity analyses on our key parameter estimates, which we discuss in Section IV. Second, we take plan offerings as given. This seems reasonable given that we are looking at small to medium-size employers, but a broader analysis of pricing ideally would incorporate plan design. Third, we do not address issues of utilization behavior or try to assess the relative social efficiency of health care utilization under the different plans in our data. Finally, our analysis is mostly based on a static model, although we do consider the interaction between risk-based pricing and dynamic insurance in Section V.

I. Health Plan Pricing and Market Efficiency

We discuss the relationship between pricing and market efficiency by adapting the model of Feldman and Dowd (1982). In their model, consumers are distinguished by their privately known forecastable health risk, denoted θ , and a consumer's health risk perfectly explains his or her preferences across health plans. We extend the model to allow for additional consumer heterogeneity in preferences, denoted by ε . Recent empirical work has emphasized the importance of preference heterogeneity in explaining insurance choices, and it seems particularly relevant when health plans offer access to different medical providers.

Each consumer chooses between two plans, A and B. Let $u_A(\theta, \varepsilon)$ and $u_B(\theta, \varepsilon)$ denote the (dollar) value a consumer of type (θ, ε) places on being covered under the two plans, so if the consumer pays p to enroll in plan j , her net benefit is $u_j(\theta, \varepsilon) - p$.⁴ Let $\Delta u(\theta, \varepsilon) = u_A(\theta, \varepsilon) - u_B(\theta, \varepsilon)$ denote the incremental willingness-to-pay for plan A. The plans' expected costs depend on consumer health risk. We denote these $c_A(\theta)$, $c_B(\theta)$ and assume the difference $\Delta c(\theta) = c_A(\theta) - c_B(\theta)$ is increasing in θ so that plan B has a comparative efficiency for high-risk consumers.

An efficient assignment places a type- (θ, ε) consumer in plan A if and only if

$$(1) \quad \Delta u(\theta, \varepsilon) - \Delta c(\theta) \geq 0.$$

⁴ Here we make the simplifying assumption, which we maintain in our econometric model, that plan preferences are additively separable in the plan premium. See Einav, Finkelstein, and Levin (2010) for an extensive discussion of this assumption.

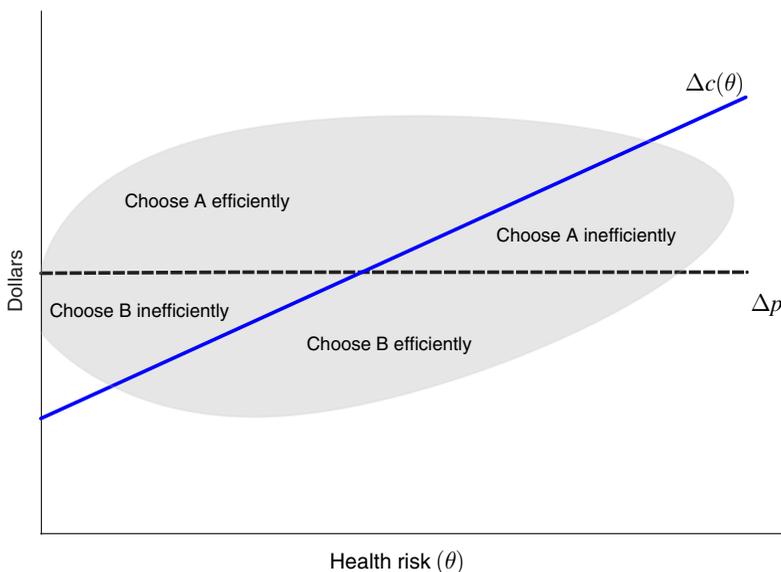


FIGURE 1. MISALLOCATION WITH UNIFORM PRICING AND HETEROGENEITY IN PREFERENCES

Notes: Figure shows misallocation from uniform pricing with heterogeneous preferences. The shaded region shows the distribution of households in preference-risk space. For these households, the y-axis value is the incremental willingness-to-pay for plan A versus plan B and the x-axis value is health risk. The solid line ($\Delta c(\theta)$) shows the incremental cost and the dashed line (Δp) shows the incremental contribution for plan A relative to plan B.

At the same time, a type- (θ, ε) consumer will select plan A if and only if

$$(2) \quad \Delta u(\theta, \varepsilon) - \Delta p \geq 0,$$

where $\Delta p = p_A - p_B$ is the incremental contribution the consumer faces for plan A.

Figure 1 provides a graphical illustration. The shaded area represents a distribution of consumers who vary in their health risk and plan preferences. Consumer health risk is on the x-axis and dollars on the y-axis. The increasing line $\Delta c(\theta)$ shows the incremental cost of covering a consumer under plan A relative to plan B. The efficient outcome is for all consumers with willingness-to-pay Δu above $\Delta c(\theta)$ to enroll in plan A. If prices vary with risk so that for all risk types, θ , the incremental contribution for plan A is set at $\Delta p(\theta) = \Delta c(\theta)$, consumers will self-select efficiently. With uniform premiums, Δp is the same for all consumers, and self-selection generally will be inefficient. As shown in Figure 1, some low-risk consumers face “too high” an incremental price and choose plan B inefficiently, while other high risks face “too low” an incremental price and choose plan A inefficiently.

The degree of inefficiency depends on two factors. The first is the plan cost structures, or specifically the slope of $\Delta c(\theta)$. The second is the distribution of consumer preferences, indicated by how Δu is distributed across the shaded area in the figure. In our empirical analysis, we essentially “fill in” Figure 1 by estimating plan costs and consumer willingness-to-pay as a function of household health status. Fixing consumer health risk, we identify preference variation by estimating price sensitivity—a more inelastic demand corresponds to a more dispersed distribution of consumer willingness-to-pay, as relatively few consumers are on the margin with

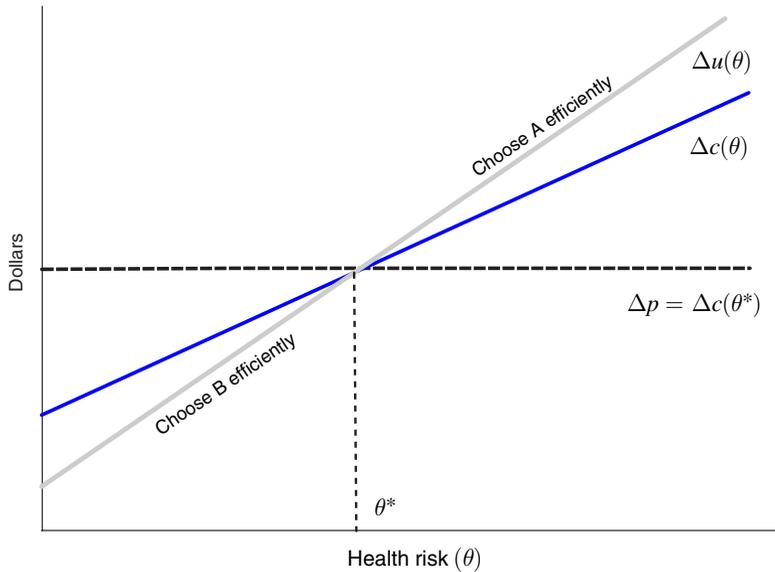


FIGURE 2. SPECIAL CASE WITH NO HETEROGENEITY AND RAPIDLY INCREASING PREFERENCES

Notes: Figure shows a special case where uniform pricing can lead to efficient allocation. The steeper solid line ($\Delta u(\theta)$) shows the homogenous relationship between the incremental willingness-to-pay for plan A versus plan B and health risk. The less-steep solid line ($\Delta c(\theta)$) shows the relationship between incremental cost and health risk. The dashed line ($\Delta p = \Delta c(\theta^*)$) shows the uniform premium that efficiently allocates households across the plans.

respect to price changes. Given these estimates, we can compute the degree of misallocation, and the extent to which it can be corrected with different price schedules.

The conclusion from Figure 1 that uniform prices generally cannot induce efficient self-selection contrasts with the standard analysis in the literature (e.g., Feldman and Dowd 1982; Cutler and Reber 1998; Cutler and Zeckhauser 2000; Miller 2005). That analysis makes two strong assumptions: that consumer willingness-to-pay Δu is perfectly correlated with health risk θ (so all consumers with a given health risk θ have identical willingness-to-pay Δu), and that $\Delta u(\theta)$ is increasing in θ more rapidly than $\Delta c(\theta)$. This situation is shown in Figure 2, in which $\Delta u(\theta)$ represents the distribution of consumer preferences. In this case, it is efficient to assign all consumers with health risk above θ^* to plan A. This can be achieved by setting $\Delta p = \Delta c(\theta^*)$. The focus of the literature in this setting has been on the consequences of poorly chosen premium differentials. For instance, if Δp is set too high, plan A attracts only very high risks, and one can end up with an adverse selection “death spiral” if prices are adjusted based on plan costs (Cutler and Reber 1998).

Both of the key assumptions in Figure 2 fail to hold in our empirical analysis. Not only do we find substantial heterogeneity in preferences, we find that the mean willingness-to-pay for the network insurer (plan A) is less sensitive to household health status than the insurer’s incremental costs. This corresponds, for consumers with “average” willingness-to-pay, to a version of Figure 2 in which $\Delta u(\theta)$ is flatter than $\Delta c(\theta)$. In this scenario, it is efficient for high-cost consumers to enroll in plan B (the integrated insurer) while low-cost consumers enroll in A, but for any uniform

premium consumers sort in the opposite direction. While this possibility is not surprising, it seems to have been neglected in prior analyses.

II. Data and Environment

A. Institutional Setting

Our analysis is based on data from a private firm that helps small and mid-sized employers manage health benefits. This firm, which we refer to as the intermediary, obtains agreements from insurers to offer plans to small employers, signs up employers, and administers their health benefit. We examine data from 11 employers who purchased coverage from the intermediary in a single metropolitan area in the western United States during 2004 and 2005.

In this market, the intermediary works with two insurers. One insurer contracts nonexclusively with a relatively broad set of providers. It offers an HMO plan (*network HMO*) that requires enrollees to choose a primary care physician and obtain a referral for specialist visits and does not cover care from out-of-network providers. It also offers a PPO plan (*network PPO*) that does not require referrals and covers providers outside the plan's network at an increased cost-share.⁵ The second insurer has an integrated and closed delivery system. It offers a standard HMO (*integrated HMO*) and a point-of-service option (*integrated POS*) that allows enrollees to seek care outside the integrated system at a higher cost.

The employers that hire the intermediary choose which plans to offer their employees. Employers may customize the basic plans to a limited degree by varying characteristics such as the deductible and the level of coinsurance, but most dimensions are fixed. Employers typically have four coverage tiers: employee only, employee plus spouse, employee plus children, and employee plus family.⁶ The level of cost sharing varies across coverage tiers. The employers do not offer any health insurance plans beyond those offered by the intermediary.

The insurers provide bids for each of the selected plans, relying on information from the intermediary. In an employer's first year with the intermediary, this information is just the distribution of employees by age and sex. In subsequent years, the insurers receive additional information on the health status of the workers, in the form of a risk score described below. The intermediary instructs the insurers to bid as if they were covering all workers within the firm. While the insurers provide bids for each tier, the bids for tiers other than employee-only are simply scaled from the employee-only bids by a constant that is very similar across employers and plans.

After the bids are received, the employer sets the employee contribution for each plan and coverage tier. The employees then make their choices, and the plans are required to accept all employees who choose to enroll. The last step is a series of payments. For each employee that enrolls in a plan, the employer pays the

⁵This insurer also offers a point-of-service (POS) plan that is the HMO with the option to go out-of-network at higher cost. We are not able to distinguish between network POS and HMO enrollees, so we simplify our analysis by dropping the three employer-years where the network POS was offered. Our results are not sensitive to alternative approaches to handling this issue.

⁶Two firms define coverage tiers based on employee only, employee plus one dependent, and employee plus two or more dependents.

intermediary the insurer's bid. The intermediary passes these payments to the insurers, implementing transfers between insurers if there is variation in the health risk of the enrollees in the different plans.

