SIEPR Discussion Paper No. 07-47

Pricing and Welfare in Health Plan Choice

By
M. Kate Bundorf
Stanford University

Jonathan D. Levin
Stanford University

Neale Mahoney
Stanford University

June 2008

Stanford Institute for Economic Policy Research
Stanford University
Stanford, CA 94305
(650) 725-1874

Pricing and Welfare in Health Plan Choice*

M. Kate Bundorf, Jonathan Levin and Neale Mahoney †

June 2008

Abstract

Prices in government and employer-sponsored health insurance markets only partially reflect insurers' expected costs of coverage for different enrollees. This can create inefficient distortions when consumers self-select into plans. We develop a simple model to study this problem and estimate it using new data on small employers. In the markets we observe, the welfare loss compared to the feasible efficient 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 relative to single-insurer options.

*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 in different health plans. This lack of information in prices 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.

A complicating factor in health insurance markets is the role played by intermediaries. 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 (Ellis and van de Ven, 2000; Keenan et al., 2001). Consumers, however, typically face prices that do not vary by individual risk. Indeed federal law prohibits U.S. employers from charging employees or their dependents different amounts based on health-related factors (GAO 2003). And public programs frequently require community rating of enrollee contributions. Moreover, even given the institutional restrictions on pricing, contributions set by employers and in regulated markets may not be optimal in terms of maximizing social welfare given the complexities of self-selection in insurance markets.

In this paper, we analyze the effect of plan pricing on allocative efficiency. We begin by making a basic theoretical point regarding the type of prices necessary to achieve 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, set to equal the incremental cost of the marginal consumer (e.g. Feldman and Dowd, 1982; Cutler and Reber, 1998; Pauly and Herring, 2000; Cutler and Zeckhauser, 2000). We demonstrate that if either (a) differences in plan cost vary with enrollee risk more than consumer preferences, or (b) consumer choice is not purely a function of health risk, achieving efficiency is not so simple. Specifically, *no* uniform contribution policy will lead to efficient consumer choices.

While this theoretical observation in principle applies broadly, its practical relevance is a empirical question. To assess this, we develop an econometric model of health plan demand and costs. We follow a recent trend by borrowing and extending standard empirical tools used to study product markets to analyze market efficiency in a setting with selection effects. We estimate the model using a novel dataset of small employers. Our estimates indicate that, at least in the setting we consider, conditions (a) and (b) both apply: cost differences among plans vary markedly with enrollee health status, and both household preferences and health status are important for plan choice. Using the model, we go on to estimate the dollar welfare costs associated with alternative pricing policies.

We estimate that, in our setting, observed employer contribution policies cause social welfare to fall short of what could be achieved with plan contributions that vary by measurable risk. The shortfall is between $60 and $325 annually per enrollee, or 2-11% of coverage costs. Employers in our data could realize approximately 1/4 of this surplus within current nondiscrimination requirements by adjusting their employee contributions to encourage more efficient choices, but capturing the remainder would require setting different prices for people in the same firm. We also find that employees select plans based on information that is relevant for coverage cost but is not captured by the risk-adjustment system used in our setting. A hypothetical social planner who incorporated this private information into prices could increase welfare by an additional $100 annually per enrollee. Despite the social inefficiencies implied by our estimates, we also find that the observed plan offerings and resulting self-selection have generated substantial benefits over any single-plan offering.

Understanding the types of coverage in our data is important for interpreting the results. The firms we observe all offer plans from two insurers. One contracts with a fairly broad network of care providers, relying primarily on patient cost sharing and, for some enrollees, primary care gatekeepers to control utilization. The other has an exclusive provider network and a tightly integrated delivery system and requires very little patient cost sharing. Using data on the plans' premium bids and realized costs to their estimate costs, we find that the two insurers have very different cost structures. Insurer costs are similar for an individual of average health status, but higher for healthy enrollees and significantly lower for less healthy enrollees in the integrated delivery system.

While our estimates of prices elasticity are broadly in line with those from other studies of health plan choice, our findings with respect to risk selection are somewhat different. Much of the older literature has found that particular plans experience highly unfavorable selection. In our setting, we find that plans experience unfavorable selection along differing components of risk (age, gender, measured health status), but no one plan experiences very unfavorable selection overall.

One explanation for this more nuanced choice behavior may be the nature of the plans. Put simply, rather than offering "more" and "less" coverage, the plans are differentiated along several dimensions. The insurer with the broader network offers more flexibility, but requires more cost-sharing than the integrated insurer. The integrated insurer, in contrast, relies more heavily on supply side mechanisms to control utilization. This product differentiation seems particularly salient given the dramatic shift over the last two decades in the types of products available in the industry as a whole. In 1987, 73% of people with employer-sponsored health insurance had conventional coverage in which plans differed primarily in the extent of patient cost sharing. By 2007, only 3% had this type of conventional coverage, and the vast majority of consumers were enrolled in either health maintenance organizations or preferred provider organizations (KFF, 2007). Limits on provider networks have become commonplace and many, if not most, plans employ a mix of supply-side and demand-side utilization management. This evolution suggests that classic insights based on purely risk-based sorting may not adequately capture the dynamics of today's market.[1]

In our setting, two features of the market lead to distortions under uniform plan pricing: heterogeneity in household plan preferences and the integrated delivery system's significant cost advantage for individuals in worse health. Our estimates suggest that although on average 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 the 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 result in a welfare loss for those lower risk households who value greater plan flexibility. While the exact magnitude of the welfare loss depends on the extent of the cost differential and the degree of heterogeneity in taste among consumers, the basic argument may extend to other markets.

