Pricing and Welfare in Health Plan Choice*

M. Kate Bundorf, Jonathan Levin and Neale Mahoney†

April 2010

Abstract

Premiums in 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 study this problem in health insurance markets. We develop a simple model and estimate it using data on small employers. In this setting, the welfare loss compared to the feasible risk-rated benchmark is around 2-11% of coverage costs. Three-quarters of this is due to restrictions on risk-rating employee contributions; the rest is due to inefficient contribution choices. Despite the inefficiency, we find substantial benefits from plan choice.

*We thank Amy Finkelstein and Will Manning for helpful suggestions. Levin gratefully acknowledges support from the National Science Foundation and the Center for Advanced Study in the Behavioral Sciences.

†Bundorf: School of Medicine, Stanford University, and NBER, bundorf@stanford.edu; Levin: Department of Economics, Stanford University and NBER, jdlevin@stanford.edu; Mahoney: Department of Economics, Stanford University, nmahoney@stanford.edu.

1 Introduction

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 U.S. private market, employers generally 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 U.S. 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 U.S., regulations prohibit employers and public programs from charging enrollees different amounts based on nearly all observable health-related factors. 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 paper, 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.[1] We show that if these assumptions are violated, it becomes very difficult to have a contribution policy

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

that leads to efficient consumer choices.

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. Based on our simulations, we estimate that, in this setting, 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% of coverage costs. Employers could realize approximately 1/4 of this surplus by adjusting their non-risk-rated contributions, but capturing the remainder would require setting different prices for people in the same firm. We also find that employees select plans based on private information about their health status. A hypothetical social planner who incorporated this private information into prices could increase welfare by an additional \$35-\$100 annually per enrollee. Despite these inefficiencies, the observed plan offerings generate substantial benefits over any single-plan offering.

The nature of the offered health plans is important for interpreting these results. One has a fairly broad network of providers 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. Our estimates suggest that these insurers have very different cost structures. While costs are similar between the two for individuals of average health, the integrated delivery system has somwhat higher costs for healthy enrollees and significantly lower costs for less healthy enrollees. We also find that both household preferences and health status affect the choice between the plans. In contrast to some other studies, however, we do not observe any single plan experiencing serious adverse selection. One possible explanation is that the plans are not ordered in a clear way by coverage level: instead, consumers face a trade-off between provider network restrictions and the degree of cost sharing.

The "horizontal" differentiation of health plans makes our setting somewhat different from many other studies of health insurance demand. The type of differentiation in our data, however, 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 a mix of supply-side and demand-side utilization management (KFF, 2007). This evolution suggests that classic insights based on purely risk-based sorting may not adequately capture the dynamics of today's market.

The inefficiencies we find are driven by two forces: heterogeneity in household preferences and the significant cost advantage of the integrated system for individuals in worse health. Our estimates suggest that although high risk households have some preference for flexibility, a large fraction would choose the integrated delivery system if they had to internalize the relevant cost differential between plans. Achieving this with a uniform contribution policy, however, would require that all households face a steep premium for the more flexible insurer. This in turn would create a welfare loss for lower risk households who value flexibility. While the exact numbers we estimate are of course specific to our setting, the trade-off we identify may be relevant more broadly. This analysis suggests that the integrated model of health care delivery faces an important challenge in health insurance markets: current pricing policies make it difficult to target households for whom the integrated model promises substantial cost savings.

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 paper 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, 2007; Handel, 2009; Einav, Finkelstein and Cullen, 2010; and Einav Finkelstein and Levin, 2010). 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. 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 5.4.

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 the last section. 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 based on a static model, so we abstract from issues of dynamic insurance. We discuss this issue in the conclusion.

2 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. Each consumer chooses between a high-cost plan A and a low cost plan B. We can think of the plans, for now, as vertically differentiated. The plans' expected costs of covering a type-θ consumer are $c_A(\theta)$ and $c_B(\theta)$. Let $\Delta c(\theta) = c_A(\theta) - c_B(\theta)$ denote the cost differential. We assume Δc is strictly positive and increasing in θ.

Let $v_A(\theta)$ and $v_B(\theta)$ denote a type θ's expected (dollar) value from being covered by each of the plans. For the moment, the benefits of coverage are determined only by forecastable health risk. We assume that contributions vary across plans but not across consumers. A consumer who makes a contribution p_j to enroll in plan $j \in \{A, B\}$ gets a net benefit $v_j(\theta) - p_j$.[2] Define $\Delta v(\theta) = v_A(\theta) - v_B(\theta)$ to be the additional amount a type-θ consumer would pay for the high-cost plan.

[2] 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.

The efficient assignment places a type-θ consumer in plan A if and only if

$$\Delta v(\theta) - \Delta c(\theta) \geq 0. \tag{1}$$

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

$$\Delta v(\theta) - \Delta p \geq 0, \tag{2}$$

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

Are there prices that lead to an efficient outcome? Assume that $\Delta v(\theta)$ is increasing in θ, which seems appropriate if plan A simply offers more generous coverage or easier access to care. Then for any incremental contribution Δp, a type-θ consumer will choose A if and only if $\theta \geq \overline{\theta}(\Delta p)$, where $\overline{\theta}(\Delta p)$ is a threshold that can be varied arbitrarily with Δp.[3] Therefore it is possible to achieve efficient sorting if and only if the efficient assignment also involves a threshold rule, i.e. if $\Delta v(\theta) - \Delta c(\theta)$ is negative up to some θ and positive above it. Roughly, the requirement for efficiency is that willingness to pay increases more quickly with risk than the cost differential between plans.

Existing analyses assume that the surplus function has the requisite single crossing property (e.g. Feldman and Dowd, 1982; Cutler and Reber 1998; Cutler and Zeckhauser 2000; Miller 2005). In this case, shown in Figure 1(a), efficiency can be achieved by setting $\Delta p = \Delta c(\theta)$. Setting this efficient policy does not even require knowledge of individual health risk (i.e. each consumer's θ), although it does require knowing the market demand and cost curves (Einav et al., 2010). The problem emphasized in the literature is that purchasers may not choose the correct premium differential. If Δp is too high, plan A attracts only very high risks, and if prices are based on past outcomes, one can even end up with an adverse selection "death spiral" (Cutler and Reber, 1998). Alternatively, if Δp is too low, too many people select plan A, including some for whom the benefits are less than the incremental social costs.