The intermediary uses a standard methodology for measuring enrollee health risk, the RxGroup model developed by DxCG, Inc. The model produces risk scores based on a person's age, sex, and chronic health conditions, where chronic conditions are inferred from prior use of prescription drugs, reported by the insurers.^{7,8} A potential concern with risk scores is that they might partially reflect how a patient's plan manages utilization, rather than the employee's health status. Our discussions with participants suggest that in this setting there were strong incentives to ensure that health risk was measured accurately. The insurers view risk adjustment as essential protection against unfavorable selection and worked with the intermediary to address potential biases. For instance, one concern was that the integrated insurer might substitute low-priced drugs more aggressively, leading the algorithm to underestimate the severity of chronic illness for its enrollees. This and related issues led to small adjustments in the risk-scoring algorithm. From what we have learned, we view it as reasonable to assume that the scores are accurate reflections of individual health-risk differences.

In addition to prescription drug utilization, each insurer also provides the intermediary with the realized costs for each employer group. The network insurer reports average claims per member per month for enrollees covered by either of the insurer's products. The integrated insurer reports similar information developed from an internal cost accounting system. Neither insurer distinguishes between its plans when reporting this information.

B. Data and Descriptive Statistics

Our data include all of the information discussed above: the plan offerings and contribution policies of each employer, the risk scores and plan choices of employees and their dependents, and the bids and reported costs of each insurer. A primary strength of the data is that it includes both demand-side information on employees and their choice behavior and supply-side information on insurer costs and bids in a setting with two very different types of insurers. In addition, many of the employers we observe offer nearly identical plans but have different risk profiles and contribution policies, which provides useful variation to identify demand and costs.

Another useful feature of the data is that we observe each employer during its first year of participation in the program. Insurers have little information on firm characteristics beyond that provided by the intermediary during the first year, allowing us to observe how plans bid when they have similar information on the likely risk of a

⁷In our analysis, we use the term "risk score" to refer to the DxCG prediction of an individual's health expenditures relative to the mean of the much larger base sample on which DxCG calibrates their model. We note that our use of the term risk refers only to the level and not to the variance of the expected expenditure, although we might naturally expect a relationship between the two.

⁸DxCG uses an internally developed algorithm to infer the presence and severity of chronic conditions from prescription drug use. The health expenditure model is estimated on a very large sample (1,000,000+) of people under 65 with private health insurance. Using the estimated model, the software predicts covered health expenditures for a given individual. A score of 1 corresponds to a mean prediction from the original estimation sample. See Zhao et al. (2001) for more detail.

TABLE 1—RISK AND DEMOGRAPHICS

	Mean	Standard deviation	Minimum	Max.
Employees ($N = 3,683$)				
Risk score	1.21	1.56	0.18	30.06
Age	40.56	12.01	18.00	72.00
Female	0.62	—	—	—
Spouse	0.28	—	—	—
Child	0.27	—	—	—
Enrollees ($N = 6,603$)				
Risk score	1.01	1.45	0.14	30.06
Age	32.13	17.67	0.00	72.00
Female	0.58	—	—	—
Spouse	0.19	—	—	—
Child	0.26	—	—	—
Firm-years ($N = 16$)				
Risk score	0.97	0.31	0.63	1.91
Age	31.67	4.63	25.71	46.09
Female	0.53	0.12	0.30	0.70
Spouse	0.19	0.07	0.08	0.27
Child	0.26	0.08	0.06	0.39
Employees	230.19	241.51	28.00	838.00
Dependents	182.50	117.51	9.00	331.00

Notes: In the first panel, spouse and child refer to the fraction of employees who enroll with a spouse or at least one child. In the second and third panels, these entries are the fraction of spouses and children in the set of enrollees. The first and second panels pool observations across firms and years. The third panel shows statistics of firm-year level averages, taken across all enrollees.

group.⁹ On the demand side, a large literature documents that health plan choices are highly persistent (e.g., Neipp and Zeckhauser 1985), so observing choice behavior in the first year likely provides a good indication of steady-state demand and allows us to observe the plan characteristics and prices at the time of initial choice.¹⁰ The data's main limitations are the fairly small number of observations and restricted set of employee characteristics relative to, say, the HR records of a large employer, and also the aggregated reporting of realized costs.

The 11 firms have 2,044 covered employees and 4,652 enrollees (employees and their dependents). We observe five of the employers for two years, creating a total of 3,683 employee-years and 6,603 enrollee-years. Table 1 provides summary statistics on the covered employees, the enrollees, and the firms. Sixty-two percent of employees are female; the average age is just over 40. Fifty-eight percent of enrollees are female, and enrollees are younger on average than employees, driven primarily by covered children. Twenty-eight percent of employees enroll in a plan that covers their spouse, and 27 percent enroll in a plan that covers at least one child.

Table 1 also presents risk scores at the employee, enrollee, and employer levels. A score of one represents an average individual in a nationally representative sample, and a score of two indicates that an individual's expected health costs are twice

⁹In a few cases, an employer had a prior contract with one of the insurers. We have examined whether incorporating this into our employee demand model affects our estimates and found it did not. One concern is that this situation could result in asymmetric information between the plans in the bidding, but we think this is unlikely to be an important problem.

¹⁰We ultimately use all 16 firm-years in the demand estimates we report here, as the estimates are more precise and very similar to the estimates that use only the 11 first firm-years.

TABLE 2—PLAN CHARACTERISTICS AND ENROLLMENT

	Network		Integrated		All
	HMO	PPO	HMO	POS	
Offering plan					
Firms	11	10	11	9	—
Firm-years	16	14	16	13	—
Bid (monthly)					
Employee	307 (64)	332 (59)	260 (30)	276 (26)	294 (54)
Employee plus spouse	645 (154)	689 (123)	544 (61)	579 (54)	616 (120)
Employee plus child(ren)	591 (143)	632 (115)	498 (58)	532 (53)	565 (111)
Employee plus family	918 (200)	989 (176)	779 (87)	832 (76)	882 (164)
Contribution (monthly)					
Employee	45 (34)	73 (54)	38 (32)	58 (40)	53 (41)
Employee plus spouse	252 (120)	303 (103)	203 (77)	255 (75)	253 (100)
Employee plus child(ren)	221 (97)	265 (86)	177 (62)	223 (55)	222 (81)
Employee plus family	418 (213)	495 (182)	342 (144)	415 (140)	418 (176)
Coinsurance (percent)					
Employee	87 (6)	86 (5)	97 (7)	78 (2)	87 (9)
Deductible (annual)					
Employee	387 (264)	440 (306)	69 (163)	336 (94)	304 (262)
Out-of-pocket max (annual)					
Employee	2,818 (462)	2,850 (474)	1,591 (625)	2,686 (731)	2,468 (775)

Notes: Mean plan characteristics with standard deviations in parentheses. Plan characteristics are pooled across years. Coinsurance, deductible, and out-of-pocket maximum are in-network values and are highly correlated ($\rho > 0.9$) with the out-of-network values. Coverage tiers based on employee plus one dependent and employee plus two or more dependents are used at two firms. Bids and costs for these coverage tiers are not shown.

the average. The average risk scores of employees and enrollees are 1.21 and 1.01, respectively. The difference reflects the lower expected expenditures for covered children. Average risk ranges widely across employers, from 0.63 to 1.91. One reason for the degree of variation is the small number of enrollees at some of the firms in our data. This variation plays a key role in our analysis. We use information on insurer bids and realized costs to estimate models of the relationship between costs and risk. Because insurers report both bids and costs at the employer level, variation across employers in average risk is necessary to identify these relationships.

Table 2 provides information on the plans offered by the employers in our sample. Most employers offer all four plans, and all offer both HMOs and at least one other plan. On average, the integrated HMO is the least expensive plan and has the lowest enrollee contribution. This plan features high rates of coinsurance (expressed as the proportion of expenditure covered by the plan), a low deductible, and a low out-of-pocket maximum. The network PPO is on average the most expensive plan and has the highest employee contribution. It features lower coinsurance rates, higher deductibles and higher maximum expenditures. Roughly speaking, the other two plans fall between these extremes. As noted above, bids for each plan vary across

TABLE 3—RISK AND DEMOGRAPHICS BY PLAN

	Network		Integrated		All
	HMO	PPO	HMO	POS	
Employees ($N = 3,683$)					
Risk score	1.19	1.22	1.22	1.19	1.21
Age	42.17	40.79	39.73	41.35	40.56
Female	0.62	0.52	0.65	0.56	0.62
Market share (percent)	22.94	7.38	58.72	10.96	100
Enrollees ($N = 6,603$)					
Risk score	1.02	1.04	0.99	1.05	1.01
Age	34.19	33.06	30.94	34.12	32.15
Female	0.58	0.54	0.59	0.55	0.58
Market share (percent)	21.24	7.84	60.35	10.57	100

Note: Employees and enrollees are pooled across firms and years.

tiers by a scaling factor that is very similar across plans and employers. Employee contributions also vary across tiers, with employees typically facing a greater fraction of the plan bid for dependent coverage. Variation in these contributions is important for the identification of our demand model. We discuss contributions in detail in the identification section.

We summarize enrollment patterns in Table 3. The integrated HMO attracts by far the most enrollees with a 59 percent market share among employees and 60 percent market share among enrollees. The integrated HMO attracts a slightly younger population, but there is little evidence of extensive risk selection. The plans have similar average enrollee risk scores. The lack of sorting is not driven by heterogeneity in plan offerings across firms. If we condition on employers that offer both the PPO and the integrated HMO, for example, the average enrollee risk is 1.04 in both plans.

III. Econometric Model

A. Consumer Preferences, Plan Costs, and Market Behavior

In this section, we develop an econometric model that allows us to jointly estimate consumer preferences and health plan costs. Note that by costs we mean overall costs to the insurer for a given enrollee in a given plan. Although we discuss factors that may create cross-plan variation in costs, overall cost is sufficient for welfare analysis, and it is not necessary to decompose whether cost differences arise from, for example, moral hazard or physician reimbursement rates or some other factor (cf., Einav, Finkelstein, and Levin 2010).

In contrast to the theoretical model above, the econometric model allows for multiple plans, varying plan characteristics, and both observable and privately known dimensions of health risk and consumer tastes. Nevertheless, we aim for the most parsimonious model that permits a credible assessment of market efficiency. In what follows, we describe the key components of the model: consumer choice, health plan costs, health plan bidding, and employer contribution setting, and the stochastic assumptions on the unobservables that permit estimation.

Consumer Choice.—We use a standard latent utility model to describe household choice behavior, where a household's (money-metric) utility from choosing a plan depends on household and plan characteristics. Specifically, household h 's utility from choosing plan j is

$$(3) \quad u_{hj} = \phi_j \alpha_{\phi} + x_h \alpha_{xj} + \psi(r_h + \mu_h; \alpha_{rj}) - p_j + \sigma_{\varepsilon} \varepsilon_{hj}.$$

In this representation, household utility depends on observable plan characteristics ϕ_j , the monthly plan contribution p_j ,¹¹ observable household demographics x_h , an idiosyncratic preference ε_{hj} , and household health risk. Our measure of household health risk is aggregated from the individual level. For each individual i , we decompose health risk into the observable risk score r_i and additional privately known health factors μ_i . The μ_i s capture information about health status that may affect choice behavior but are not subject to risk adjustment. Equivalently, we can interpret μ_i as measurement error in the risk score. We assume that each μ_i is an i.i.d. draw from a normal distribution with mean zero and variance σ_{μ}^2 , and that the idiosyncratic tastes ε_{hj} are i.i.d. type I extreme value random variables.

We handle heterogeneity in household size and composition by assuming that, apart from the treatment of health risk, each household behaves as if it had a representative member with characteristics equal to the average of those of its members.¹² We parameterize household risk using two variables: the average risk of household members (i.e., the average of the $r_i + \mu_i$) and an indicator of whether the household includes a high-risk member. We define high risk as being above 2.25, which corresponds to the 90th percentile of the observed risk-score distribution. The other household characteristics in the model are the averages of age and the male indicator among covered household members as well as imputed household income.¹³

In addition to the employee contribution, plan characteristics ϕ_j include a dummy variable for plan (the network HMO and PPO and the integrated HMO and POS), the relevant coinsurance rate and deductible for the given employee, and an indicator of nonstandard drug coverage.¹⁴ To be consistent with our approach to household aggregation, we divide both the contribution and the deductible by the number of enrollees covered by the contract.