Our analysis ties in to a large literature studying health plan choice, and a smaller, more recent

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

literature quantifying the efficiency of health insurance markets. Work on health plan choice, particularly the role of risk selection, is well summarized by Glied (2000) and Cutler and Zeckhauser (2000). We extend this literature in two ways. First, we demonstrate that risk selection across plans takes place on both characteristics of consumers that are observable using existing methods to measure risk and those that are not observable ex-ante either to insurers or intermediaries.[2] This implies that current methods of risk adjusting payments across plans may be inadequate to fully counteract the incentives of plans to select enrollees through plan design. Second, as noted above, we document somewhat more nuanced risk selection, which we suggest may relate to the greater horizontal product differentiation in today's market for health insurance.[3]

More directly related is a smaller literature quantifying the efficiency of health insurance markets, starting with Cutler and Reber (1998), and including more recent papers by Carlin and Town (2007) and Einav, Finkelstein and Cullen (2008). We discuss these papers in more detail below, but one key difference is the benchmark we use to define efficiency. These papers implicitly define efficiency to be best allocation that can be observed with uniform pricing. One point we make is that this is a rather constrained notion of efficiency, as uniform pricing may preclude a large fraction of the welfare gains that could in principle be achieved with available information, and also ignores the welfare loss that is inevitably created by private information about health status.[4] At a broader level, however, we view these papers as highly complementary. They analyze quite different settings so comparing results reveals some interesting cross-market differences in plan differentiation and consumer behavior.

Our analysis may also shed light on two puzzles in the health insurance literature. One is why employers have not systematically adopted "Enthoven-style" contribution policies that expose employees to the full premium increment of choosing higher cost plans in order to promote more efficient plan choices (Enthoven and Kronick 1989). In our data, only a small fraction of the firms use such a policy. Nevertheless, our results suggest that the efficiency gains from such a change

[2]Our approach here follows Cardon and Hendel (2001), who found no evidence that private information about health status was relevant for choice behavior.

[3]With respect to the choices made by group purchasers, our paper also relates to Goldstein and Pauly's (1976) theoretical work on group health insurance as a local public good. They focused on the incentives facing an employer choosing a single plan for a group of workers with heterogeneous preferences for coverage. Our analysis looks at optimal contribution setting with multiple plans and alternative pricing constraints.

[4]Cutler and Reber's (1998) paper is a bit of an exception here because, lacking any cost data, they simply assume a fixed dollar difference in plan costs, irrespective of health status. If this were in fact the case, uniform pricing would suffice to achieve a first-best outcome, even with private health status information, which they do not consider.

would be relatively modest both because demand is relatively price inelastic and because, with a uniform contribution, any efficiency gains from moving higher risk enrollees to the integrated plan are offset by the efficiency losses experienced by lower risk enrollees who highly value the greater flexibility of the network plan. The second puzzle is why the integrated model of health care delivery has not been more successful. We find that the integrated insurer achieves the greatest savings for people in poor health, but that current pricing institutions make it difficult to target these households although it might lead to overall efficiency gains.

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. In practice, one problem with risk-rated contributions is that they can create dynamic reclassification risk for individuals. We discuss this issue in the conclusion.

2 Health Plan Pricing and Market Efficiency

We illustrate the relationship between pricing and the efficiency of plan selection by adapting the model of Feldman and Dowd (1982). In their model, consumers are distinguished by their forecastable health risk, denoted θ. Each consumer chooses between a high-cost plan (plan A) and a low cost plan (plan B). While plans can be differentiated along many dimensions, it is probably easiest to think of them, for the moment, 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 that 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, we assume that 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 $v_j(\theta) - p_j$. We make the simplifying assumption, which we maintain in our econometric model, that while consumers may be highly risk-averse to uncertainty about their health status, care and future medical expenditures, they do not exhibit diminishing marginal utility over the range of potential premiums.[5] 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.

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

$$\Delta v(\theta) - \Delta c(\theta) \geq 0.$$

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

$$\Delta v(\theta) - \Delta p \geq 0,$$

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 than plan B. Then for any incremental contribution Δp, there is a threshold type $\overline{\theta}(\Delta p)$ such that a consumer of type θ chooses plan A if and only if $\theta \geq \overline{\theta}(\Delta p)$.[6] This threshold can be varied arbitrarily with Δp. Therefore it will be possible to achieve efficient sorting *if and only* if the efficient assignment also involves a threshold rule, i.e. if the surplus function $\Delta v(\theta) - \Delta c(\theta)$ is negative up to some θ^* and positive above it. Intuitively, the requirement for efficiency is that willingness to pay increases more quickly with risk than does the cost differential between plans.

Existing analyses assume either explicitly or implicitly 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, depicted in Figure 1(a), the efficient assignment can be decentralized by setting $\Delta p = \Delta c(\theta^*)$. 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 the very highest risks. If prices are set by looking at past outcomes, one can even end up with an adverse selection "death spiral," where a higher incremental premium for plan A leads to severe adverse

[5] Though traditional models often derive risk-aversion from a globally concave utility function defined over wealth, there are many reasons to distinguish aversion to large risks, such as health status and health expenditures from diminishing marginal utility over the range of small expenditures, even apart from the modeling simplicity it affords.

[6] 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.

selection, which in turns leads to an even greater contribution gap (Cutler and Reber, 1998). Alternatively, if $\Delta p < \Delta c(\theta^*)$, too many people will select plan A, including some for whom the benefits do not exceed the incremental social costs.

There are at least two reasons to expand on this familiar analysis. First, even in the model we have been considering—where consumers are differentiated only by health status and plans are more or less vertically differentiated—it may be socially efficient for high risks to enroll in a cost-conscious plan. Arguably the benefits of delivering care efficiently are largest for the chronically ill.[7] The cost savings from a plan that more actively manages utilization might more than compensate these consumers for the loss of flexibility. In this case, depicted in Figure 1(b), the efficient assignment cannot be achieved because high risk consumers always enroll in plan A even though it is efficient for them to enroll in plan B.