This familiar analysis can be questioned on several levels. First, even if we continue to assume that consumers differ only in their health status and plans differ mainly in the amount of coverage they offer, it could be socially efficient for high risks to be in a cost-

[3] An empirical prediction of this model is that plan A will experience unfavorable selection, and its risk composition will be worse the larger is Δp.

conscious plan.[4] For instance, the cost savings from a plan that actively manages utilization might more than compensate these consumers for the loss of flexibility. In this case, shown in Figure 1(b), uniform pricing is inherently inefficient because there is no way to induce only the relatively high risks to self-select into plan B when everyone faces the same prices.

Perhaps the more obvious issue from an empirical perspective is that health plans are often differentiated well beyond offering "more" or "less" coverage, and consumer health risk does not perfectly proxy for consumer preferences. For instance, it is increasingly common for firms to offer employees an HMO option that places greater restriction on provider choice, and a PPO option that allows broader access to provider, with greater cost sharing. Consumers in relatively poor health may value flexibility but be wary of increased cost-sharing. As a result, heterogeneity in tastes or income may be at least as important as health status in driving choice.

To capture this, think of plan A as a PPO and plan B as an HMO, and suppose that consumers vary in both forecastable risk and taste. Specifically, let ε denote a consumer's preference for flexibility, so that $\Delta v(\theta, \varepsilon)$ is the extra willingness to pay for plan A. A consumer of type (θ, ε) is efficiently assigned to plan A if and only if $\Delta v(\theta, \varepsilon) - \Delta c(\theta) > 0$ and chooses plan A if and only if $\Delta v(\theta, \varepsilon) - \Delta p$. Except in very special cases, it is clear that an efficient allocation will be very difficult to achieve. Indeed, if consumers cannot be priced on the basis of their tastes, and tastes matter sufficiently, achieving efficient sorting requires measuring and pricing health risk so that consumers of type θ face a contribution differential $\Delta c(\theta)$.

The potential inefficiency under a uniform pricing policy is shown in Figure 2. The figure assumes that Δv is strictly increasing in θ as well as ε, and the curve $\Delta v(\theta, \varepsilon) = \Delta p$ depicts consumers who are just indifferent between plans for a given contribution differential. Consumers above and to the right choose the PPO; those below and to the left choose the HMO. Similarly, $\Delta v(\theta, \varepsilon) = \Delta c(\theta)$ defines the set of consumers who are marginal in the efficient allocation. Figure 2(a) shows a situation where, holding tastes constant, higher risk consumers are efficiently assigned to the PPO. Figure 2(b) shows the reverse situation where higher risk consumers are efficiently assigned to the HMO. In both cases, there is a critical

[4] Arguably the benefits of delivering care efficiently may be largest for the chronically ill. The most detailed analysis of differences in utilization between traditional Medicare coverage and Medicare managed care plans found that the reductions in utilization generated by managed care plans were concentrated among high risk beneficiaries and that these reductions in utilization were not associated with differences in short term health outcomes (Brown 1993).

type θ^0 (defined so that $\Delta c(\theta^0) = \Delta p$) that faces an actuarily fair price differential for the plans and sorts efficienctly. Consumers with risk types above θ^0 are effectively subsidized to choose the PPO and some do so inefficiently. Consumers with risk types below θ^0 face too high a price premium for the PPO and some inefficiently opt for the HMO.

These figures suggest some straightforward observations. First, the potential inefficiencies depend on whether $\Delta c(\theta)$ varies with consumer risk. If $\Delta c(\theta)$ does not vary much, one can approximate the efficient risk-adjusted contribution with a uniform contribution. Second, the potential inefficiencies also depend on the price and risk sensitivity of consumers. If consumer demand is relatively elastic, the welfare gain from risk-rating contributions may be substantial. If demand inelastic, pricing distortions simply may not matter much for consumer choices. Finally, the nature of consumer heterogeneity is important. For example, in Figure 2, it is clear that the welfare effects depend on how consumers are distributed across the various regions of risk/preference space.

Each of these issues is inherently quantitative in nature. We use the econometric model developed in the next section to identify the relevant cost and demand parameters and then evaluate empirically how various pricing arrangements affect social welfare.

3 Data and Environment

3.1 Institutional Setting

Our analysis is based on data from a private firm that helps small and mid-sized employers manage health benefits. This firm, who 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 non-exclusively 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

[5] This insurer also offers a point-of-service (POS) plan that is the HMO with the option to go out-of-

that facilitates supply side utilization management. 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 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

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.

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

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 under-estimate 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.

3.2 Data and Descriptive Statistics

Our data includes 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.

[7] In 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.

[8] DxCG 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. (2005) for more detail.

Another useful feature of the data is that we observe each employer during their 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 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. 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 forty. 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 the average. The average risk scores of employees and enrollees are 1.25 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.

Panel A of 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

[9] In 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.

plan. On average, the integrated HMO is the least expensive plan and has the lowest enrollee contribution. This plan features high rates of coinsurance, 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. While bids for each plan vary substantially across tiers, reflecting differences in expected expenditures, the bids for tiers other than employee only are simply scaled by a factor that is very similar across both 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 Panel B. The integrated HMO attracts by far the most enrollees with a 59% market share among employees and 60% market share among enrollees. We also find little evidence of extensive risk selection across the plans. The integrated HMO attracts a slightly younger population, and women, particularly women employees, disproportionately choose the network and integrated HMOs. But the differences across the plans in both average age and average risk score are small. This lack of sorting is not driven by heterogeneity across firms in the choice sets. 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.

4 Econometric Model

4.1 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. It should be noted that by costs we mean overall costs to the insurer for a given enrollee in a given plan. Although we discuss factors that may explain the variation in costs below, overall cost is sufficient for welfare analysis and it is not necessary to decompose whether these cost differences arise from, for example, moral hazard or physician reimbursement rates or some other factor (c.f. Einav et al., 2010).

In contrast to the simple theoretical model discussed 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 identify 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 a combination of household and plan characteristics. Specifically, household h's utility from choosing plan j is:

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

In this representation, household utility depends on observable plan characteristics ϕ_j, the monthly plan contribution p_j,[10] 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_is capture information about health status that may affect choice behavior, but is 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 (i.e. logit errors).

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 household members.[11] 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 90% percentile of the observed

[10] We convert employee contributions, which are made with pre-tax 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 $p_h = (1 - \tau_h)\rho_h$.

[11] 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 Appendix includes individual enrollee estimates.

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.[12]

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 non-standard drug coverage.[13] 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.

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,

$$q_{hj} = 1 \iff u_{hj} \geq u_{hk} \text{ for all } k \in \mathcal{J}_h. \tag{4}$$

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 $v_{hj} = u_{hj} - \varepsilon_{hj}$, and let X_{hj} denote the full set of relevant observables, we have the familiar formula for choice probabilities:

$$\Pr\left(q_{hj} = 1 \mid X_{hj}, \mu_h\right) = \frac{\exp\left(v_{hj}\right)}{\sum_{k 2 J_h} \exp\left(v_{hk}\right)}. \tag{5}$$

Health Plan Costs

We model each plan j's cost of enrolling a given individual as a function 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

[12] 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 $73,00 and 41%, respectively.