¹¹ We convert employee contributions, which are made with pretax dollars, to post-tax dollars by adjusting them by the marginal tax rate (see footnote 12 for discussion). For a given household h , let ρ_h be the nominal contribution and τ_h the household's marginal tax rate. The tax adjusted contribution is $p_h = (1 - \tau_h)\rho_h$.

¹² We experimented with estimating different weights for household members, and also with restricting the sample to individual enrollees. Neither has much effect on our results. The online Appendix includes individual enrollee estimates.

¹³ We impute taxable income for each household in our sample by estimating a model of household income as a function of worker age, sex, family structure, firm size and industry using data from the Current Population Survey for 2004 and 2005 on workers with employer-sponsored health insurance in the corresponding state. We then use the model to impute household income for each employee in our data incorporating random draws from the posterior distributions of the regression coefficients and the standard deviation of the residuals. Based on these predictions, we use Taxsim to calculate marginal tax rates based on federal, state, and FICA taxes, making some assumptions on the correlation of coverage tier with filing status and number of dependents. The average taxable family income and marginal tax rate for workers in our sample are about \$73,000 and 41 percent, respectively.

¹⁴ While the prescription drug coverage for each plan is complicated and involves both formularies and tiered cost sharing, it is generally standardized within plans across employers. This variable is an indicator of the two employers whose coverage deviates from the standard, in both cases being less generous.

For each household h , we observe the set of available plans \mathcal{J}_h and the plan chosen. Let q_{hj} be a dummy variable indicating whether household h chooses plan $j \in \mathcal{J}_h$. Given our specification,

$$(4) \quad q_{hj} = 1 \Leftrightarrow u_{hj} \geq u_{hk} \text{ for all } k \in \mathcal{J}_h.$$

Recall that the utility function includes two unobservables: the idiosyncratic taste ε_{hj} and the private health risks of household members μ_h . Conditional on the μ_h s, however, we have a standard logit specification. In particular, if we define $v_{hj} = u_{hj} - \sigma_\varepsilon \varepsilon_{hj}$, and let \mathbf{X}_{hj} denote the full set of relevant observables, we have the familiar formula for choice probabilities:

$$(5) \quad \Pr(q_{hj} = 1 | \mathbf{X}_{hj}, \mu_h) = \frac{\exp(v_{hj})}{\sum_{k \in \mathcal{J}_h} \exp(v_{hk})}.$$

Health Plan Costs.—We model each plan j 's cost of enrolling a given individual as a function of the plan's base cost for a "standard" enrollee with risk score 1, an adjustment based on how the forecastable risk varies from the baseline, and an idiosyncratic health shock. Specifically, we write j 's cost of enrolling individual i as

$$(6) \quad c_{ij} = a_j + b_j \cdot (r_i + \mu_i - 1) + \eta_{ij}.$$

In this specification, a_j represents plan j 's baseline expected cost for a standard enrollee, and b_j is the marginal cost of insuring a higher (or lower) health risk. Again we decompose forecasted health risk into the observable risk score r_i and the private information component μ_i . We allow both the base cost a_j and the marginal cost b_j to depend on plan characteristics, most importantly the underlying plan type. We assume that each η_{ij} is an independent mean-zero random variable.

Our cost data are aggregated to the insurer-firm-year level so we aggregate the individual cost model accordingly. Let \mathcal{I}_{jf} denote the set of enrollees in plan j in firm-year f , and let \mathcal{J}_{kf} be the set of plans offered by insurer k . (To keep subscripts manageable, we use f rather than ft to index firm-years.) Aggregated costs are then

$$(7) \quad C_{kf} = \sum_{j \in \mathcal{J}_{kf}} \sum_{i \in \mathcal{I}_{jf}} \{a_j + b_j \cdot (r_i + \mu_i - 1) + \eta_{ij}\}.$$

Health Plan Bidding.—The next component of our model is the plan bids. As described above, in a firm's first year of participation, each insurer had the same limited information about each firm, namely the age and sex of employees but not dependents. The intermediary instructed insurers to bid assuming they were covering all workers within the firm, assuring them that the payments they received would be adjusted based on the risk scores of actual enrollees.

We assume that the insurers bid roughly as instructed, submitting a marked-up estimate of their costs for insuring all employees at each given firm under a particular plan. We also assume that insurers bid based only on the information available from the intermediary. To ensure the validity of this assumption, we limit the data to first-year bids when the insurers had no experience with a particular employer. The

fact that each firm represents only a tiny fraction of each insurer’s business also supports the plausibility of this assumption. To the extent that providers were concerned about unfavorable risk selection, it seems likely that they would simply bid a larger profit margin for all coverage sold through the intermediary rather than investing effort to collect additional information to fine-tune each bid.

To formalize the model, let \mathcal{I}_f denote the set of employees in firm f , and \mathbf{x}_i the demographic information about employee i that was available to the insurers, i.e., age and sex. The expected cost for plan j to cover a representative employee of firm f is

$$(8) \quad \frac{1}{|\mathcal{I}_f|} \sum_{i \in \mathcal{I}_f} \mathbb{E}[c_{ij} | \mathbf{x}_i] = a_j + b_j(\mathbb{E}[\bar{r}_f | \mathbf{x}_f] - 1),$$

where \bar{r}_f denotes the average risk of employees in firm f , which the insurer forecasts using the available demographic information, \mathbf{x}_f .¹⁵

We model expected plan bids as a markup over expected cost. So plan j ’s bid for firm f is

$$(9) \quad B_{jf} = \delta_j \cdot (a_j + b_j \cdot (\mathbb{E}[\bar{r}_f | \mathbf{x}_f] - 1)) + \nu_{jf},$$

where ν_{jf} is an independent mean zero random variable. The new parameter introduced in the bid model is the markup, δ_j . We constrain the markup to be constant across the plans offered by a particular insurer. Although in theory an insurer could vary the markup across its different plans, because the cost data are at the insurer-firm level, we are unable to identify separately the markup and the fixed costs for each plan offered by an insurer. Naturally we expect the markup parameters to be larger than one.

Employer Contribution Setting.—The last part of our model specifies how employers set plan contributions. We adopt a simple model in which employers pass on a fraction of their cost for the lowest cost plan, and then a fraction of the incremental cost for higher cost plans. We allow these fractions, denoted β and γ , to vary across firm-years and coverage tiers.

Let \underline{B}_{lf} denote the minimum bid received for coverage tier l in firm-year f , denote plan j ’s bid for coverage tier l in firm-year f as $B_{jl f}$. We model the required contribution as

$$(10) \quad p_{jl f} = \beta_{lf} \cdot \underline{B}_{lf} + \gamma_{lf} \cdot (B_{jl f} - \underline{B}_{lf}) + \xi_{jl f}.$$

This model describes employer behavior in our data remarkably well. The residuals from the linear regression (10) have a standard deviation of 7.64, and the R^2 is 0.99. Approximately half of the firms in our data choose a “proportional pass-through” strategy where $\beta = \gamma$. The others choose an “incremental pass-through” strategy in which $\beta < \gamma$.

¹⁵We construct $\mathbb{E}[r | \mathbf{x}]$ by regressing risk score on fully interacted dummy variables for age group and sex.

B. Discussion of Model and Identification

The key quantities in our model are the structure of plan costs, and the distribution of consumer preferences—in particular, the extent to which household demand varies with plan prices and the household's forecastable risk. We now discuss the variation in the data that identifies each of these quantities in the estimation.

The effect of forecastable risk on plan costs is identified by variation across firms in the average risk scores of employees and dependents, and how it affects insurer bids and realized costs.¹⁶ We identify the markup parameters (δ_j) by the difference between the plan bids and reported costs. A maintained assumption in estimating markups is that insurers base their bids on only the information about employees that is provided by the intermediary. We discuss this assumption more below, but we believe it is reasonable given the small size of the contracts and the fact that we obtain very similar estimates using only the first year of plan bids, when additional information was less likely to be available.

More subtle identification issues arise on the demand side in estimating consumer sensitivity to plan contributions. Plan contributions are the result of plan bids and employer pass-through decisions. Our model allows four sources of variation in contributions: cross-firm variation in demographics (x_f) that leads plans to submit different bids, idiosyncratic variation in plan bids (v_{jf}), cross-firm and cross-tier variation in employer pass-through rates (γ_{jif}), and idiosyncratic variation in the plan contributions (ξ_{jif}).¹⁷

Figure 3 demonstrates this variation by plotting the incremental contributions against the incremental bids for each plan relative to the integrated HMO, which is usually the plan requiring the lowest employee contribution. We plot contribution rates for two tiers, employee only and employee plus spouse, to demonstrate how contributions vary across tier. For the employee plus spouse data, we divide both the contributions and the bids by two to obtain per-enrollee prices. The incremental bid for employee-only coverage for the network PPO ranges from \$50 to \$150 per month, with a large fraction due to cross-firm variation in demographic risk. Combinations of incremental contributions and bids that lie along the 45-degree line in Figure 3 represent employers who pass on the full marginal cost of higher plan bids to employees. A subset of employers adopt this approach. Another subset of employers fully subsidize the higher cost plans, setting incremental contributions of zero. Between these two extremes are employers who partially subsidize higher cost plans through contribution policies. In general, employers tend to pass on a greater portion of incremental costs for plans with dependent coverage.

The availability of multiple sources of variation permits some flexibility in estimating price elasticities. Recall that accurate identification requires using price variation that is not correlated with idiosyncratic household tastes ε_{ij} or privately known health risk μ_{hi} . Our baseline estimates use all four sources of variation. We also employ instrumental variables to isolate different sources of variation. The

¹⁶We have experimented with including demographic covariates in the cost specification but found that it does not improve predictive power. This is not surprising, as the risk score measure already accounts for the effects of age and gender on expected utilization.

¹⁷We also introduce variation in employee contributions through the imputed marginal tax rates, but we control for imputed income and relevant household demographics in the demand equation.

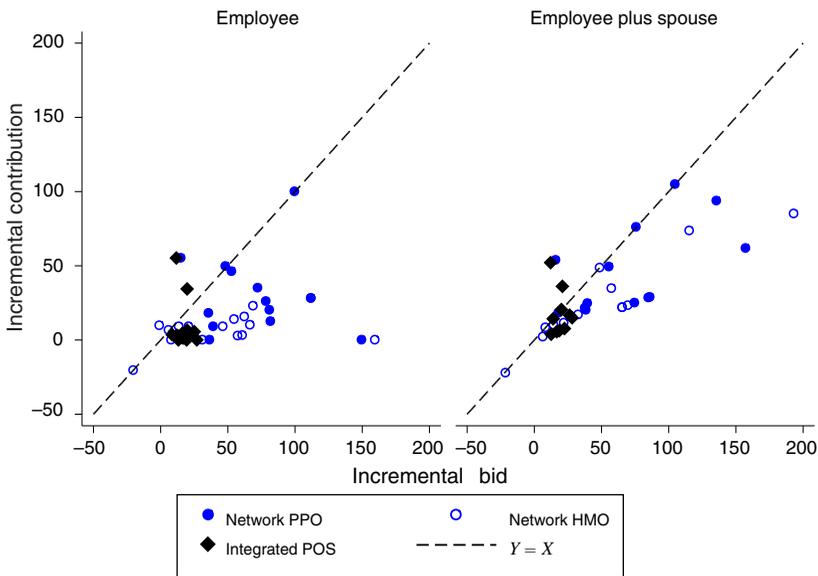


FIGURE 3. CONTRIBUTIONS AND BIDS RELATIVE TO INTEGRATED HMO

Notes: Incremental contribution and incremental bid are relative to integrated HMO. In Employee plus spouse, numbers are divided by two for comparability.

instruments are predicted plan contributions based on alternative covariates. The bottom line from these specifications is that our price elasticity estimates are quite robust to focusing on different sources of variation in contributions. This robustness, despite our relatively small sample, suggests that endogeneity may not be an important concern, at least in this setting. Nevertheless, we now discuss the issues in detail.