Perhaps a more general issue, however, is that the differences between health plans often extend beyond "more" versus "less" coverage, and the differences between consumers often extend beyond "more" versus "less" health risk. For instance, firms increasingly offer employees both an HMO and a PPO option. An HMO may place greater restriction on provider choice and attempt to control care using supply-side mechanisms, while requiring relatively little cost sharing. A PPO typically provides access to a broader set of providers and asserts less direct control over care, but requires greater cost sharing. Consumers with health problems may place greater value on provider flexibility, but may also 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 explicitly as a PPO and plan B as an HMO. We allow consumers to vary in at least two dimensions: forecastable risk and taste. Specifically, let ε denote a consumer's preference for provider choice, so that $\Delta v(\theta, \varepsilon)$ is the consumer's extra willingness to pay for plan A. To make the extension non-trivial, suppose that tastes matter: $\Delta v(\theta, \varepsilon)$ is increasing in ε. 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 $\Delta v(\theta, \varepsilon) - \Delta p$. Clearly, uniform pricing does not generate the efficient allocation: assuming consumers cannot be priced on the basis of their tastes, achieving efficient sorting *requires* risk-based pricing so that consumers of type θ face a contribution differential $\Delta c(\theta)$.[8]

[7] 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).

[8] One way to see this in relation to the classic analysis is that for any uniform differential Δp, there are *many*

The potential matching inefficiencies under uniform pricing are depicted in Figures 2(a) and 2(b). In both, we assume Δv is strictly increasing in θ as well as ε, although as we have argued, that need not be so. The curve $\Delta v(\theta, \varepsilon) = \Delta p$ represents the set of 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, the proportion of consumers for whom the PPO is efficient increases with risk. In this case, the PPO is efficient for consumers above and to the right of $\Delta v(\theta, \varepsilon) = \Delta c(\theta)$, and the HMO is efficient for consumers below and to the left. Figure 2(b) shows the reverse situation, where the efficient proportion of consumers in the PPO declines with risk. In this case, the PPO is efficient for consumers above and to the left of $\Delta v(\theta, \varepsilon) = \Delta c(\theta)$ In both cases, there is a critical type θ' such that $\Delta c(\theta') = \Delta p$. For this risk type, consumers efficiently allocate across plans because the contribution is equal to the cost differential between plans. All consumers with risk types above θ' are effectively subsidized to choose the PPO and some do so inefficiently. All consumers with risk types below θ' face a price differential above what is actuarially fair, and some inefficiently opt for the HMO.

These figures suggest some straightforward observations. First, the relationship between efficient and equilibrium matching depends crucially on how the cost differential $\Delta c(\theta)$ varies with consumer risk. If $\Delta c(\theta)$ does not vary much, one can approximate the efficient risk-adjusted contribution with a uniform contribution. In Figures 2(a) and 2(b), reducing the extent to which the cost differential varies by risk will cause the curve defining the efficient allocation to rotate toward the curve defining the market allocation, reducing the proportions of consumers who choose the HMO and the PPO inefficiently. Second, the responsiveness of consumers to price as well as to risk when choosing among plans will determine the slope of each curve. If consumer demand is relatively price and risk elastic, the welfare gain from risk-rating contributions may be substantial. On the other hand, if consumer demand is highly inelastic, with consumers sorting primarily on the basis of taste, changing to the efficient risk-rated contribution policy may not have a large effect. Finally, the distribution of consumers based on their risk and their tastes will determine the degree of inefficiency. For example, in Figure 2(a), the welfare loss associated with uniform contributions will be greater if consumers are concentrated in the areas in which they choose plans inefficiently

marginal consumers, each with a different cost differential Δc.

rather than if they are spread out across risk/preference space.

Because each of these issues is inherently quantitative in nature, an empirical analysis is needed to assess market efficiency and social welfare. In what follows, we develop an econometric version of the model and estimate its parameters. This allows us to evaluate empirically the degree to which 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 sells a dual-carrier, choice-based health insurance product to small and mid-sized employers. The firm obtains agreements from insurers to offer their plans as a single product, markets the product to employers, and administers the benefit for those who purchase it. We refer to this firm as the intermediary.

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 offers products from two insurers. One insurer contracts non-exclusively with a relatively broad set of providers in the local market, offering two plans, which we refer to as the *network HMO* and the *network PPO*.[9] The network HMO requires enrollees to choose a primary care physician and to obtain a referral to visit a specialist, and does not cover care from out-of-network providers. The network PPO does not require referrals for specialist visits and covers care from providers outside the plan's network, although with increased cost-sharing. The second insurer features an exclusive provider network and a highly integrated delivery system that facilitates greater supply side utilization management. It also offers two plans: its standard HMO (*integrated HMO*) and a point-of-service option (*integrated POS*) that allows enrollees to seek care outside the plan's network at a higher cost.

The intermediary generally follows a standard process when dealing with employers. The employer first chooses which plans to offer to its employees. Employers may customize the basic plans described above to a limited degree by varying characteristics such as the deductible and the level of

[9] This insurer also offers a point-of-service (POS) plan that is the HMO with the option to go out-of-network at higher cost. Unfortunately we are not able to distinguish between network POS and HMO enrollees. As a result, 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.

coinsurance, but most dimensions of the coverage are fixed. Employers typically offer four coverage tiers: employee only, employee plus spouse, employee plus children, and employee plus family.[10] The level of cost sharing varies across coverage tiers.

The insurers then provide quotes, which we refer to as bids, for each plan on the employer's menu. The intermediary provides information on the composition of the group to help insurers form their bids. In an employer's first year with the intermediary, this information is limited to 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 each 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.

Given the bids, the employer sets the employee contribution by coverage tier for each plan on the menu. While neither the intermediary nor the insurers place any restrictions on how employers set their contributions, the intermediary encourages them to use a "managed competition" approach in which employees face the full marginal cost for more expensive plans. Employees make their choices after observing the menu of plans and the required contributions. If an employee selects a plan, the plan must allow the employee to enroll.