[13] While the prescription drug coverage for each plan is complicated, comprised of both formulate restrictions 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, the coverage is less generous.

cost of enrolling individual i as

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

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 forecas 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:

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

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 the 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 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:

$$\frac{1}{|\mathcal{I}_f|}\sum_{i \in \mathcal{I}_f} \mathbb{E}[c_{ij}|x_i] = a_j + b_j(\mathbb{E}[\overline{r}_f|x_f] - 1), \tag{8}$$

where $\overline{r}_f$ denotes the average risk of employees in firm f, which the insurer forecasts using the available demographic information, x_f.[14]

We model expected plan bids as a mark-up over expected cost. So plan j's bid for firm f is:

$$B_{jf} = \delta_j \cdot (a_j + b_j \cdot (\mathbb{E}[\overline{r}_f|x_f] - 1)) + \nu_{jf}, \tag{9}$$

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

Employer Contribution Setting

The last part of our model specifies how employers set required 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_{jlf}. We model the required contribution as:

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

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-squared is 0.99. As noted above, 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 $\beta < \gamma$.

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

4.2 Discussion of Model and Identification

The key quantities in our model are plan costs and plan demand as functions of forecastable risk, and the price elasticity of demand. The former determine the efficient allocation of households to plans, while the latter determines how price changes affect self-selection. We now discuss the variation in the data that identifies each of these quantities in estimation.

Identifying plan costs is straightforward. The effect of forecastable risk on plan costs is identified by variation across firms in the average risk scores of workers and dependents, and how it affects insurer bids and realized costs. We identify the mark-up parameters, δ_j, by the difference between the plan bids and reported costs. A maintained assumption in estimating mark-ups 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 consider only the first year of plan bids.

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 variance of the private information, σ_μ^2, which is identified by the correlation between consumers' enrollment decisions and plans' realized costs. This identification is aided by cross-firm variation in contribution policies and demographics that, conditional on observable health risk, affect enrollment but not realized costs.[15] The identification obtained from price variation is similar to the identification in standard selection models, and relies on the exclusion restriction if a given individual i is enrolled in a given plan j, his or her utilization does not depend on the per-month premium (although of course it may depend on other elements of the plan such as the copayment rate).

The most subtle identification issues arise in estimating the effect of plan contributions on demand. 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 (ν_{jf}), cross-firm and cross-tier variation in employer pass-through rates (γ_{jlf}), and idiosyncratic variation in the plan contributions (ξ_{jlf}).[16]

[15] Our 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.

[16] We also introduce variation in employee contributions through the imputed marginal tax rates, but we

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. Difference in the bids for the integrated HMO and 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_{hj} or privately known health risk μ_h. Our baseline estimates use all four sources of variation. We also employ instrumental variables to isolate different sources of variation. The 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_{jlf} and household preferences ε_{hj}. To mitigate this concern, we instrument for the actual plan contribution using the predicted value ($\hat{p}_{jlf}$) 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.

---

control for imputed income and relevant household demographics in the demand equation.

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_{jlf}. 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 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.[17]

4.3 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.

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

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

$$\mathbb{E}_{\varepsilon}\left[q_{hj} - \Pr\left(q_{hj} = 1 \mid X_h\right) \mid Z_h, \mu_h\right] = 0. \tag{11}$$

In this equation, the X_h are the household covariates, and 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 $\Pr\left(q_{hj} = 1 | X_h, \mu_h\right)$.

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

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

Here 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:

$$\mathbb{E}_{\nu}\left[B_{jf} - \delta_j \cdot \left(a_j + b_j \cdot \left(\mathbb{E}[\overline{r}_f | x_f] - 1\right)\right) \mid X_f\right] = 0. \tag{13}$$

Here X_f contains the demographic information on firm f available to the insurers.

Each conditional expectation is of the form $\mathbb{E}\left[h^{\tau}(\theta, X_n, \mu_n) \mid Z_n, \mu_n\right]$, where θ are the unknown parameters, X_n are the observables for observation n, Z_n are instruments and μ_n the unobserved health risk. We let $\tau = 1, 2, 3$ index the choice, cost and bid equations, respectively.[18] Following the standard GMM approach, we create moments $m^{\tau}\left(\theta, X_n, Z_n, \mu_n\right) = Z_n' \cdot h^{\tau}\left(\theta, X_n, \mu\right)$, with the property that $\mathbb{E}[m^{\tau}\left(\theta_0, X_n, Z_n, \mu_n\right)] = 0$. Let $m\left(\theta, X, Z, \mu\right)$ denote the vector obtained by stacking all of 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 $m\left(\theta, X, Z\right) = \int m^{\tau}\left(\theta, X, Z, \mu\right) dF_{\mu}(\mu)$. Again these moments have the property that $\mathbb{E}[m\left(\theta_0, X_n, Z_n\right)] = 0$.

In practice, we construct $m(\theta, X, Z)$ using simulation to approximate the integral. For

[18] We 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.

each individual in the data, we take s draws from F_μ and average across them to obtain $\tilde{m}(\theta, X_n, Z_n) = \frac{1}{S}\sum_{s=1}^{S} m\left(\theta, X_n, Z_n, \mu_{n,s}\right)$. We then obtain parameter estimates in typical fashion by constructing the sample analogue $\hat{m}(\theta) = \frac{1}{N}\sum_{n=1}^{N} \tilde{m}(\theta, X_n, Z_n)$, and choosing $\hat{\theta}$ =argmin$_{\theta \in \Theta}\hat{m}(\theta)' W \hat{m}(\theta)$. For efficiency, we set $W = \left\{\mathbb{E}\left[\hat{m}(\theta)\hat{m}(\theta)'\right]\right\}^{-1}$ following the standard two-step process.

4.4 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 easily 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, calculating choice probabilities and expected plan costs, and then averaging over simulation 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 p to p' is:

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

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 the various transfers between them. The change in the producer surplus resulting from choices by household h:

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

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

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

$$\Delta S\left(p, p'\right) = \sum_h \left\{ \Delta U_h\left(p, p'\right) + \Delta \Pi_h\left(p, p'\right) \right\}.$$

To calculate ΔS in practice, we draw values of μ for each individual in the data, calculate

$\Delta U_h\left(p, p^0, \mu_h\right)$ and $\Delta \Pi_h\left(p, p^0, \mu_h\right)$ for each simulation draw, and average over the draws to obtain $\Delta U_h\left(p, p^0\right)$ and $\Delta \Pi_h\left(p, p^0\right)$. 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.