Perhaps the most obvious identification concern is that employers believe their employees will prefer a particular plan and price accordingly. This could mean catering to employees with a low contribution or setting a high contribution to pass on costs. Either would generate a correlation between the idiosyncratic part of the contribution ξ_{jif} and household preferences ε_{hj} . To mitigate this concern, we instrument for the actual plan contribution using the predicted value (\hat{p}_{jif}) from the contribution model (10). We take this as our preferred specification in performing welfare analysis, although the results are similar to the baseline case with no instruments.

A second concern is that plan bids are correlated with unobserved household tastes. This could happen if an insurer believed its plan was attractive due to, say, a nearby clinic location. It would generate a correlation between the idiosyncratic bid component, ν_{jf} , and household preferences ε_{hj} . We view this problem as most likely of marginal importance given the limited information on the part of insurers. Nevertheless, we check our estimates by instrumenting for plan contribution with a predicted value that is constructed by plugging the predicted bid \hat{B}_{jf} from (9) into the contribution model (10). This specification purges the variation in both ν_{jf} and ξ_{jif} . The results are similar to our preferred specification.

A third issue for identification is that employer pass-through rates might be systematically influenced by employee preferences. This also seems unlikely, mainly

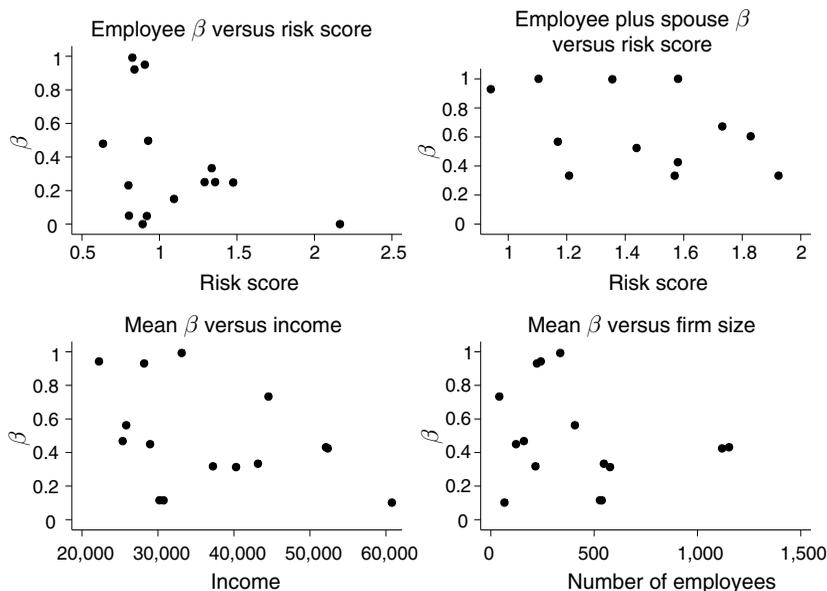


FIGURE 4. EMPLOYER CONTRIBUTIONS AND EMPLOYEE CHARACTERISTICS

Notes: Each point represents a firm-year. β is the incremental pass-through of bids. See text for details. Employee β versus risk score plots the β for employees against their mean risk score. Employee plus spouse β versus risk score plots the β for those in employee plus spouse plans against their risk score. Scatterplots for other coverage tiers look similar. Mean β versus income plots the mean β (across coverage tiers) against mean employee income. Mean β versus firm size plots the mean β (across coverage tiers) against the number of employees in the firm. Separate scatterplots by coverage tier look similar.

because pass-through rates in our data are uncorrelated with *observable* differences across firms. Figure 4 plots employer pass-through rates against employee health status, dependent health status, worker income and firm size. There is no correlation, suggesting that cross-firm differences in contribution policies may be due more to idiosyncratic factors, such as management philosophy, than employee tastes. Nevertheless, we again use an IV strategy to verify that our results are not driven by a correlation between the pass-through coefficients γ_{jlf} and unobserved preferences ε_{hj} . To this end, we instrument for plan contribution using predicted values from a variant of the contribution model (10) in which pass-through coefficients are restricted to be identical across firms. This purges cross-firm variation in γ_{jlf} as well as the variation in ξ_{jlf} . The results are again similar, although with large standard errors.¹⁸

The remaining demand parameters are less troublesome. The effect of household risk on choice behavior (i.e., the coefficients α_{rj} in the demand equation) is identified by variation in observable risk across households. Our model also allows private information about health status to affect choice. The key parameter here is the

¹⁸A final identification concern is that household choices may be influenced by the health status of their co-employees, leading to a correlation between \bar{r}_f and ε_{hj} and, hence, between p_{hj} and ε_{hj} . To check on this issue, we tried including \bar{r}_f as an explanatory variable in our baseline demand model. The results were again similar.

variance of the private information, σ_{μ}^2 . It is identified by the correlation between consumers' enrollment decisions and plans' realized costs. As in a standard selection model, one may be concerned about whether this type of identification is sensitive to our assumption that μ_h is normally distributed. Our identification is strengthened, however, by the variation in contributions discussed above. Because this variation shifts employees across plans but does not affect costs directly, it identifies the cost of households on the margin between plans.¹⁹ This is similar to the usual type of identification from an exclusion restriction in selection models.

C. Estimation Strategy

We estimate the model using the method of simulated moments. A method of moments estimator is useful because it allows us to combine the information in consumer choices, plan costs, and plan bids, each of which is observed at a different level of aggregation. We estimate the employer contribution model separately, by OLS regression, and use it to construct instruments for the plan prices as discussed above.

Our first set of moments come from consumer choice. For each household h , we have

$$(11) \quad \mathbb{E}_{\varepsilon}[q_{hj} - \Pr(q_{hj} = 1 | \mathbf{X}_h, \mu_h) | \mathbf{Z}_h, \mu_h] = 0.$$

In this equation, the \mathbf{X}_h are the household covariates, and \mathbf{Z}_h denotes the same vector with plan contributions replaced by the relevant predicted contributions for the IV specifications. Equation (5) above provides the logit formula for $\Pr(q_{hj} = 1 | \mathbf{X}_h, \mu_h)$.

The second set of moment conditions come from the model of realized insurer costs. For each firm-insurer-year, we have

$$(12) \quad \mathbb{E}_{\eta} \left[C_{kf} - \sum_{j \in \mathcal{J}_{kf}} \sum_{i \in \mathcal{I}_{jf}} \{a_j + b_j \cdot (r_i + \mu_i - 1)\} | \mathbf{X}_{kf}, \mu_{kf} \right] = 0.$$

Here \mathbf{X}_{kf} contains the relevant characteristics of enrollees and plans in the given firm-insurer-year, including the observed risk r_{kf} of insurer k 's enrollees, and μ_{kf} are the unobserved risks of these enrollees.

The final moment conditions come from plan bids. For each firm-plan during a firm's first year of participation, we have

$$(13) \quad \mathbb{E}_{\nu}[B_{jf} - \delta_j \cdot (a_j + b_j \cdot (\mathbb{E}[\bar{r}_f | \mathbf{x}_f] - 1)) | \mathbf{X}_f] = 0.$$

Here \mathbf{X}_f contains the demographic information on firm f available to the insurers.

Each conditional expectation is of the form $\mathbb{E}[h^{\tau}(\theta, \mathbf{X}_n, \mu_n) | \mathbf{Z}_n, \mu_n]$, where θ are the unknown parameters, \mathbf{X}_n are the observables for observation n , \mathbf{Z}_n are instruments, and μ_n the unobserved health risk. We let $\tau = 1, 2, 3$ index the choice,

¹⁹Our demand model also includes plan characteristics such as coinsurance and deductible. Their coefficients are identified off cross-firm and cross-tier variation in the characteristics.

cost, and bid equations, respectively.²⁰ Following the standard GMM approach, we create moments $m^\tau(\theta, \mathbf{X}_n, \mathbf{Z}_n, \mu_n) = \mathbf{Z}'_n \cdot h^\tau(\theta, \mathbf{X}_n, \mu_n)$, with the property that $\mathbb{E}[m^\tau(\theta_0, \mathbf{X}_n, \mathbf{Z}_n, \mu_n)] = 0$. Let $m(\theta, \mathbf{X}, \mathbf{Z}, \boldsymbol{\mu})$ denote the vector obtained by stacking all of the moment conditions. This vector depends on the unobserved health risks, but we can integrate over the distribution of those risks (assumed normal with mean zero and variance σ_μ^2) to obtain $m^\tau(\theta, \mathbf{X}, \mathbf{Z}) = \int m^\tau(\theta, \mathbf{X}, \mathbf{Z}, \mu) dF_\mu(\boldsymbol{\mu})$. Again the stacked moments have the property that $\mathbb{E}[m(\theta_0, \mathbf{X}_n, \mathbf{Z}_n)] = 0$.

In practice, we construct $m(\theta, \mathbf{X}, \mathbf{Z})$ using simulation to approximate the integral. For each individual in the data, we take s draws from F_μ and average across them to obtain $\tilde{m}(\theta, \mathbf{X}_n, \mathbf{Z}_n) = \frac{1}{s} \sum_{s=1}^s m(\theta, \mathbf{X}_n, \mathbf{Z}_n, \mu_{n,s})$. We then obtain parameter estimates in typical fashion by constructing the sample analog $\hat{m}(\theta) = \frac{1}{N} \sum_{n=1}^N \tilde{m}(\theta, \mathbf{X}_n, \mathbf{Z}_n)$ and choosing $\hat{\theta} = \text{argmin}_{\theta \in \Theta} \hat{m}(\theta)' \mathbf{W} \hat{m}(\theta)$. For efficiency, we set $\mathbf{W} = \{\mathbb{E}[\hat{m}(\theta)\hat{m}(\theta)']\}^{-1}$ following the standard two-step process.

D. Welfare Measurement

We use the estimated model to compare market allocations and social welfare under alternative contribution policies. Here we explain briefly these calculations. For any given set of plan prices, household choice probabilities and expected plan costs can be computed using the above formulas so long as the private risks (μ_s) are known. As we do not observe μ , we integrate over its distribution by taking simulation draws for each individual and later averaging over these draws.

Changes in social welfare are calculated in similar fashion. The expected change in the money-metric utility of household h following a price change from \mathbf{p} to \mathbf{p}' is

$$(14) \quad \Delta U_h(\mathbf{p}, \mathbf{p}') = \int n_h \cdot \left\{ \ln\left(\sum_{j \in J} \exp(v_j(p'_{hj}))\right) - \ln\left(\sum_{j \in J} \exp(v_j(p_{hj}))\right) \right\} dF_\mu(\mu_h),$$

which is the formula derived by Small and Rosen (1981), scaled by the number of members in each household n_h and integrated over private risk μ .

To calculate changes in producer surplus, it is convenient to treat the employer and the insurers together, netting out transfers between them. The change in the producer surplus following a price change from \mathbf{p} to \mathbf{p}' for household h is

$$(15) \quad \Delta \Pi_h(\mathbf{p}, \mathbf{p}') = \int \left\{ \sum_{j \in J} \Pr(q_{hj} = 1 | p'_{hj}, \mu_h)(p'_{hj} - c_{hj}^e) - \sum_{j \in J} \Pr(q_{hj} = 1 | p_{hj}, \mu_h)(p_{hj} - c_{hj}^e) \right\} dF_\mu(\mu_h),$$

where c_{hj}^e is the expected cost of covering household h in plan j .

With these pieces in place, the overall change in social welfare is

$$(16) \quad \Delta S(\mathbf{p}, \mathbf{p}') = \sum_h \{ \Delta U_h(\mathbf{p}, \mathbf{p}') + \Delta \Pi_h(\mathbf{p}, \mathbf{p}') \}.$$

²⁰We slightly abuse notation by letting n index observations on choices, costs and bids, despite the fact that the level of aggregation is different for each equation, and, hence, we have different numbers of observations.

To calculate ΔS we draw values of μ for each individual in the data, calculate $\Delta U_h(\mathbf{p}, \mathbf{p}', \mu_h)$ and $\Delta \Pi_h(\mathbf{p}, \mathbf{p}', \mu_h)$ for each simulation draw, and average over the draws to obtain $\Delta U_h(\mathbf{p}, \mathbf{p}')$ and $\Delta \Pi_h(\mathbf{p}, \mathbf{p}')$. Adding up across households yields the desired quantities.