The last step is a series of payments. For each employee enrolled in a particular product, the employer pays the intermediary the insurer's bid. The intermediary passes on these payments to the insurers, implementing a system of transfers between insurers to compensate for differential selection across plans based on the health status of enrolled employees and their dependents.

The intermediary uses a standard methodology for measuring enrollee health status, the Rx-Group model developed by DxCG, Inc. The model produces risk scores conditional on a person's age, sex, and health status. Health status is determined by using prescription drug utilization to identify chronic conditions.[11] In our setting, the insurers report prescription drug utilization from the current year to the intermediary. The intermediary uses the DxCG algorithm to predict ex-

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

[11] DxCG uses an internally-developed mapping of prescription drugs to their therapeutic indication to identify chronic conditions. 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.

penditures for each enrollee and makes corresponding transfers across the insurers. In our analysis, we use the term "risk score" to refer to the DxCG prediction, conditional on age, sex and health status, 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.

Each insurer also provides the intermediary with information on their 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.

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.[12] 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

[12] 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.

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. This variation plays a key role in our analysis. We use information on insurer bids and realized costs to estimate models of the relationship between costs and risk. Because insurers report both bids and costs at the employer level, variation across employers in average risk is necessary to identify these relationships.

Table 2 provides information on the plans offered by the employers in our sample. Most employers offer all four plans, and all offer both HMOs and at least one other plan. On average, the integrated HMO is the least expensive plan and has the lowest enrollee contribution. This plan features high rates of coinsurance, a low deductible, and a low out-of-pocket maximum. The network PPO, which is offered by all but one employer, is on average the most expensive plan and has the highest employee contribution. This plan 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 based on family structure, as indicated earlier, 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 as well as within tiers. In general, employee contributions represent a fraction of the bid and the fraction is smallest for employee only coverage.

In our demand model, we identify price elasticities based on variation across both firm-years and enrollment tiers in the relative contributions for different plans. Figure 3 demonstrates the extent and sources of variation in relative plan contributions by plotting the incremental contribution against the incremental bid 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.

There is significant heterogeneity in contributions across employers, and across tiers. 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. Figure 3 also demonstrates the significant variation across employers in the bids they receive for similar plans. As we demonstrate later, this variation is driven in large part by differences across firms in the demographic composition of employees.

We summarize enrollment patterns in Table 3. 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, and 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. 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:

$$u_{hj} = \phi_j \alpha_\phi + x_h \alpha_{xj} + \psi(r_h + \mu_h; \alpha_{rj}) - p_j + \sigma_\varepsilon \varepsilon_{hj}. \quad (1)$$

In this representation, household utility depends on observable plan characteristics ϕ_j, the monthly plan contribution p_j,[13] 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.[14] 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 risk score distribution. The

[13] 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 10 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 $p_h = (1 - \tau_h)\rho_h$.

[14] 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.

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.[15]

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.[16] 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,

$$q_{hj} = 1 \iff u_{hj} \geq u_{hk} \text{ for all } k \in \mathcal{J}_h. \tag{2}$$

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:

$$\Pr\left(q_{hj} = 1 \mid X_{hj}, \mu_h\right) = \frac{\exp\left(v_{hj}\right)}{\sum_{k \in J_h} \exp\left(v_{hk}\right)}. \tag{3}$$

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 cost of enrolling individual i as

$$c_{ij} = a_j + b_j \cdot (r_i + \mu_i - 1) + \eta_{ij}. \tag{4}$$

[15] 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.

[16] 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.

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 forecastable 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:

$$C_{kf} = \sum_{j \in \mathcal{J}_{kf}} \sum_{i \in \mathcal{I}_{jf}} \left\{ a_j + b_j \cdot (r_i + \mu_i - 1) + \eta_{ij} \right\}. \tag{5}$$

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:

$$\frac{1}{|\mathcal{I}_f|}\sum_{i\in\mathcal{I}_f}\mathbb{E}[c_{ij}|x_i] = a_j + b_j(\mathbb{E}[\overline{r}_f|x_f]-1), \tag{6}$$

where $\overline{r}_f$ denotes the average risk of employees in firm f, which the insurer forecasts using the available demographic information, x_f.[17]

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

$$B_{jf} = \delta_j \cdot (a_j + b_j \cdot (\mathbb{E}[\overline{r}_f|x_f]-1)) + \nu_{jf}, \tag{7}$$

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:

$$p_{jlf} = \beta_{lf}\cdot \underline{B}_{lf} + \gamma_{lf}\cdot (B_{jlf} - \underline{B}_{lf}) + \xi_{jlf}. \tag{8}$$

This model describes employer behavior in our data remarkably well. The residuals from the linear regression (8) 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 $\beta < \gamma$.

[17] 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.[18]

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}).[19] There is substantial variation in each of the first three variables. For instance, the difference in the bids for the integrated HMO and the network PPO ranges from \$50 to \$150 per month (Figure 3), with a large fraction due to cross-firm variation in demographic risk. Similarly, some employers in our data pass through the full incremental difference in plan bids, while others pass through only a fraction or in some cases none at all (Figure 3).

[18] 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.

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

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 (8). We take this as our preferred specification in performing welfare analysis although the results are similar to the baseline case with no instruments.

A second concern is that plan bids are correlated with unobserved household tastes. This could happen if an insurer believed its plan was attractive due to, say, a nearby clinic location. It would generate a correlation between the idiosyncratic bid component, ν_{jf}, and household preferences ε_{hj}. We view this problem as most likely of marginal importance given the limited information on the part of insurers. Nevertheless, we check our estimates by instrumenting for plan contribution with a predicted value that is constructed by plugging the predicted bid $\hat{B}_{jf}$ from (7) into the contribution model (8). 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 (8) 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.[20]

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.