5 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.

5.1 Model Estimates

Table 3 presents parameter estimates from three different specifications of the demand model.[19] 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.

E¤ect of Demographics and Risk on Choice

The demand estimates indicate that overall sorting on the basis of risk is rather modest, but that different plans experience unfavorable selection across differing components risk. Older employees, who on average cost more to insure, prefer the network HMO and the integrated POS plan to the integrated HMO. An additional year of age is associated with an increase in the willingness to pay for the network HMO relative to the integrated HMO of \$1.75 per month (Column 1). Because older people are often in worse health, they are likely to place a higher value on the broader provider network of the network plan which would give them greater freedom in choosing among providers. Women, who at the age of workers

[19] 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 non-standard drug coverage.

in our data typically cost more to insure than men, prefer the integrated HMO to either the integrated POS plan or the network PPO. Women 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 cares since, 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).

We find some sorting on the basis of health status conditional on age and sex, 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 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 2/3 of the health status information that factors into plan choice.

While our findings with respect to risk selection are not inconsistent with existing research, they suggest a relatively complex pattern of sorting. Much of the existing literature finds evidence of unfavorable selection into more generous plans (Cutler and Zeckhauser, 2000; Glied, 2000). We also find that the highest risk enrollees favor the most flexible plan, the network PPO. Overall, however, the average risk across plans is quite similar due to off-setting selection along different demographic dimensions, including age and gender, that are correlated with risk. This finding is consistent with the idea that the plans cater to individuals with different tastes for health care, rather than offering different quantities of care, or targeting individuals of different health status.

E¤ect 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 dollar 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 other estimates in 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 3 also include the estimated value of plan characteristics other than price, such as coinsurance rate and deductible. Enrollees appear to be moderately sensitive to both. We estimate that a 10 percentage point increase in the coinsurance rate is valued at approximately \$276 dollars 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.[20]

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 4.2) 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 is a scatterplot of enrollee risk scores against realized costs. Each point corresponds to an insurer-firm-year. The x-axis is the average risk of the insurer's enrollees; the y-axis is the reported costs per enrollee. The lines represent the model's prediction (shown in Table 4) of expected costs for the network PPO and the integrated HMO. Figure 6 displays corresponding information for bids. It shows the average risk of a firm's employees plotted against plan bids, with each observation at the plan-firm-year level.

[20] The results are unchanged when out-of-pocket maximum are included as plan characteristics.

As the figures illustrate, the plans seem 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 integrated HMO, compared to \$507 for the network HMO and \$413 for the network PPO. Network plans do relatively well for low risks. The expected monthly cost for an enrollee with a risk score of 0.5 is \$198 for the integrated HMO, as opposed to \$151 for the network PPO and \$74 for the network HMO.

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 mechanisms may be more effective at limiting costs conditional on receiving services (see, e.g., Newhouse 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 healthcare 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 mark-up 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 model illustrated, this will not happen under self-selection without a mechanism that allows different risk groups to face different premium differentials. In our setting, prices do not have this feature, suggesting the potential for inefficiency. We return to this point in the next section, when we quantify social welfare.

Note that the interpretation of these results depends on the risk score being an accurate measure of individual health status. As discussed above, we view this as a reasonable assumption. If one were to question it, probably the main concern would be that risk scores are pushed down for enrollees in the integrated plan due to its more aggressive management of drug utilization. If this type of bias were present, the actual health of the integrated enrollees would be worse than the risk scores suggest. We would then be understating the cost advantage of the integrated plan.

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 non-integrated 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 or prices. 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.

5.2 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. Table 5 presents the results of these simulations. The left-hand panels 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 panels present information on the change in surplus relative to the observed allocation for each scenario.

The Welfare Cost of Observed Prices

In the top panels, we calculate the inefficiency of observed pricing policies relative to two risk-rated benchmarks. The first is individual risk rating based on the observed risk scores. This pricing policy, which we refer to as "feasible risk-rated contributions", maximizes social welfare conditional on knowledge of the risk scores, but not each household's private information. The third panel of the Table reports outcomes when prices are first-best, i.e. risk-rated based on both public and private information.

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 re-allocation 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 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% 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 6 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 re-calculate 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% 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, of nondiscriminatory pricing within firms. Nevertheless, it appears that employers could increase social surplus by around 1-3% 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 1/2 of the firms in our sample use this approach for at least some workers. The last entry in Table 5 considers the effect of moving all the firms to an Enthoven-style approach. Perhaps surprisingly, this has relatively little effect on overall welfare, or on 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 effect on choices.

This last observation raises an important point for our pricing experiments. The relatively low elasticity of demand means that the contribution differentials needed to re-allocate households in the direction of efficiency are sizeable. For instance, maximizing welfare while keeping contributions uniform within firm-tiers would lead to some households seeing a $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.

5.3 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. 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 7 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 consumer surplus (gross of plan contributions) rather than to a reduction in insurer costs. Indeed, insurer costs would be lowest if all households were enrolled in the network HMO but the reduction in social surplus would be large due the reduction in consumer surplus.

One caveat to this calculation is 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 the benefits of plan choice are real, we urge some caution in interpreting the magnitude of the measured effect.

5.4 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 (2007), 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, and also similar to our estimates. Cutler and Reber estimate that observed prices at Harvard University reduce welfare by around 2-4% of coverage costs relative to optimal uniform prices. Einav et al. estimate that in their setting average cost pricing has a welfare cost of roughly 2% relative to optimal uniform pricing. Carlin and Town find much smaller welfare effects, due to very low demand elasticity.[21] Note that these

[21] One explanation for their inelastic demand estimate is that they rely on time-series variation in contri-

papers all focus 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% 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 guage 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% 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.

6 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. A main finding is that observed contribution policies distort enrollment decisions from their efficient level. We calculate that the welfare loss is on the order of 2-11% 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% of coverage costs. A key point to emphasize is that despite these distortions, there appear to be substantial gains in our context from introducing plan choice at the employer level.

One important point about risk-rated premiums in health insurance is that coverage is typically purchased on an annual basis. While risk-rating might increase static efficiency, it exposes households to reclassification risk as their health status changes over time. This is one argument for community rating or nondiscrimation in employer contributions. It is interesting to ask whether there are ways to promote static efficiency that nevertheless

butions. As discussed above, employees appear to be more price sensitive in making initial choices than in making changes once they are enrolled.

mitigate reclassification risk. One natural approach is to ensure that consumers have a baseline option whose price is independent of risk, and allow incremental purchases to be risk-rated. An alternative would be longer-term contracts that provide dynamic insurance (e.g. Cochrane, 1995). Understanding the dynamic aspects of health insurance are an important challenge for future empirical work.