Below, we also solve for prices that are optimal given various constraints (e.g., not risk rated, risk rated based on observable risk, etc.) To do this, we nest the social welfare calculation inside a gradient-based optimization routine in Matlab, solve for optimal prices, and then use a grid search to check for global optimality.

IV. Empirical Results

In this section, we report estimates of the model parameters and calculations of market allocation and social welfare under alternative pricing policies and choice sets.

A. Model Estimates

Table 4 presents parameter estimates from three different specifications of the demand model.²¹ The first column is a baseline model where we do not instrument for plan contributions and do not allow for private information about household risk. The second and third columns instrument for plan contributions using the predicted values from the contribution model (10). The third column, which is our preferred specification, allows for private information about risk. To scale the utility to money-metric form, we divide each coefficient by the coefficient on the monthly contribution and adjust the standard errors accordingly. We report the price effects as semi-elasticities at the bottom of the table.

Effect of Demographics and Risk on Choice.—The demand estimates show a number of relationships between demographics and plan preferences. Conditional on risk score, older employees prefer the network HMO and the integrated POS plan to the integrated HMO. Women prefer the integrated HMO to either the integrated POS plan or the network PPO and are willing to pay \$35 per month less than men for the network PPO relative to the integrated HMO (column 1). Women may have stronger preferences for the integrated HMO if they perceive that it is more effective in providing preventive care as, in this age group, more preventive services are recommended for women than for men. The effects of age and sex are not particularly sensitive to the use of instruments for the employee contribution (column 2) or the incorporation of unobserved risk (column 3).

The demand estimates indicate that sorting on the basis of risk is rather modest, and to the extent it exists is driven primarily by having a very high-risk household member. The effects of the linear risk score on plan choice are generally small and imprecise. Households with a high-risk member, however, are less likely to enroll in the network HMO and more likely to enroll in the network PPO than the integrated HMO. In our preferred specification (column 3), an employee with a high-risk

²¹ The table does not report every parameter. The parameters not reported are the plan fixed effects, and the coefficients on imputed household income and an indicator for nonstandard drug coverage.

TABLE 4—DEMAND MODEL

	Non-IV (1)		IV (2)		IV and private risk (3)	
<i>Rescaled coefficients</i>						
Contribution	-1.00	(0.28)	-1.00	(1.28)	-1.00	(1.43)
Coinsurance (percent)	1.91	(0.49)	1.41	(2.0)	2.31	(1.28)
Deductible	-0.01	(0.02)	-0.01	(0.02)	0.01	(0.03)
NHMO						
× Risk score	-1.24	(3.40)	-0.92	(2.04)	-1.59	(1.58)
× Age	1.75	(0.27)	1.27	(0.24)	1.82	(0.30)
× Female	4.93	(9.18)	4.34	(8.34)	7.20	(9.86)
× High risk	-21.27	(12.76)	-15.14	(7.05)	-17.17	(5.59)
NPPO						
× Risk score	-11.07	(6.76)	-8.32	(4.37)	-3.93	(2.64)
× Age	0.75	(0.45)	0.54	(0.49)	0.51	(0.48)
× Female	-34.64	(14.36)	-26.36	(7.66)	-32.44	(9.54)
× High risk	49.38	(19.87)	36.89	(11.11)	28.11	(8.78)
IPOS						
× Risk score	-6.10	(5.41)	-4.44	(2.90)	-5.29	(2.15)
× Age	1.58	(0.39)	1.15	(0.24)	1.56	(0.35)
× Female	-35.24	(12.85)	-25.54	(10.52)	-32.63	(10.55)
× High risk	16.40	(17.28)	12.23	(8.81)	9.43	(7.30)
σ_ϵ	109.29	—	79.26	—	102.33	—
σ_μ					0.68	(0.65)
Observations	3,683		3,683		3,683	
Semi-elasticities						
NHMO	-0.09		-0.09		-0.07	
NPPO	-0.10		-0.05		-0.05	
IHMO	-0.05		-0.13		-0.10	
IPOS	-0.09		-0.09		-0.07	
Total	-0.07		-0.09		-0.07	

Notes: Table reports estimates from the logit demand model: the dependent variable is an indicator for the plan chosen, and IHMO is the baseline choice. Specifications (2) and (3) use predicted contributions as an instrument, and (3) allows for privately known risk. The estimated coefficients (including the standard deviation of the logit error, which is not an estimated parameter) are rescaled so the coefficient on the monthly contribution is one. Contribution is in tax adjusted dollars and coinsurance is in percentage points. Plan fixed effects, income and a dummy variable for nonstandard prescription drug coverage are included but not shown. Semi-elasticities are the change in market share for a hundred-dollar increase in the annual premium, calculated as $(100 \times \text{Marginal effect}) / (12 \times \text{Market share})$.

family member is willing to pay \$28 per month more than an employee without a high-risk family member to enroll in the network PPO relative to the integrated HMO. This is consistent, once again, with those who are more likely to use care placing a greater value on less restrictive provider networks.

Our results also suggest that private information about health risk plays a role in plan choice, although the estimate is not precise. We estimate that the standard deviation of private risk information σ_μ is 0.68, which is substantial relative to the standard deviation of the observed risk scores (1.56 in Table 1). Roughly speaking, observed risk scores appear to pick up just over two-thirds of the health status information that factors into plan choice.

While our findings with respect to risk selection are not inconsistent with existing research, we do not find the sharp selection of high-risk employees into more generous plans reported in some studies (Cutler and Zeckhauser 2000; Glied 2000). We do find that the highest-risk enrollees favor the most flexible plan, the network PPO,

but the average risk across plans is quite similar. This finding is consistent with the idea that the plans in our setting cater to individuals with different tastes for health care delivery, rather than offering different quantities of care, or targeting individuals of different health status.

Effect of Plan Prices on Choice.—In the bottom panel, we present price semi-elasticities of demand, defined as the percentage decrease in market share resulting from a \$100 increase in the annual enrollee contribution, evaluated at the mean choice probability for each plan. On average, a \$100 increase in the annual enrollee contribution decreases market share by 7 to 9 percent. While instrumenting for the contribution reduces the precision of the estimate, it has relatively little effect on its magnitude. These estimates suggest that demand is relatively inelastic, in line with the literature. In the online Appendix, we discuss studies in settings similar to ours. Across these studies, a \$100 increase in the annual contribution reduces market share by 1.6 to 9.6 percent; our estimate is toward the higher end of this range.

The results in Table 4 also include the estimated value of plan characteristics other than price, such as coinsurance rate and deductible. Enrollees appear to be moderately sensitive to the coinsurance rate. We estimate that a 10 percentage point increase in the coinsurance rate is valued at approximately \$276 annually, which is about 10 percent of the annual cost per enrollee reported by the insurers. Our estimates indicate that enrollees are not particularly sensitive to the deductible when choosing among plans.²²

Because the estimates of risk and price elasticity are the key parameters for our welfare calculations, we have examined the sensitivity of these estimates to a variety of issues. In the online Appendix we present estimates where we vary the sample of households and use different instruments (discussed in Section IIIB) for the employee contributions. We also discuss specifications with alternative sets of controls. The bottom line is that the estimates are robust to variation across these dimensions.

Structure of Plan Costs.—The difference in cost structures for the integrated and network insurer can be seen in the raw data depicted in Figures 5 and 6. Figure 5 plots average enrollee risk scores against realized costs for each insurer-firm-year. The lines represent the model's prediction (based on the estimates in Table 5) of expected costs for the network PPO and the integrated HMO. Figure 6 shows the average risk of a firm's employees plotted against plan bids, with each observation at the plan-firm-year level.²³

The plans appear to have similar costs for enrollees with average health risk and divergent costs for enrollees in good and poor health. The expected monthly cost for an enrollee with a risk score of 1 is \$235 for the integrated HMO, \$236 for the integrated POS, \$218 for the network HMO, and \$238 for the network PPO. For less healthy enrollees, the integrated insurer has a substantial cost advantage. The expected monthly cost of an enrollee with a risk score of two is \$309 for the

²² The results are unchanged when out-of-pocket maximums are included as plan characteristics.

²³ Figures 5 and 6 show that there is an outlying firm with a higher risk score than the others. In online Appendix Table A2, we show that the estimates of the costs model are very similar when this firm is dropped from the sample.

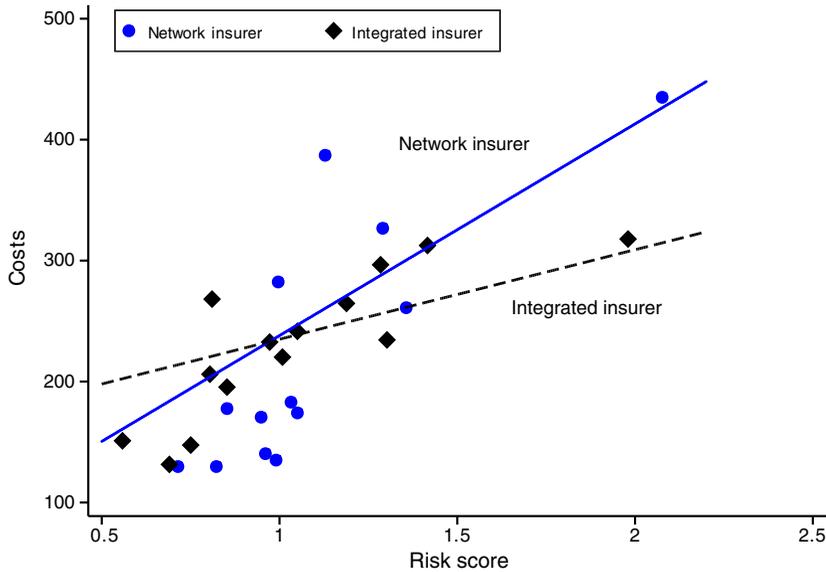


FIGURE 5. COSTS BY RISK SCORE

Notes: Each point represents an insurer-firm-year. Risk score is the average enrollee risk score in a firm-year. Costs is the insurers' monthly cost per enrollee. Fitted lines represent the network HMO and integrated HMO.

TABLE 5—COSTS AND BIDS

	(1)		(2)	
Network insurer markup	1.29	(0.12)	1.27	(0.07)
Integrated insurer markup	1.08	(0.04)	1.07	(0.03)
NHMO	218.42	(21.35)	195.08	(9.48)
× (Risk score – 1)	288.25	(86.05)	265.36	(30.29)
× Coinsurance			0.32	(0.94)
NPPO	238.32	(22.65)	204.59	(9.82)
× (Risk score – 1)	174.92	(34.34)	167.73	(23.81)
× Coinsurance			1.23	(1.29)
IHMO	234.86	(9.71)	228.77	(6.77)
× (Risk score – 1)	73.67	(22.01)	104.80	(18.80)
× Coinsurance			0.38	(0.54)
IPOS	236.37	(14.13)	216.74	(13.01)
× (Risk score – 1)	188.64	(69.35)	200.78	(61.60)
× Coinsurance			1.65	(0.82)
Observations	91		91	

Notes: GMM estimates of cost parameters. See text for details. Coinsurance is de-measured at the plan level.

integrated HMO, compared to \$507 for the network HMO and \$413 for the network PPO. Network plans are relatively low cost for low risks. The expected monthly cost for an enrollee with a risk score of 0.75 is \$216 for the integrated HMO, as opposed to \$146 for the network HMO and \$195 for the network PPO.