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

$$\mathbb{E}_{\varepsilon}\left[q_{hj} - \Pr\left(q_{hj} = 1 \mid X_h\right) \mid Z_h, \mu_h\right] = 0. \tag{9}$$

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 (3) 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:

$$\mathbb{E}_{\eta}\left[C_{kf} - \sum_{j \in \mathcal{J}_{kf}} \sum_{i \in \mathcal{I}_{jf}} \left\{a_j + b_j \cdot (r_i + \mu_i - 1)\right\} \mid X_{kf}, \mu_{kf}\right] = 0. \tag{10}$$

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

[20] 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.

of participation, we have:

$$\mathbb{E}_{\nu}\left[B_{jf}-\delta_j\cdot\left(a_j+b_j\cdot\left(\mathbb{E}[\overline{r}_f|x_f]-1\right)\right)\mid X_f\right]=0. \tag{11}$$

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.[21] 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 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 $\hat{\theta}=\text{argmin}_{\theta\in\Theta}\hat{m}(\theta)'W\hat{m}(\theta)$. 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 how these effects are calculated. 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. Following Small and Rosen (1981), the expected change in the money-metric utility of household h following a price change from p to

[21] 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.

p' is:

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

A subtlety here is that our model of household choice invokes a representative household member, so to give each individual equal welfare weight, we scale each household by the number of members n_h. Again, this formula is conditioned on μ, so define:

$$\Delta U_h\left(p, p'\right) = \int \Delta U_h\left(p, p', \mu_h\right) dF_\mu\left(\mu_h\right).$$

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:

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

where c^e_{hj} is the expected cost of covering household h in plan j. This cost, and the choice probabilities are again conditional on household unobserved risk, μ_h, but we can integrate over the distribution of μ_h to obtain $\Delta \Pi_h\left(p, p'\right)$.

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 $\Delta U_h\left(p, p', \mu_h\right)$ and $\Delta \Pi_h\left(p, p', \mu_h\right)$ for each simulation draw, and average over the draws to obtain $\Delta U_h\left(p, p'\right)$ and $\Delta \Pi_h\left(p, p'\right)$. 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 4 presents parameter estimates from three different specifications of the demand model.[22] 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 (8). The third column, which is our preferred specification, allows for private information about risk. To scale the utility to money-metric form, we divide each coefficient by the coefficient on the monthly contribution and adjust the standard errors accordingly. We report the price effects as semi-elasticities at the bottom of the table.

Effect of Demographics and Risk on Choice

The demand estimates 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). Women, who at the age of workers 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). 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

[22] 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.

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.

Our results also suggest that private information about health risk plays a role in plan choice, although the estimate is not very 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 (e.g. 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.

Effect of Plan Prices on Choice

In the bottom panel, we present price semi-elasticities of demand, defined as the percentage decrease in market share resulting from a $100 increase in the annual enrollee contribution, evaluated at the mean choice probability for each plan. On average, a $100 dollar increase in the annual enrollee contribution decreases market share by 7 to 9 percent. While instrumenting the contribution reduces the precision of the estimate of the price effect, it has relatively little effect on its magnitude. These estimates suggest that demand is relatively inelastic; a finding that accords with most prior studies of plan choice. Existing studies estimate that a $100 (inflation adjusted) increase in the annual contribution reduces market share by 1.6 to 9.6 percent, placing our estimate in the middle to high end of those of existing studies.[23]

[23] Studies in setting similar to ours vary in the method for reporting elasticities. Following Chernew et. al. (2007), we reconcile the estimates of price elasticity across studies by converting them to semi-elastiticities (calculated as the percent change in market share in response to an increase of $100 in the employee contribution). Because the studies cover different time periods, we adjust the change in the employee contribution for inflation. After making these adjustments, most studies estimate semi- elastiticities in the range of -1.5 to -3.5 (Chernew et. al 2007; Chernew et. al 2003; Strombom, et. al 2002; Wedig and Tai-Seale 2002; Cutler and Reber 1998). Estimates from Royalty and

The results in Table 4 also include the estimated value of plan characteristics other than price, such as coinsurance rate and deductible. Enrollees appear to be moderately sensitive to 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.[24]

Demand Model Sensitivity Analyses

Because the estimates of risk and price elasticity are the key parameters for our welfare calculations, we have examined the sensitivity of these estimates with respect to a variety of issues. We briefly summarize these analyses here, providing some more detailed results in Appendix Table A1. Our main specifications pool data from enrollees in different coverage tiers. This increases the sample size and generates additional variation in plan contributions due to the higher pass-through rates for dependent coverage tiers, but requires us to make assumptions about household aggregation. We test whether our estimates are sensitive to these assumptions by restricting the sample to employees purchasing employee-only coverage (Table A1 - Column 2) and also by controlling for family structure in our full-sample specification (Table A1 - Column 3). Both variations give similar results to our preferred specification (Table A1 - Column 1).[25]

As discussed in Section 4.2, we developed different instruments to exploit alternative sources of identifying variation in employee contributions and hence test the robustness of our estimates to different assumptions regarding potential sources of endogeneity. In Table 4 (Column 3), we demonstrate that the estimate of the price effect is robust to idiosyncratic variation in employer contribution setting. The results in Table A1 demonstrate that the estimate of the price effect does not appear to be affected by endogeneity in either insurer bidding (Column 5) or employer choice of pass through rates (Column 4).

Solomon 1999 are somewhat higher (ranging from -2.4 to -21.1). Similarly, Jin and Sorensen (2005), who analyze retirees, have an estimate that implies a semi-elastiticity of -9.0. At the extremes, estimates from Feldman, Dowd and Cassou (1989) imply semi-elastiticies of -19.0 to -63.0, and Barringer and Mitchell (1984) find semi-elastiticies of less than -0.15.