References

Akerlof, G. "The Market for Lemons: Quality Uncertainty and the Market Mechanism, Quarterly Journal of Economics, August 1970, 84(3), pp. 488-500.

Brown, R. S., J. W. Bergeron, D. G. Clement, J. W. Hill, and S. M. Retchin, "The Medicare Risk Program for HMOs–Final Summary Report on Findings from the Evaluation." Princeton, NJ: Mathematica Policy Research, 1993.

Cardon, J.H. and I. Hendel, "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey," Rand Journal of Economics, 2001.

Carlin, C. and R. Town, "Adverse Selection, Welfare and Optimal Pricing of Employer-Sponsored Health Plans," U. Minnesota Working Paper, 2007.

Cochrane, J., "Time-Consistent Health Insurance," Journal of Political Economy, 1995, 103, pp. 445-473.

Cutler, D. M., A. Finkelstein and K. McGarry, "Preference Heterogeneity and Insurance Markets: Explaining a Puzzle of Insurance," American Economic Review, May 2008, 98(2).

Cutler, D. M. and S. Reber, "Paying for health insurance: The trade-off between competition and adverse selection." Quarterly Journal of Economics, 1998, 113(2): 433-66.

Cutler, D. M. and R. J. Zeckhauser, "The Anatomy of Health Insurance." In Handbook of Health Economics, edited by A. J. Cuyler and J. P. Newhouse, pp. 563-637. Amsterdam: Elsevier, 2000.

Enthoven, A. and R. Kronick, "A Consumer-Choice Health Plan for the 1990s: Universal Health Insurance in a System Designed to Promote Quality and Economy," New England Journal of Medicine, 1989, 320(1): 29-37.

Einav, L., A. Finkelstein and M. Cullen, "Using Price Variation to Estimate the Welfare Costs of Adverse Selection," Quarterly Journal of Economics, 2010.

Einav, L., A. Finkelstein and J. Levin, "Beyond Testing: Empirical Models of Insurance Markets," Annual Review of Economics, 2010.

Feldman, R. D. and B. E. Dowd, "Simulation of a Health Insurance Market with Adverse Selection." Operations Research, 1982, 30(6):1027-42.

Glied, S., "Managed Care ". Handbook of Health Economics. A. J. Culyer and J. P. Newhouse. New York, Elsevier Science B. V. 1A: 708-53, 2000.

Handel, B., "Adverse Selection and Switching Costs in Health Insurance Markets: When Nudging Hurts," Northestern University Working Paper, 2009.

Kaiser Family Foundation, "Employer Health Benefits: 2007 Annual Survey." Retrieved June 1, 2008, from http://www.kff.org/insurance/7672/upload/76723.pdf.

Keenan, P. S., M. J. B. Buntin, T. G. McGuire, and J. P. Newhouse, "The Prevalence of Formal Risk Adjustment in Health Plan Purchasing." Inquiry, 38, Fall 2001, pp. 245-59.

Miller, N. H., "Pricing health benefits: A cost-minimization approach." Journal of Health Economics 24, 2005, pp. 931-49.

Neipp, J. and R. Zeckhauser, "Persistence in the Choice of Health Plans." In Advances in Health Economics and Health Services Research, edited by R. Scheffler and L. Rossiter, pp. 44-74. Greenwich, CT: JAI Press, 1985.

Newhouse, J. P., Free for All? Lessons from the RAND Health Insurance Experiment, Harvard University Press, 1993.

Newhouse, J. P., "Reimbursing Health Plans and Health Providers: Efficiency in Production Versus Selection." Journal of Economic Literature, 34(3), 1996, pp. 1236-1263.

Pauly, M. V. and B. Herring, "An efficient employer strategy for dealing with adverse selection in multiple-plan offerings: an MSA example," Journal of Health Economics, 2000, 19(4): 513-528.

Rothschild, M. and J. Stiglitz, "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information," Quarterly Journal of Economics, 90, 1976, pp. 629-49.

Small, K. A. and Rosen, H. S., "Applied Welfare Economics with Discrete Choice Models." Econometrica, 41, 1981, pp. 105-130, pp. 89-116.

van de Ven, W. P. M. M. and R. P. Ellis, "Risk adjustment in Competitive Health Plan Markets," in Handbook of Health Economics, edited by A. J. Culyer and J. P. Newhouse. New York, Elsevier, 2000.

Zhao, Y., R. P. Ellis, A. S. Ash, D. Calabrese, J. Z. Ayanian, J. P. Slaugher, L. Weyuker, and B. Brown. "Measuring Population Health Risks Using Inpatient Diagnoses and Outpatient Pharmacy Data." Health Services Research, 36, 2001, pp. 180-93.

Table 1: Risk and Demographics

	Mean	Sd.	Min.	Max.
Employees (N = 3683)				
Risk Score	1.21	1.56	0.18	30.06
Age	40.56	12.01	18.00	72.00
Female	0.62	0.48	-	-
Spouse	0.28	0.45	-	-
Child	0.27	0.44	-	-
Enrollees (N = 6603)				
Risk Score	1.01	1.45	0.14	30.06
Age	32.13	17.67	0.00	72.00
Female	0.58	0.49	-	-
Spouse	0.19	0.39	-	-
Child	0.26	0.44	-	-
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.

Table 2: Plan Characteristics and Enrollment

	Network		Integrated		
	HMO	PPO	HMO	POS	All
Panel A: Plan Characteristics					
Offering Plan					
Firms	11	10	11	9	-
Firm-Years	16	14	16	13	-
Bid (Monthly)					
Employee	307	332	260	276	294
	(64)	(59)	(30)	(26)	(54)
Employee plus spouse	645	689	544	579	616
	(154)	(123)	(61)	(54)	(120)
Employee plus child(ren)	591	632	498	532	565
	(143)	(115)	(58)	(53)	(111)
Employee plus family	918	989	779	832	882
	(200)	(176)	(87)	(76)	(164)
Contribution (Monthly)					
Employee	45	73	38	58	53
	(34)	(54)	(32)	(40)	(41)
Employee plus spouse	252	303	203	255	253
	(120)	(103)	(77)	(75)	(100)
Employee plus child(ren)	221	265	177	223	222
	(97)	(86)	(62)	(55)	(81)
Employee plus family	418	495	342	415	418
	(213)	(182)	(144)	(140)	(176)
Coinsurance (%)					
Employee	87	86	97	78	87
	(6)	(5)	(7)	(2)	(9)
Deductible (Annual)					
Employee	387	440	69	336	304
	(264)	(306)	(163)	(94)	(262)
Out-of-Pocket Max (Annual)					
Employee	2818	2850	1591	2686	2468
	(462)	(474)	(625)	(731)	(775)
Panel B: Enrollment					
Employees (N=3683)					
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 (%)	22.94	7.38	58.72	10.96	100
Enrollees (N=6603)					
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 (%)	21.24	7.84	60.35	10.57	100

Notes: Panel A shows 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 ($\rho > .9$) 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. Panel B shows mean employee and enrollee characteristics. The sample is pooled across firms and years.