The structure of plan costs we estimate is consistent with the basic idea that patient cost sharing may be effective at limiting provider visits, while supply-side

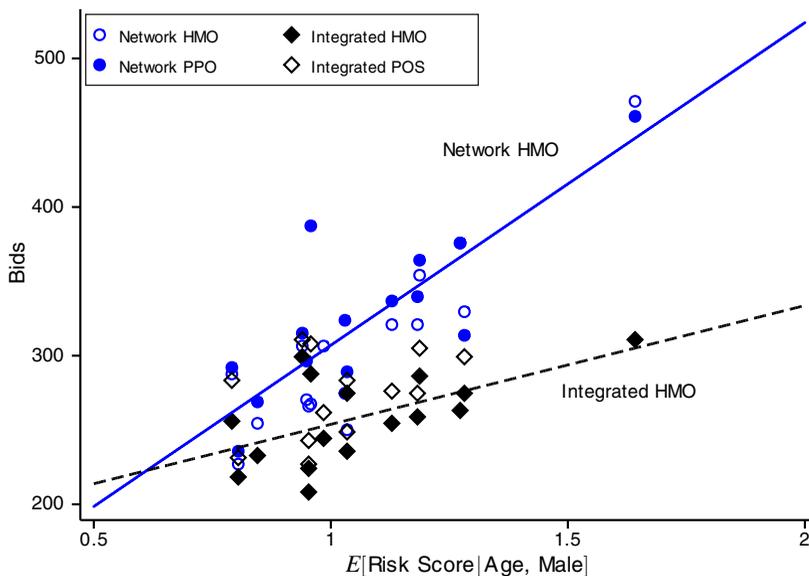


FIGURE 6. BIDS BY RISK SCORE

Notes: Each point represents a plan-firm-year. $E[\text{Risk} | \text{Age, Male}]$ is the average predicted risk score of potential enrollees of an employer. Bids is the per-month bid. Fitted lines represent the network HMO and integrated HMO.

mechanisms may be more effective at limiting costs conditional on receiving services (see, e.g., Newhouse and The Insurance Experiment Group 1993). While we do not have visit-level data to support the claim, the steep cost curves for the network plans are consistent with cost sharing limiting visits, particularly for low risks, but having little effect on the high risks who consume health care on the intensive margin. In contrast, the integrated plans with their relatively low cost sharing but stronger supply-side utilization controls may be less effective at limiting provider visits for low risks but more effective at managing costs conditional on provider visits for the high risks. Another factor explaining the relatively high costs for low risks in the integrated plan may be the greater use of preventive services. We also note that the estimated markup of bids over expected costs varies across insurers: 24 percent for the network insurer and 8 percent for the integrated insurer.

The sensitivity of cost differentials as a function of enrollee risk, compared to the relatively modest effect of risk on plan preferences, has an important implication. It indicates that as consumer risk varies, changes in relative plan costs rather than changes in preferences will drive the efficient allocation. As our simple theory explained, this will not happen under self-selection unless different risk groups face different premium differentials. In our setting, prices do not have this feature, suggesting the potential for inefficiency.

The interpretation of our cost estimates depends on the risk score's being an accurate measure of individual health status. As discussed above, we view this as a reasonable assumption. If one were to question it, perhaps the main concern would be that reported risk scores are too low for enrollees in the integrated plan due to its

more aggressive management of drug utilization. If this were the case, the actual health of the integrated plan enrollees would be worse than the data suggest, and we would be underestimating the plan's cost advantage.

Two specific features of our cost estimates are a bit surprising and would be interesting to explore with additional data. One is that our estimates for the integrated POS plan costs are closer to those of the network plans than to the integrated HMO, although the POS estimate is somewhat imprecise. We do not have data to indicate whether enrollees in this plan utilize nonintegrated services, which would help to illuminate this. A similar point is that for high-risk enrollees, the network HMO does not generate cost savings relative to the PPO product, although again this is not statistically significant. One's initial guess might be that the network HMO has lower costs for all risk types. It's possible that this finding is driven by our relatively small data sample on costs, which necessitates rather strong functional form assumptions.

A further factor to keep in mind when evaluating our estimates of plan costs is that we observe the *insurers'* costs of coverage, not the overall dollars spent on care. The distinction is important because, in plans with copayments and deductibles, enrollees bear a share of the cost of care that we do not capture in our data. These payments are largest at the network PPO and smallest at the integrated HMO. While our model assumes that these payments will be internalized in making plan choices, they do affect the interpretation of the effects of the different plan types on utilization of care. In particular, the reduction in insured costs for low risks in the network plan may represent, at least in part, a shift from insured to uninsured payments, rather than a reduction in utilization. For high risks, in contrast, the difference in insured costs between the plans likely underestimates the extent to which the integrated plan reduces total costs.

B. *Quantifying Social Welfare Inefficiencies*

In this section, we use the estimated demand and cost model to compute the inefficiency associated with observed contribution policies relative to alternative efficient benchmarks. We also compare welfare between the observed policies and alternative uniform contribution policies to demonstrate the extent to which the inefficiency associated with a uniform contribution could be reduced within the current institutional constraints.

To get some intuition, Figure 7 provides a graphical summary of the model that parallels Figure 1. To construct the figure, we simulate the unobserved risk μ and idiosyncratic preference ε for each individual in the data and plot the joint distribution of risk scores and willingness-to-pay for the network HMO relative to the integrated HMO. As the figure shows, consumers are relatively dispersed in terms of willingness-to-pay, so small price changes do not result in major market share shifts. Figure 7 also shows the incremental cost of coverage for the network HMO and the average incremental contribution, as well as the misallocation that results from setting premiums uniformly across risk types. In the next section we translate this misallocation into quantitative efficiency costs.

The Welfare Cost of Observed Prices.—To obtain numerical welfare estimates, we account for all four plans and simulate consumer choices and realized costs under different pricing regimes. Table 6 presents the results of these simulations.

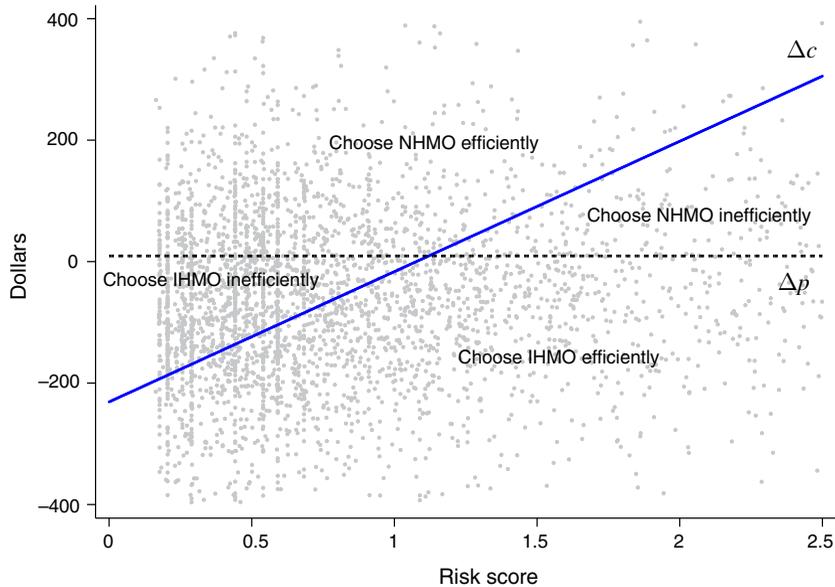


FIGURE 7. PREFERENCES AND COSTS BY RISK SCORE

Notes: Figure shows the distribution of preferences and costs by risk score. Each point represents a household. For these points, the y -axis value is the incremental willingness-to-pay for the network HMO relative to the integrated HMO, and the x -axis value is the household's risk score. The solid line (Δc) shows the incremental cost, and the dashed line (Δp) shows the incremental contribution for the network HMO relative to the integrated HMO. The points are calculated using the demand estimates from Table 3, column 3, with the μ_s and ε_s simulated from the estimated distributions. The solid line is calculated using the cost estimates from Table 4, column 2. The dashed line shows the average incremental premium across firm-years.

The left-hand columns present the market share, average enrollee risk, and the average incremental contribution for each plan under five different pricing scenarios. The incremental contribution represents the monthly contribution per enrollee relative to the integrated HMO averaged across all households. The right-hand columns present information on the change in surplus relative to the observed allocation for each scenario.

We compare welfare from the observed pricing policies to two risk-rated benchmarks. The first is individual risk rating based on the observed risk scores, or “feasible risk-rated contributions.” This involves setting contributions to maximize social welfare conditional on knowledge of the risk scores but not each household's private information. This contrasts to “optimal risk-rated contributions” which use both public and private risk information to obtain a first-best allocation.

Overall, under risk-rated contributions, high-risk households face higher premiums and low-risk households face somewhat lower premiums for the network plans relative to observed contribution policies. In both the feasible and first-best scenarios, this leads to a substantial reallocation of enrollees across plans, although overall market shares change modestly. With feasible risk rating, the average enrollee risk at the integrated HMO increases from its observed level of 0.99 to 1.49, and the network HMO experiences a decline in average enrollee risk from 1.03 to 0.58. This

TABLE 6—MATCHING AND WELFARE UNDER ALTERNATIVE CONTRIBUTION POLICIES

	Matching				Welfare ^a			Truncated
	NHMO	NPPO	IHMO	IPOS	Gross surplus ^b	Insurer costs ^b	Social surplus ^b	Social surplus ^b
Observed								
Market shares	0.25	0.09	0.54	0.12	0.00	0.00	0.00	0.00
Risk score	1.03	1.07	0.99	1.02				
Incremental contribution	9.30	23.70	0.00	5.00				
Feasible risk-rated contributions								
Market shares	0.37	0.09	0.43	0.11	-16.60	-43.70	27.10	5.00
Risk score	0.58	0.78	1.49	0.74				
Incremental contribution	-14.70	11.80	0.00	-1.30				
Optimal risk-rated contributions								
Market shares	0.38	0.08	0.44	0.10	-22.10	-57.50	35.50	7.80
Risk score	0.60	0.79	1.46	0.76				
Incremental contribution	-14.90	11.80	0.00	-1.60				
Uniform by tier within firms								
Market shares	0.31	0.09	0.49	0.12	-6.10	-12.80	6.70	1.40
Risk score	0.86	1.02	1.11	0.97				
Incremental contribution	-16.50	8.90	0.00	-1.10				
Enthoven rule								
Market shares	0.22	0.08	0.58	0.13	-1.10	-0.80	-0.30	-0.50
Risk score	1.01	1.05	1.00	1.02				
Incremental contribution	28.70	39.90	0.00	10.80				

Notes: Feasible risk-rated contributions implements efficient matching by setting incremental contributions equal to incremental costs, conditional on observable risk but not privately known risk. Optimal risk-rated contributions sets incremental contributions equal to incremental costs, conditional on both observable and privately known risk. Uniform by tier within firms maximizes social surplus subject to the constraint that contributions vary only by coverage tier and by firm, but not by individual risk. Enthoven rule is implemented by setting incremental contributions equal to incremental bids. Reported risk score is conditional on plan choice. The truncated results hold cost differentials between plans for risk scores lower than 0.75 and higher than 2.0 at these boundary levels.

^aIncremental contribution, gross surplus, insurer costs, and social surplus are averaged across enrollees and denominated in \$ per month.

^bGross surplus, insurer costs, and social surplus are normalized to zero under the observed allocation. Other scenarios show gross surplus as social surplus relative to the observed allocation. Under the observed allocation, costs average \$241.70 per enrollee per month. Gross and social surplus are not pinned down.