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

[25] While the estimate of the effect of being a high risk individual on demand for the network PPO is smaller in magnitude in the employee-only than in the full sample, the estimate is imprecise likely because it is identified by a small number of enrollees (Column 1). When we include controls for family structure in the model estimated on the full sample, we find that families with a spouse seem to prefer the PPO relative to the average household (Column 2).

Finally, we note that while we present results from relatively parsimonious model, our estimates do not appear to be sensitive to including a variety of different control variables. Additional controls we have explored include plan out-of-pocket maximums, whether a plan was offered to an employer group prior to the employer hiring the intermediary, and whether an employee is making his or her initial plan choice or a later plan choice.

Structure of Plan Costs

The estimate from the bid and cost models in Table 5 indicate that the integrated and network insurers have quite different cost structures. Expected costs do not differ much across plans for an enrollee of average risk. The expected monthly cost for an enrollee with a risk score of one is \$235 for the integrated HMO, \$236 for the integrated POS, \$218 for the network HMO, and \$238 for the network PPO. The integrated insurer, however, has a substantial cost advantage for higher risk enrollees. The expected monthly cost of an enrollee with a risk score of two is \$309 for the integrated HMO, \$425 for the integrated POS, \$507 for the network HMO and \$413 for the network PPO. The mark-up varies across insurers. We estimate that the network insurer bids 24% over expected costs and the integrated insurer bids 8% over costs.

The difference in cost structures can be seen even 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 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. As the Figures illustrate, the integrated plans appear to have a cost advantage in covering higher risk enrollees. In contrast, the network plans do relatively well for low risks. The clear patterns in the figures also indicate why we obtain fairly precise cost estimates despite a small number of observations.

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 a story where cost sharing limits visits, particularly for low risks, but has 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 visit for the high risks.

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 prices 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.

A 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 6 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 7 decomposes the welfare calculation by household risk quintiles. The lowest and highest risk quintiles (average household risk below 0.36 and above 1.33) generate about three-quarters of the welfare effect. This raises a concern that our calculation might be driven in part by extrapolating plan costs out of sample. As Figures 5 and 6 illustrate, we observe plan bids and costs only for average risk scores between 0.75 and 2.0. In contrast, household risk ranges from 0.16 to 30.1. To address this, we truncate the cost differentials between plans at their 0.75 and 2.0 levels and re-calculate the welfare numbers. These calculations appear in the final columns of Tables 6 and 7. 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, a social planner could achieve approximately 70% of the potential welfare gains associated with individualized pricing using only observable information on risk.

Social Welfare without Risk-Rating

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

The next scenario in Table 6 addresses the question of what is possible without individualized pricing by considering contributions that maximize social welfare subject to being uniform within each firm-tier. As in the case of fully risk-rated prices, optimizing uniform within firm contributions leads to a reallocation of high-risk households into the integrated plans and away from the network plans, particularly the PPO. The shift is much less dramatic, however, than under full risk rating. Overall social surplus is $1.40-6.70 higher per enrollee-month than under the observed policies, but still $3.60-20.40 below the efficient level. This indicates that about 3/4 of the observed inefficiency is due to the requirement of nondiscriminatory pricing within firms. Nevertheless, it appears that employers could increase social surplus by around 1-3% 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." About 1/2 of the firms in our sample use this approach. The last entry in Table 6 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, we compare 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 results indicate that employers' observed contribution policies lead to a notable inefficiency in plan enrollment. Our baseline estimate is that the welfare loss is on the order of 2-11% of the costs of coverage. This loss arises because households do not face incremental plan contributions that accurately reflect incremental costs of coverage. Because our results are based on a small set of employers in a particular geographic area, it is not clear whether our findings are broadly generalizable. Thus, it is useful to compare our results to those from the relatively small literature quantifying the efficiency of health insurance markets (Cutler and Reber 1998, Carlin and Town 2007, and Einav, Finkelstein and Cullen 2008).

As noted above, one difference between our study and these other recent papers is the choice of the efficient benchmark. In particular, the papers just mentioned consider only pricing schemes in which contributions are uniform across individuals. Unless preferences are perfectly correlated with health risk, or plan cost differentials are precisely the same in dollar terms for each consumer, this is a constrained notion of efficiency. In our context, the analogous exercise is to compare welfare under observed pricing to optimal uniform pricing. We found that this implies a welfare loss of approximately 1-3% of coverage costs, or about 25% of the welfare difference between the observed outcome and the feasible efficient outcome. It would be interesting to know if this conclusion is specific to our environment or more general.

What we can compare across studies is the estimated welfare loss from non-optimal uniform pricing, and there appears to be a reasonable amount of consistency in this regard. Cutler and Reber calculate that a change in Harvard University's contribution policy generated a welfare loss on the order of 2-4% of coverage costs.[26] Using data on plan demand and realized costs from a different and even larger employer, Einav, Finkelstein and Cullen calculate that moving from zero-profit plan contributions to the most efficient uniform contribution could increase welfare by up to 2%. Finally, Carlin and Town (2007) study a third large employer and find the plan pricing has little effect on aggregate welfare, primarily because demand is quite price inelastic in the setting they examine.

Underlying the similarities in these estimates, however, are very different market environments. For example, in the episode studied by Cutler and Reber (1998), Harvard modified its contribution policy to link the employee contribution for the PPO relative to HMO plans to the observed premium differential between the plans. While they estimate a price elasticity very similar to ours — a $100 increase in the annual employee contribution translates into about a 3% decrease in plan market share — they find that the PPO experienced unfavorable selection prior to the introduction of the policy and the degree of unfavorable selection into the PPO increased after the relative price increase. Differences in the plan offerings may explain their finding of substantial risk-selection. In the Harvard market, the PPO and the HMO may have been primarily vertically differentiated, with PPO enrollees getting "more" coverage. As we have discussed, plans in our setting seem to be more horizontally differentiated. A related point is that Cutler and Reber do not have a health status measure, so they measure risk selection based only on age. Our estimates suggest that differing components of risk may affect selection in different ways, particularly when products are horizontally differentiated.