Table 3: Demand Model

	Non-IV (1)		IV (2)		IV and private risk (3)	
Rescaled Coefficients						
Contribution	-1.00	(0.28)	-1.00	(1.28)	-1.00	(1.43)
Coinsurance (%)	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						
X Risk Score	-1.24	(3.40)	-0.92	(2.04)	-1.59	(1.58)
X Age	1.75	(0.27)	1.27	(0.24)	1.82	(0.30)
X Female	4.93	(9.18)	4.34	(8.34)	7.20	(9.86)
X High Risk	-21.27	(12.76)	-15.14	(7.05)	-17.17	(5.59)
NPPO						
X Risk Score	-11.07	(6.76)	-8.32	(4.37)	-3.93	(2.64)
X Age	0.75	(0.45)	0.54	(0.49)	0.51	(0.48)
X Female	-34.64	(14.36)	-26.36	(7.66)	-32.44	(9.54)
X High Risk	49.38	(19.87)	36.89	(11.11)	28.11	(8.78)
IPOS						
X Risk Score	-6.10	(5.41)	-4.44	(2.90)	-5.29	(2.15)
X Age	1.58	(0.39)	1.15	(0.24)	1.56	(0.35)
X Female	-35.24	(12.85)	-25.54	(10.52)	-32.63	(10.55)
X 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)
N	3683		3683		3683	
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: Specifications (2) and (3) use predicted contributions as an instrument. Specification (3) allows for unobservable risk. Coefficients are rescaled so that the coefficient on monthly contribution is one. The dependent variable is a dummy variable for the plan chosen. IHMO is the omitted category. 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-Elasticites are the percent change in market share for a hundred dollar increase in the annual premium, calculated as $(100XMarginalEffect)/(12XMarketShare)$ in percent. The standard deviation of the logit error (σ_ϵ) is not an estimated parameter and does not have a standard error.

Table 4: 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)
X (Risk Score - 1)	288.25	(86.05)	265.36	(30.29)
X Coinsurance			0.32	(0.94)
NPPO	238.32	(22.65)	204.59	(9.82)
X (Risk Score - 1)	174.92	(34.34)	167.73	(23.81)
X Coinsurance			1.23	(1.29)
IHMO	234.86	(9.71)	228.77	(6.77)
X (Risk Score - 1)	73.67	(22.01)	104.80	(18.80)
X Coinsurance			0.38	(0.54)
IPOS	236.37	(14.13)	216.74	(13.01)
X (Risk Score - 1)	188.64	(69.35)	200.78	(61.60)
X Coinsurance			1.65	(0.82)
N	91		91	

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

Table 5: Matching and Welfare under Alternative Contribution Policies

	Matching				Welfare†			Truncated
	NHMO	NPPO	IHMO	IPOS	Gross Surplus‡	Insurer Costs‡	Social Surplus‡	Social Surplus‡
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 holds cost differentials between plans for risk scores lower than 0.75 and higher than 2.0 at these boundary levels.

† Incremental contribution, gross surplus, insurer costs, and social surplus are averaged across enrollees and denominated in $ per month.

‡ Gross 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 6: Matching and Welfare by Risk Score Quintile

	Feasible Risk Rated Contributions versus Observed							
	Matching				Welfare			Truncated
Quintile (Risk Score range)	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.)

Table 7: 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.
‡ Gross 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.