TABLE 7—MATCHING AND WELFARE BY RISK SCORE QUINTILE

Quintile (risk score range)	Feasible risk-rated contributions versus observed							
	Matching				Welfare			Truncated
	NHMO	NPPO	IHMO	IPOS	Δ Gross surplus	Δ Insurer costs	Δ Social surplus	Δ Social surplus
Quintile 1 (< 0.36)								
Δ Market share	0.332	0.000	-0.330	-0.002	-27.2	-56.9	29.8	4.3
Δ Incremental contribution	-179.4	-93.4	0.0	-86.6				
Quintile 2 (0.36, 0.54)								
Δ Market share	0.265	0.003	-0.266	-0.001	-16.6	-35.6	18.9	3.4
Δ Incremental contribution	-141.6	-75.9	0.0	-65.6				
Quintile 3 (0.54, 0.79)								
Δ Market share	0.181	0.006	-0.189	0.002	-7.7	-17.1	9.3	1.3
Δ Incremental contribution	-99.1	-53.4	0.0	-44.6				
Quintile 4 (0.79, 1.33)								
Δ Market share	0.040	0.004	-0.037	-0.007	-0.8	-2.4	1.6	0.4
Δ Incremental contribution	-21.0	-19.7	0.0	-1.2				
Quintile 5 (> 1.33)								
Δ Market share	-0.184	-0.047	0.299	-0.069	-30.3	-105.9	75.6	15.4
Δ Incremental contribution	324.8	154.5	0.0	179.3				
Total								
Δ Market share	0.128	-0.007	-0.106	-0.015	-16.6	-43.8	27.1	5.0
Δ Incremental contribution	-23.9	-11.9	0.0	-6.3				

Notes: Δ Market share, Δ incremental contribution, Δ gross surplus, Δ insurer costs, and Δ social surplus are calculated as the difference between the feasible risk-rated and observed values of these variables. Truncated fixes cost differentials between plans for risk scores outside of 0.75 and 2.0. Values averaged across enrollees within each quintile and denominated in \$ per month. (Total values are averaged across all enrollees.)

reallocation of households across plans substantially reduces overall insurer costs, by \$44 per enrollee-month, and increases total social surplus by just over \$27 per enrollee-month. The increase in social welfare represents approximately 11 percent of average insurer costs in our sample.

A substantial fraction of the welfare gain is due to the highest and lowest-risk households making more efficient plan choices. Table 7 decomposes the welfare calculation by household-risk quintiles. The lowest- and highest-risk quintiles (average household risk below 0.36 and above 1.33) generate about three-quarters of the welfare effect. This raises a concern that our calculation might be driven in part by extrapolating plan costs out of sample. As Figures 5 and 6 illustrate, we observe plan bids and costs only for average risk scores between 0.75 and 2.0. In contrast, household risk ranges from 0.16 to 30.1. To address this, we truncate the cost differentials between plans at their 0.75 and 2.0 levels and recalculate the welfare numbers. These calculations appear in the final columns of Tables 5 and 6. We view the numbers based on truncated cost differences as a lower bound on welfare differences, and the baseline numbers based on straight-line extrapolation as closer to an upper bound. Truncating the cost differentials has little effect on the resulting assignment of households to plans, but, as one might expect, it reduces the welfare cost of observed pricing to \$5 per enrollee-month, or 2 percent of insurer costs, relative to the feasible optimum.

It is also interesting to compare what is possible using prices based on observed risk scores to what, in principle, could be achieved using both observed risk scores *and* households' private information. This calculation captures the extent to which private information on risk constrains the efficiency of feasible relative to optimal risk-rated pricing. Changing from feasible risk-rated contributions to the first-best scenario increases social surplus by between \$2 and \$8 per enrollee-month, depending on the treatment of costs for extreme risk, or roughly 1–3 percent of insurer costs. One way to interpret this is that, in our sample, a social planner could achieve approximately 70 percent of the potential welfare gains associated with individualized pricing using only observable information on risk.

Social Welfare without Risk Rating.—The calculations above indicate that the observed prices fall well short of the efficient benchmark. A natural question is whether efficiency gains could be realized even without risk-rated contributions. That is, to the extent reallocating high- and low-risk households would increase social welfare, is it possible to induce this reallocation given current institutional pricing constraints? At first glance, the answer is unclear. After all, current institutions require uniform pricing within firm-tiers, but this still allows a fair amount of pricing flexibility within our sample. For example, average risk varies substantially across the firms in our data, suggesting that cross-firm variation in contribution policies could alleviate some of the inefficiency associated with uniform contributions.

The next scenario in Table 6 addresses the question of what is possible without individualized pricing by considering contributions that maximize social welfare subject to being uniform within each firm-tier. As in the case of fully risk-rated prices, optimizing uniform within-firm contributions leads to a reallocation of high-risk households into the integrated plans and away from the network plans, particularly the PPO. The shift is much less dramatic, however, than under full risk rating.

Overall social surplus is \$1.40–\$6.70 higher per enrollee-month than under the observed policies, but still \$3.60–\$20.40 below the efficient level. This indicates that about three-quarters of the observed inefficiency is due to the requirement of nondiscriminatory pricing within firms. Nevertheless, it appears that employers could increase social surplus by around 1–3 percent of average insurer costs simply by adjusting their contributions to better reflect differences in underlying plan costs.

One difficulty for employers, of course, is that matching contributions to plan costs may be a fairly complex exercise. Many benefits consultants, including the intermediary in our data, suggest a simpler approach, which is to pass on the full incremental premium for all but the lowest priced plan. We refer to this as the “Enthoven Rule” (Enthoven and Kronick 1989). About half of the firms in our sample use this approach for at least some workers. The last entry in Table 6 considers the effect of moving all the firms to an Enthoven-style approach. Perhaps surprisingly, this has little effect on overall welfare or household choices. The reason is that demand is not very price elastic and from a practical standpoint most firms already pass through a substantial fraction of the premium differentials. So relative to the price changes needed to move substantial numbers of households across plans, a change to an Enthoven policy has only a modest impact.

This last observation raises an important point for our pricing experiments. The relatively low elasticity of demand means that the contribution differentials needed to reallocate households in the direction of efficiency are sizable. For instance, maximizing welfare while keeping contributions uniform within firm-tiers would lead to some households seeing an \$87 per-enrollee monthly premium for the network PPO relative to the integrated HMO. A move to efficient risk-rated prices would increase this differential even more for some high-cost households. For instance, an individual employee with a risk score of 3 would face a monthly premium differential of between \$101 and \$202 depending on our cost extrapolation. These large price differentials indicate that achieving efficient allocations may raise issues of fairness or affordability of coverage for particular subgroups.

C. The Value of Plan Choice

By choosing to offer benefits through the intermediary, each of the firms in our sample moved from offering a single health plan to offering multiple plans from two carriers. A clear benefit of plan choice is that households with different preferences can select their preferred plan (Bundorf 2010). Our estimates indicate a substantial amount of preference heterogeneity and, hence, suggest substantial welfare gains from giving households multiple plan options.

To illustrate this, Table 8 compares aggregate surplus under the observed offerings to the surplus that would be obtained if all the households in our sample were enrolled in one of the four plans. The most natural benchmark is the integrated HMO, as it would be the most efficient single-plan offering for every firm in our data. Relative to the integrated HMO benchmark, the observed plan offerings increase social welfare by almost \$70 per enrollee-month for the firms in our data. Virtually all of this is due to an increase in gross surplus rather than to a reduction in insurer costs. Indeed, insurer costs would be lowest if all households were enrolled

TABLE 8—THE VALUE OF PLAN CHOICE

	Welfare		
	Gross surplus	Insurer costs	Social surplus
Observed	0.0	0.0	0.0
All enrolled in:			
NHMO	-148.8	-9.2	-139.7
NPPO	-216.9	5.8	-222.7
IHMO	-71.4	-2.1	-69.4
IPOS	-180.7	4.5	-185.2

Notes: Gross surplus, insurer costs, and social surplus are averaged across enrollees and denominated in \$ per month. These values are normalized to zero under the observed allocation. Other scenarios show gross surplus as social surplus relative to the observed allocation. Under the observed allocation, costs average \$241.70 per enrollee per month. Gross and social surplus are not pinned down.

in the network HMO, but the reduction in social surplus would be large due to the reduction in gross surplus.

One caveat to the exact numbers in this calculation is that the logit demand specification is notorious for generating large “new product” welfare gains. Roughly speaking, the problem is that each new product adds a new preference dimension, and some households invariably enjoy a large welfare gain from this addition due to the logit distributional assumption. So while we think that preference heterogeneity and, hence, the benefits of plan choice is quite a robust finding, we urge caution in interpreting the exact magnitudes in Table 8.

D. Discussion and Sensitivity Analysis

Our estimates of market inefficiencies are based on a particular set of employers in a particular geographic area. One way to address external validity is to compare our estimates with some other studies of specific environments, such as Cutler and Reber (1998); Carlin and Town (2008); and Einav, Finkelstein, and Cullen (2010). These studies all rely on data from individual large employers, and in each case, the plans are plausibly distinguished by their level of generosity, making the environments a bit different from ours. All three studies find evidence that more generous plans are adversely selected. Cutler and Reber document this by using enrollee age as a proxy for risk. The latter two studies, like ours, use data on realized costs.

Despite the difference in institutional settings, the bottom line welfare estimates from these studies are fairly similar to our estimates. Cutler and Reber estimate that observed prices at Harvard University reduce welfare by around 2–4 percent of coverage costs relative to optimal uniform prices. Einav, Finkelstein, and Cullen estimate that in their setting average cost pricing has a welfare cost of roughly 2 percent relative to optimal uniform pricing. Carlin and Town find much smaller welfare effects, due to very low demand elasticity.²⁴ Note that these papers all focus

²⁴One explanation for their inelastic demand estimate is that they rely on time-series variation in contributions. As discussed above, employees appear to be more price sensitive in making initial choices than in making changes once they are enrolled.

on uniform pricing, which we have noted is generally inefficient except in special cases. When we use our estimates to compare observed pricing to optimal uniform prices, we find welfare costs of approximately 1–3 percent of coverage costs. In this sense, there appears to be a fair amount of agreement between studies.

As a group, these studies also reinforce our earlier observation that inefficiencies from pricing can be driven both by the nature of sorting and risk selection, and by the price elasticity of demand, which determines the extent to which implicit subsidies or taxes affect choices. To gauge the sensitivity of our own estimates to these factors, we recalculated the surplus difference between the observed and the feasible efficient allocation assuming that demand was twice as sensitive to price as we have estimated, and half as sensitive. We performed a similar analysis varying the risk sensitivity of demand. These analyses increase the range of the welfare gains to 1–13 percent of total coverage costs. Given the range of demand estimates in the literature, one may want to assign a corresponding range of uncertainty to the potential welfare costs of price distortions.

One simplified feature of our modeling approach is that we allow individual health status to vary only along a single health status dimension (the sum of the observed risk score and private health information). In practice, individuals may differ not just in expected health costs, but in the variance of these costs, and in their need for particular types of chronic care. A model with richer household heterogeneity could provide further insights into individual choices and plan incentives, perhaps with significant implications for sorting and efficiency.

V. Risk-Adjusted Pricing and Reclassification Risk

A potential concern with risk-rated premiums is that households can face reclassification risk if their premiums adjust annually along with their health status. The amount of risk depends on both the persistence of health shocks and the nature of risk rating. For instance, if health shocks are completely persistent and prices actuarially fair, households will see each health shock reflected in future premiums. On the other hand, if individual risk scores evolve in a predictable fashion and unexpected health shocks are transitory, reclassification risk may be minimal. Moreover, the type of risk rating we are envisioning is a bit different from what is usually discussed because it requires risk adjustment of only incremental premiums. It therefore seems possible to provide substantial intertemporal insurance by maintaining a uniform contribution for a base plan and risk-adjusting incremental contributions above this base.

To address these issues, we combine our model with data on risk-score transitions and calculate the potential variation in premiums and welfare cost of reclassification risk under three different pricing regimes: uniform pricing of plans independent of health risk, full risk rating where employee premiums for each plan are set equal to the household's expected costs under the plan, and incremental risk rating where the integrated HMO plan has a uniform price, but the premiums for the other plans are adjusted to reflect each household's expected incremental costs.²⁵

²⁵ We use a single uniform price structure for this analysis, calculated as the values that maximize social surplus for the sample.

TABLE 9—PREMIUMS AND UTILITY RISK UNDER ALTERNATIVE CONTRIBUTION POLICIES

	Mean within-household standard deviation				
	Premiums		Money-metric utility		
	NHMO (1)	IHMO (2)	Dominating preference for NHMO (3)	Dominating preference for IHMO (4)	Estimated preferences (5)
Uniform	0.00	0.00	8.22	3.55	3.57
Feasible risk rated	84.45	21.58	87.47	21.41	39.48
Feasible risk rated with uniform IHMO	62.87	0.00	66.04	3.14	20.13

Notes: Table shows means of the within-household standard deviations in premiums and money-metric utility under alternative contribution policies, denominated in \$ per enrollee per month. Columns 1 and 2 show premiums for the network and integrated HMO plans. Columns 3 and 4 show the variation in money-metric utility when households having a dominating preference for these plans. Column 5 shows the variation in money-metric utility under estimated preferences from the demand model. Uniform sets contributions for each plan equal to the constant value that maximizes social surplus. Feasible risk rated sets contributions equal to estimated costs. Feasible risk rated with uniform IHMO holds the IHMO contribution constant while setting the incremental contributions for the other plans equal to their incremental costs.