Carlin and Town's (2007) paper is similar to ours in attempting to estimate both a demand system and cost structures for different plans, although the setting is quite different. Like us, they find that plans have quite different cost structures, so there are substantial cost savings to be realized by efficiently allocating enrollees to plans. They also uncover evidence of risk-based self-selection. At the same time, however, they find that demand is quite inelastic. Roughly, a $100 increase in the annual employee contribution translates into only about a 1% decrease, or less, in market share. This leads them to conclude that prices don't have much affect on the plan choice and hence on aggregate welfare. One explanation for their estimate of highly inelastic demand is that they rely on time-series variation in contributions. The literature on health plan choice suggests

[26] Cutler and Reber lacked data on plan costs, so they assumed that the PPO's cost of covering individuals was roughly $200 per year more than the HMOs, independent of risk. This difference was more or less the incremental contribution for the PPO under Harvard's original policy, so they designate that policy to be the efficient benchmark.

that employees may be more price sensitive in making initial choices than in making changes once they are enrolled.

Einav, Finkelstein and Cullen (2008) also use data from a single large employer to estimate both plan demand and plan costs. In their case, the plans are PPOs with different levels of cost-sharing. As in earlier studies of vertically differentiated plans, they find substantial adverse selection for the more generous plan, once again pointing to the potential importance of sources of product differentiation. Methodologically, their paper is also a bit different in that they simply specify demand and cost as a function of price, rather than using a characteristics-based model as we do here.

Taken together, these studies reinforce our earlier observation that the welfare consequences of inefficient pricing are driven by two factors: first, the cost of inefficient sorting, which is a function of employee preferences and plan costs; and second, the price elasticity of demand, which determines the sensitivity of self-selection to price. While we have gone to some length to document the robustness of our estimates along both dimensions, our sample is nonetheless small and our model imposes a variety of assumptions. Therefore as a further sensitivity check, we also recalculated the welfare difference between the observed and the efficient allocations 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 elasticity of demand. These sensitivity analyses increase the range of the welfare gains to 1-13% of total coverage costs. Given the variety of estimates of risk and price sensitivity in the literature, one may want to assign a corresponding range of uncertainty to the potential welfare costs of distorted prices.

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. In the absence of such policies, employee contributions that more accurately reflected cost differentials between plans might still increase welfare by 1-3% of coverage costs. A key point to emphasize is that despite the observed pricing distortions, there appear to be substantial gains in our context from introducing plan choice at the employer level.

Our findings raise the question of how one might fully evaluate the costs and benefits of risk-rated pricing. Obviously many insurance markets (e.g. the private market for life insurance) have the feature that prices reflect enrollee risk. A difficulty with risk rating in health insurance is that coverage is typically purchased on an annual basis. In such an environment, risk-rated premiums expose households to reclassification risk as their health status changes over time. Indeed the dynamic insurance provided by uniform pricing within firms may be a significant benefit of an employer-based system.

There are also ways to promote static efficiency while mitigating reclassification risk. Efficiency requires risk-rating only the incremental contributions for higher-cost plans. So one way to insure against changes in health status is to have a baseline option where the price is independent of risk. An alternative would be longer-term contracts that provide dynamic insurance (e.g. Cochrane, 1995). It would be interesting to extend our approach into a more dynamic setting and try to pinpoint the value of dynamic insurance that is provided by uniform contributions and balance this value against the static distortions.

References

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

Barringer, M. and O. S. Mitchell, "Workers' Preferences Among Company-Provided Health Insurance Plans." *Industrial and Labor Relations Review* 48(1):141-52, 1994.

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.

Chernew, M., G. Gowrisankaran, and D. P. Scanlon, "Learning and the Value of Information: Evidence from Health Plan Report Cards," 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." Work in progress, 2008.

Feldman, R., M. Finch, B. Dowd, and S. Cassou, "The Demand for Employment-based Health Insurance Plans." *Journal of Human Resources*, 1989, 24(1):117-42.

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.

Goldstein, G. S. and M. V. Pauly, "Group Health Insurance as a Local Public Good," in .*The Role of Health Insurance in the Health Services Sector,* ed. R. N. Rosett. New York: National Bureau of Economic Research, 1976, pp. 3-110.

GAO, "Private Health Insurance: Federal and State Requirements Affecting Coverage Offered by Small Businesses". pp. 1-63. United States General Accounting Office, 2003.

Jin, G. Z. and A. Sorensen, "Information and consumer choice: The value of publicized health plan ratings." *Journal of Health Economics,* 2005, 25(2):248-75.

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.

Royalty, A. B. and N. Solomon, "Health Plan Choice: Price Elasticities in a Managed Competition Setting," *Journal of Human Resources* 34(1), 1999, pp. 1-41.

Scanlon, D. P., M. Chernew, C. McLaughlin, and G. Solon, "The Impact of Health Plan Report Cards on Managed Care Enrollment," *Journal of Health Economics,* 21, 2002, pp. 19-41.

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

Strombom, B. A., T. C. Buchmueller, and P. J. Feldstein, "Switching costs, price sensitivity and health plan choice." *Journal of Health Economics*, 21, 2002, 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.

Wedig, G. J. and M. Tai-Seale, "The effect of report cards on consumer choice in the health insurance market." *Journal of Health Economics*, 21, 2002, pp. 1031-48.

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

	Network		Integrated		
	HMO	PPO	HMO	POS	All
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)

Notes: Mean plan characteristics, with standard deviations in parentheses. Plan characteristics are pooled across years. Coinsurance, deductible, and out-of-pocket maximum are in-network values and are highly correlated ($\rho > .9$) with out-of-network coinsurance, deductible and out-of-pocket maximum. 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.