Figure 1A: Efficient allocation assigns high-risk to plan B

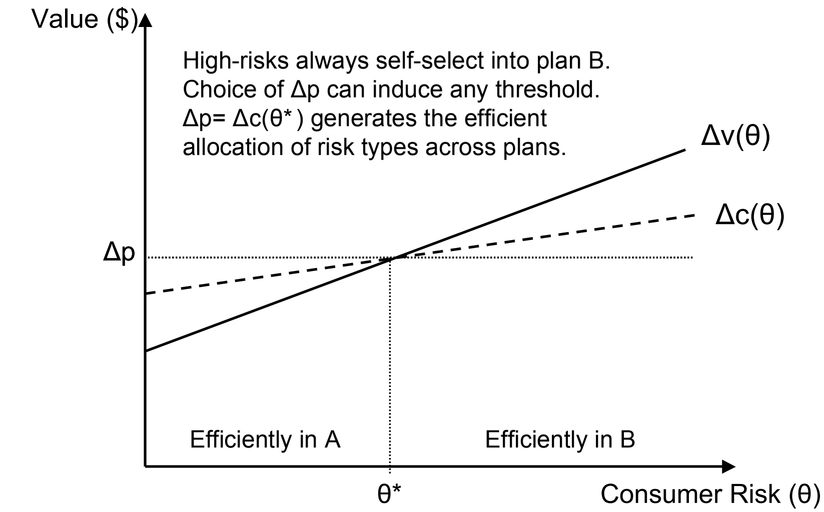


Figure 1B: Efficient allocation assigns high-risk to plan A

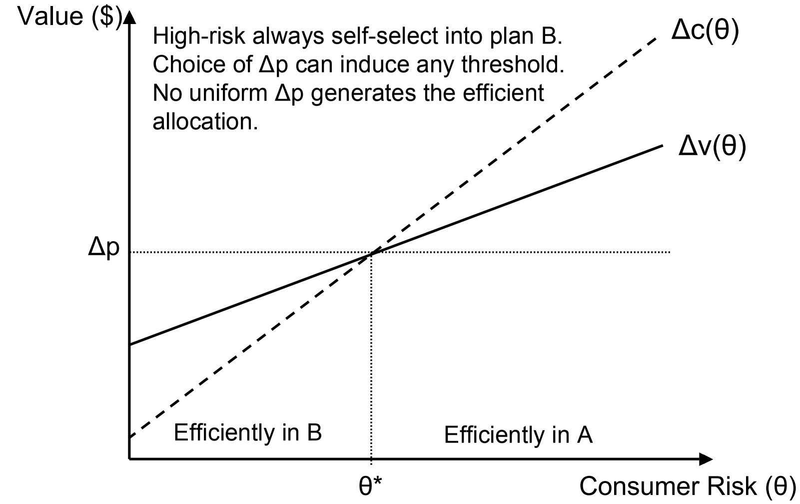


Figure 2A: Efficient allocation assigns high-risk to PPO

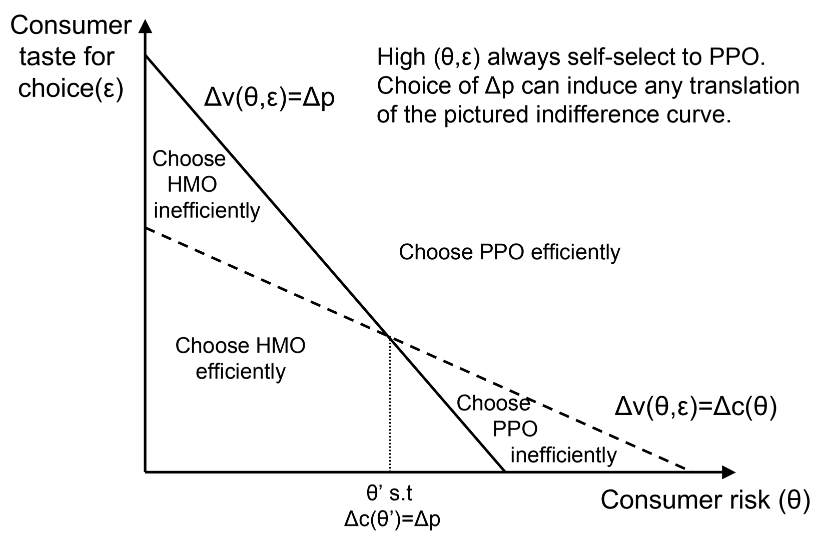


Figure 2B: Efficient allocation assigns high-risk to HMO

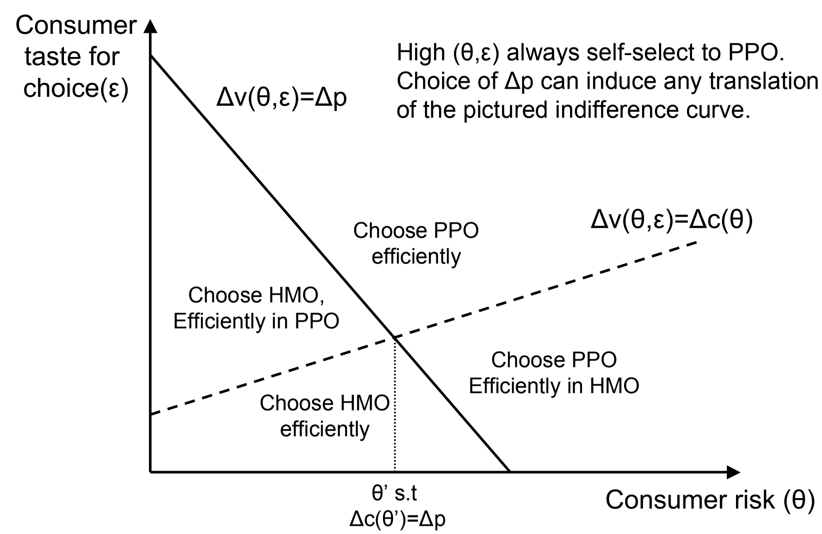
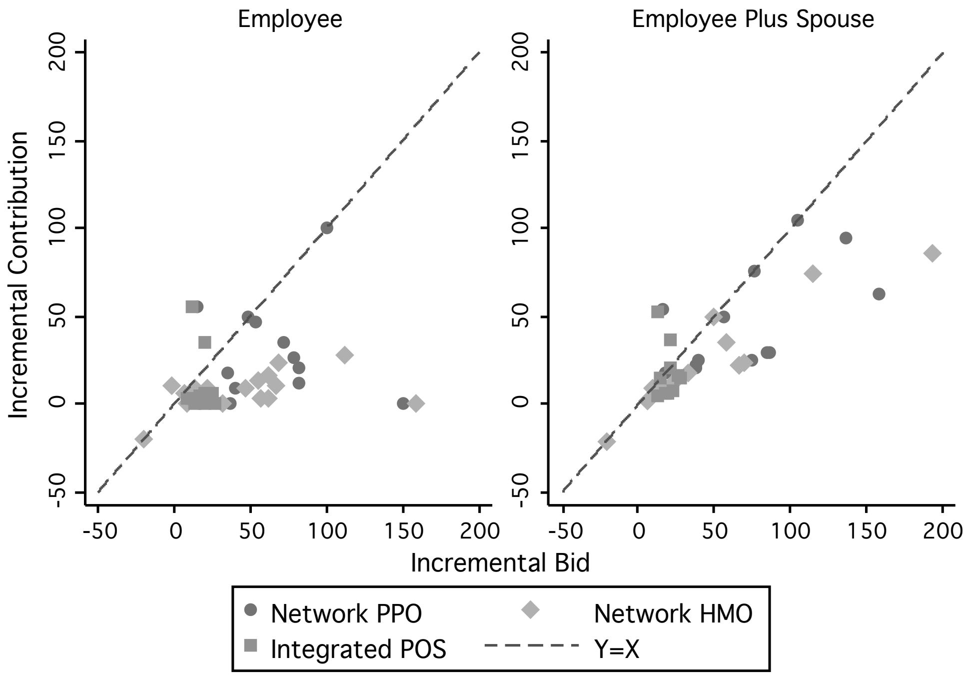
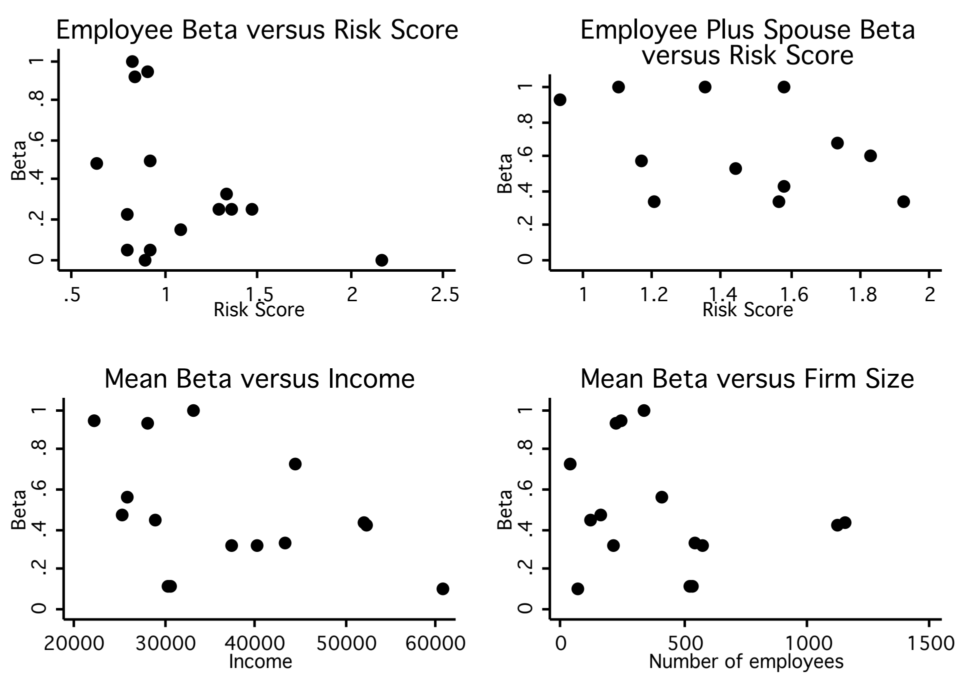