We restrict attention for this exercise to the 930 households that we observe in consecutive years.²⁶ For these households we can model the evolution of risk scores. About 17 percent of households have constant risk scores across years. For the households that switch, a log-log model with a normally distributed error fits the data quite well.²⁷ Simulating risk score transitions from this model, and combining them with our model of plan costs, we calculate the distribution of possible contributions for each employee under each pricing regime.

Columns 1 and 2 of Table 9 report the average within-household standard deviation of employee contributions under each regime. We show values for the two HMO plans that together account for over 80 percent of the market. As in our earlier tables, values are denominated in dollars per enrollee-month and cost differentials for risk scores outside of 0.75 to 2.0 are truncated to avoid relying on projections outside of the variation in the data. When all plan premiums are risk rated, households face on average an \$84 standard deviation in their monthly premium for the network HMO plan, and \$22 for the integrated HMO.²⁸ The variance is lower for the integrated plan because its expected costs are less sensitive to the risk score. The variation is reduced if the premium for the integrated HMO is set uniformly and only the incremental premiums above this base are risk adjusted. Then there is no variation in the integrated HMO premium, and an average standard deviation across households of $\$84 - \$22 = \$62$ for the network HMO.

This variation somewhat overstates the risk to households because they have the opportunity to choose across plans. To incorporate this, we simulate household

²⁶ We also exclude households that change coverage tiers or number of family members over this period. The mean risk score in this sample is slightly higher than average. Because the variance is increasing in the mean, this suggests that the variation in risk scores is slightly higher in this sample than in the entire data.

²⁷ The model $\ln r_{h,t+1} = \alpha_0 + \alpha_1 \ln r_{h,t} + \epsilon_h$ has an R^2 of 0.49. We estimate coefficients (standard errors) of $\alpha_0 = -0.006$ (0.026) and $\alpha_1 = 0.708$ (0.025). The standard deviation of the error term is $\sigma_\epsilon = 0.592$. We experimented with and rejected models that allow for heterogeneity in the coefficients and heteroskedasticity in the error term, although the results that follow are very similar under alternative risk-score specifications.

²⁸ For comparison, actuarially fair employee contributions average \$51 and \$40 for these plans.

choices and money-metric utility under the different pricing regimes. Columns 3 and 4 report the variation in money-metric utility for households that have a dominating preference for a particular plan and always choose it. Even with uniform premiums these households face a small amount of risk because utility varies directly with health status, but the numbers are very similar to columns 1 and 2. In column 5, we simulate health realizations and optimal choices for all households in the sample. With risk rating for all plans, the within-household standard deviation of money-metric utility averages \$39 per enrollee per month. This drops roughly in half, to \$20 per enrollee-month, if the integrated HMO contribution is uniform and only the incremental premiums are risk adjusted.

The importance of this variation depends on household risk attitudes. To this end, we can think of the realized utility following a household's health shock and choice of plan as the (monetary) payoff from a lottery, and ask how much households would be willing to pay to insure this risk.²⁹ Specifically, for each household h we can solve for the risk premium π_h that solves

$$\int u\left(\max_{j \in J} \{v_{hj}(r_h)\}\right) dF_{r_h} = u\left(\int \max_{j \in J} \{v_{hj}(r_h)\} dF_{r_h} - \pi_h\right),$$

where $\max_{j \in J} \{v_{hj}(r_h)\}$ is the household's money-metric utility from its most preferred plan under the given pricing regime, F_{r_h} is the one-year-ahead distribution of risk scores for the household, and $u(\cdot)$ is a concave function that captures aversion to risk.

We choose a constant absolute risk aversion specification for $u(\cdot)$. To be conservative, we parameterize the coefficient of absolute risk aversion to be 9.16×10^{-5} , which is equivalent to a constant relative risk aversion parameter of 4 when divided by mean income of \$43,669 and is near the top of the range of estimates in the literature (see Cohen and Einav 2007). With the integrated HMO plan priced uniformly, and incremental plan prices risk adjusted, we calculate that households would be willing to pay on average \$0.45 per enrollee per month to insure against one-year reclassification risk.³⁰ The number increases to \$1.19 per enrollee per month if all plan prices are risk adjusted. These numbers are relatively small compared to the static efficiency gain of \$7.80 per enrollee per month that we estimated from risk adjusting incremental (or all) plan prices.

These calculations suggest that the costs of reclassification risk are modest compared to the efficiency benefits of risk adjusting incremental plan contributions. Of course, our sample population is relatively young and employed. For an older, less healthy population, the variance and persistence of health shocks could be greater. As a rough way of assessing the sensitivity of our estimates, we tried increasing the variance of health shocks in our calculations by a factor of 10. This raises the cost

²⁹ We focus on a one-year time horizon both because we can observe only one-year risk-score transitions in the data and because issues such as firm turnover and changes in family structure, which are import shocks for our sample population, are outside the scope of our analysis. We note, however, that the mean reversion in the risk scores suggests that the long-run cost of reclassification is likely to be less than the discounted sum of the one-year values.

³⁰ We use the one-year risk because our static efficiency calculations are also for a single year. In principle, one could imagine doing a more involved exercise with a longer term horizon. This probably would require starting to think about savings decisions, and labor market turnover, among other factors.

of reclassification risk to \$1.15 per enrollee month with risk-rated incremental premiums, or about 15 percent of the static efficiency gains.³¹ Although we would be hesitant to draw conclusions too far out of sample, it seems potentially possible to implement risk adjustment of incremental prices without creating excessive reclassification risk.

VI. Conclusion

Economists have long understood that competition in health insurance markets is no guarantee of efficiency. This paper contributes to a nascent literature that attempts to quantify market inefficiencies and identify their sources. We find that observed contribution policies do distort enrollment decisions from their efficient level, creating a welfare loss on the order of 2–11 percent of the total cost of coverage. Capturing these gains in full would require the use of risk-rated contribution policies. Absent such policies, optimally set employee contributions might increase welfare by 1–3 percent of coverage costs. Despite these distortions, there do appear to be gains from allowing plan choice because of heterogeneous preferences for different plans.

A substantive idea that emerges from our analysis is the possibility of risk-adjusting incremental plan prices to encourage efficient self-selection. In our empirical context we estimated that the short-term reclassification risk under this type of arrangement might be relatively modest. It would be interesting to explore whether there are other ways to provide dynamic insurance while encouraging households to match efficiently into plans, perhaps through the use of longer-term insurance contracts (e.g., Cochrane 1995).

REFERENCES

- Akerlof, George A. 1970. "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism." *Quarterly Journal of Economics* 84 (3): 488–500.
- Bundorf, M. Kate. 2010. "The Effects of Offering Health Plan Choice within Employment-Based Purchasing Groups." *Journal of Risk and Insurance* 77 (1): 105–27.
- Bundorf, M. Kate, Jonathan Levin, and Neale Mahoney. 2012. "Pricing and Welfare in Health Plan Choice: Dataset." *American Economic Review*. <http://dx.doi.org/10.1257/aer.102.7.3214>.
- Cardon, James H., and Igal Hendel. 2001. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *RAND Journal of Economics* 32 (3): 408–27.
- Carlin, Caroline, and Robert Town. 2008. "Adverse Selection, Welfare and Optimal Pricing of Employer-Sponsored Health Plans." Unpublished.
- Cochrane, John H. 1995. "Time-Consistent Health Insurance." *Journal of Political Economy* 103 (3): 445–73.
- Cohen, Alma, and Liran Einav. 2007. "Estimating Risk Preferences from Deductible Choice." *American Economic Review* 97 (3): 745–88.
- Cutler, David M., Amy Finkelstein, and Kathleen McGarry. 2008. "Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance." *American Economic Review* 98 (2): 157–62.
- Cutler, David M., and Sarah J. Reber. 1998. "Paying for Health Insurance: The Trade-Off between Competition and Adverse Selection." *Quarterly Journal of Economics* 113 (2): 433–66.

³¹ The increase in risk premiums is relatively small because households can substitute towards the integrated HMO plan after receiving a large health shock. For households with such a strong network HMO preference that they would never substitute to another plan, the risk premium is \$7.74, fully offsetting the gains from static risk rating.

- Cutler, David M., and Richard J. Zeckhauser.** 2000. "The Anatomy of Health Insurance." In *Handbook of Health Economics Volume 1A*, edited by Anthony J. Culyer and Joseph P. Newhouse, 563–643. Handbooks in Economics, vol. 17. Amsterdam: Elsevier Science, North-Holland.
- Enthoven, Alain, and Richard Kronick.** 1989. "A Consumer-Choice Health Plan for the 1990s." *New England Journal of Medicine* 320 (1): 29–37.
- Einav, Liran, Amy Finkelstein, and Mark R. Cullen.** 2010. "Estimating Welfare in Insurance Markets Using Variation in Prices." *Quarterly Journal of Economics* 125 (3): 877–921.
- Einav, Liran, Amy Finkelstein, and Jonathan Levin.** 2010. "Beyond Testing: Empirical Models of Insurance Markets." *Annual Review of Economics* 2 (1): 311–36.
- Feldman, Roger D., and Bryan E. Dowd.** 1982. "Simulation of a Health Insurance Market with Adverse Selection." *Operations Research* 30 (6): 1027–42.
- Glied, Sherry.** 2000. "Managed Care." In *Handbook of Health Economics Volume 1A*, edited by Anthony J. Culyer and Joseph P. Newhouse, 707–53. Handbooks in Economics, vol. 17. Amsterdam: Elsevier Science, North-Holland.
- Handel, Benjamin.** 2012. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." Unpublished.
- Kaiser Family Foundation.** 2007. "Employer Health Benefits: 2007 Annual Survey." <http://www.kff.org/insurance/7672/upload/76723.pdf> (accessed June 1, 2008).
- Miller, Nolan H.** 2005. "Pricing Health Benefits: A Cost-Minimization Approach." *Journal of Health Economics* 24 (5): 931–49.
- Neipp, Joachim, and Richard J. Zeckhauser.** 1985. "Persistence in the Choice of Health Plans." In *Advances in Health Economics and Health Services Research Volume 6*, edited by Richard M. Scheffler and Louis F. Rossiter, 47–72. Greenwich, CT: JAI Press.
- Newhouse, Joseph P.** 1996. "Reimbursing Health Plans and Health Providers: Efficiency in Production versus Selection." *Journal of Economic Literature* 34 (3): 1236–63.
- Newhouse, Joseph P., and The Insurance Experiment Group.** 1993. *Free for All? Lessons from the RAND Health Insurance Experiment*. A RAND Study. Cambridge, MA: Harvard University Press.
- Pauly, Mark V., and Bradley J. Herring.** 2000. "An Efficient Employer Strategy for Dealing with Adverse Selection in Multiple-Plan Offerings: An MSA Example." *Journal of Health Economics* 19 (4): 513–28.
- Rothschild, Michael, and Joseph E. Stiglitz.** 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics* 90 (4): 630–49.
- Small, Kenneth A., and Harvey S. Rosen.** 1981. "Applied Welfare Economics with Discrete Choice Models." *Econometrica* 49 (1): 105–30.
- van de Ven, Wynand P. M. M., and Randall P. Ellis.** 2000. "Risk Adjustment in Competitive Health Plan Markets." In *Handbook of Health Economics Volume 1A*, edited by Anthony J. Culyer and Joseph P. Newhouse, 755–845. Handbooks in Economics, vol. 17. Amsterdam: Elsevier Science, North-Holland.
- Zhao, Yang, Randall P. Ellis, Arlene S. Ash, David Calabrese, John Z. Ayanian, James P. Slaughter, Lori Weyuker, and Bruce Bowen.** 2001. "Measuring Population Health Risks Using Inpatient Diagnoses and Outpatient Pharmacy Data." *Health Services Research* 36 (6): 180–93.