Figure 3: Contributions and Bids Relative to Integrated HMO



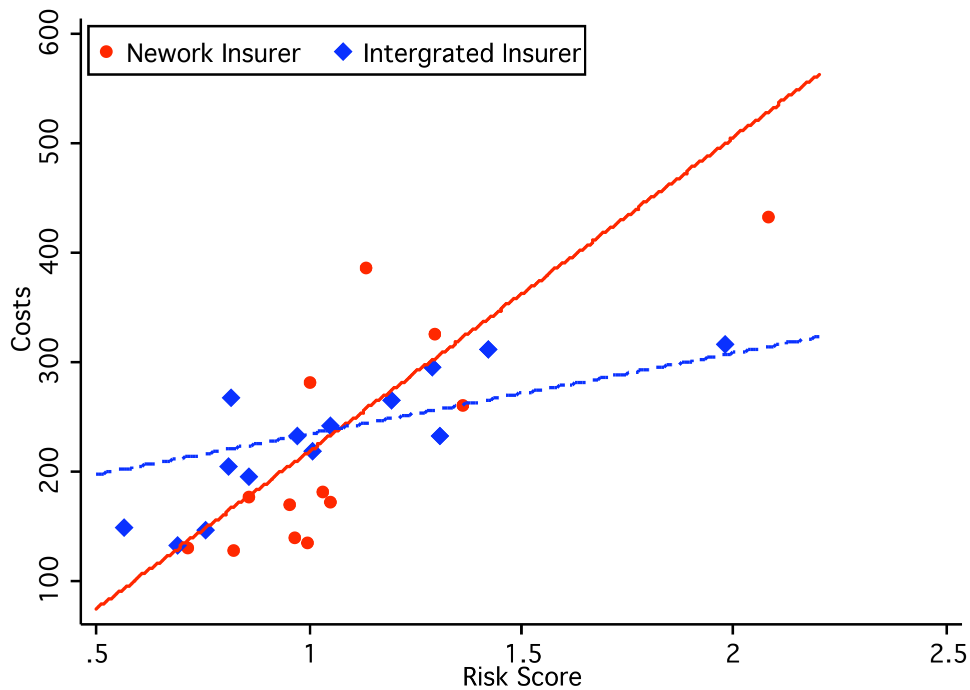
Notes: Incremental Contribution and Incremental Bid are relative to Integrated HMO. In Employee Plus Spouse, numbers are divided by two for comparability.

Figure 4: Employer Contributions and Employee Characteristics



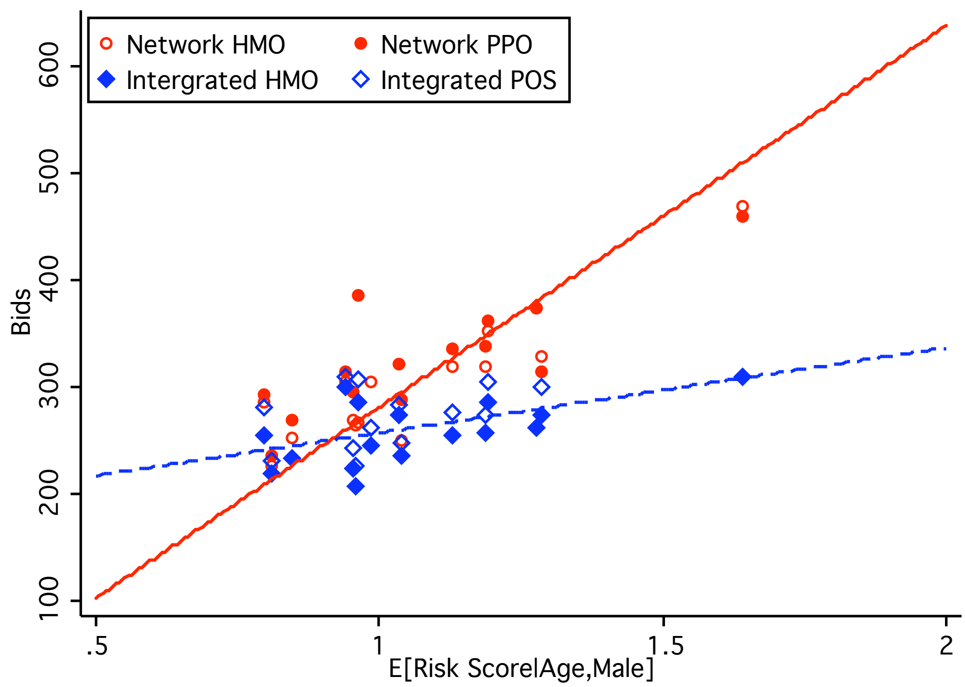
Notes: Each point represents a firm-year. Beta is the incremental pass-through of bids. See text for details. Employee Beta versus Risk Score plots the beta for employees against their mean risk score. Employee Plus Spouse Beta versus Risk Score plots the beta for those in employee plus spouse plans against their risk score. Scatter plots for other coverage tiers look similar. Mean Beta versus Income plots the mean beta (across coverage tiers) against mean employee income. Mean Beta versus Firm Size plots the mean beta (across coverage tiers) against the number of employees in the firm. Separate scatter plots by coverage tier look similar.

Figure 5: Costs by Risk Score



Notes: Each point represents a insurer-employer-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.

Figure 6: Bids by Risk Score



Notes: Each point represents a plan-employer-year. E[Risk | Age, Male] is the average predicted risk score of potential enrollees in a firm-year. Bids is the per-month bid. Fitted lines represent the Network HMO and Integrated HMO.