Table 3: Risk and Demographics by Plan

	Network		Integrated		
	HMO	PPO	HMO	POS	All
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: Employees and enrollees are pooled across firms and years.

Table 4: Demand Model

	Non-IV (1)	IV (2)	IV and μ (3)
Rescaled Coefficients			
Contribution	-1.00	-1.00	-1.00
	(.28)	(1.28)	(1.43)
Coinsurance (%)	1.91	1.41	2.31
	(.49)	(2.00)	(1.28)
Deductible	-0.01	-0.01	0.01
	(.02)	(.02)	(.03)
NHMO			
X Risk Score	-1.24	-0.92	-1.59
	(3.40)	(2.04)	(1.58)
X Age	1.75	1.27	1.82
	(.27)	(.24)	(.30)
X Female	4.93	4.34	7.20
	(9.18)	(8.34)	(9.86)
X High Risk	-21.27	-15.14	-17.17
	(12.76)	(7.05)	(5.59)
NPPO			
X Risk Score	-11.07	-8.32	-3.93
	(6.76)	(4.37)	(2.64)
X Age	0.75	0.54	0.51
	(.45)	(.49)	(.48)
X Female	-34.64	-26.36	-32.44
	(14.36)	(7.66)	(9.54)
X High Risk	49.38	36.89	28.11
	(19.87)	(11.11)	(8.78)
IPOS			
X Risk Score	-6.10	-4.44	-5.29
	(5.41)	(2.9)	(2.15)
X Age	1.58	1.15	1.56
	(.39)	(.24)	(.35)
X Female	-35.24	-25.54	-32.63
	(12.85)	(10.52)	(10.55)
X High Risk	16.40	12.23	9.43
	(17.28)	(8.81)	(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 privately known risk. Coefficients are rescaled so that the coefficient on 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 $(100X\text{MarginalEffect})/(12X\text{MarketShare})$ in percent.

Table 5: Costs and Bids

	(1)	(2)
Network Insurer Markup	1.29	1.27
	(.12)	(.07)
Integrated Insurer Markup	1.08	1.07
	(.04)	(.03)
NHMO	218.42	195.08
	(21.35)	(9.48)
X (Risk Score - 1)	288.25	265.36
	(86.05)	(30.29)
X Coinsurance		0.32
		(.94)
NPPO	238.32	204.59
	(22.65)	(9.82)
X (Risk Score - 1)	174.92	167.73
	(34.34)	(23.81)
X Coinsurance		1.23
		(1.29)
IHMO	234.86	228.77
	(9.71)	(6.77)
X (Risk Score - 1)	73.67	104.80
	(22.01)	(18.80)
X Coinsurance		0.38
		(.54)
IPOS	236.37	216.74
	(14.13)	(13.01)
X (Risk Score - 1)	188.64	200.78
	(69.35)	(61.6)
X Coinsurance		1.65
		(.82)
N	91	91

Notes: Estimates of cost parameters from joint demand and cost model. See text for details.

Table 6: 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. Risk Score conditional on plan choice. Incremental Contribution with respect to the contribution for IHMO and unconditional on plan choice. Truncated fixes cost differentials between plans for risk scores outside of 0.75 and 2.0. Monthly average incremental condition, gross surplus, costs, and social surplus are shown.

† 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 7: 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 8: The Value of Plan Choice

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

Notes: † Gross surplus, insurer costs, and social surplus are averaged across enrollees and denominated in $ per month.
‡ 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

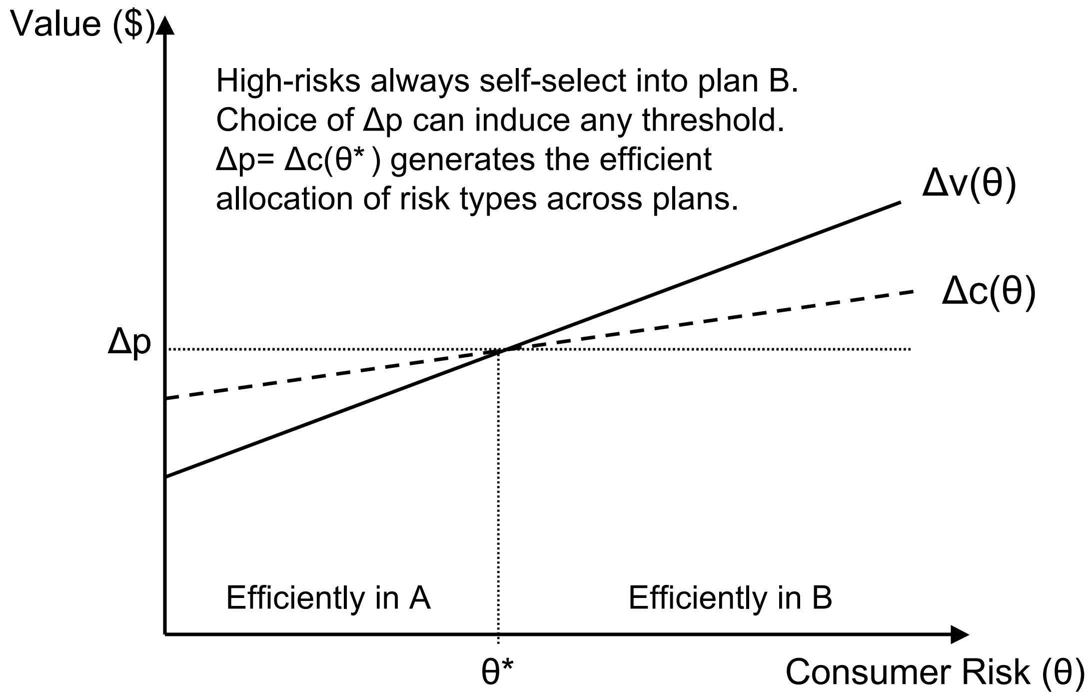


Figure 1B: Efficient allocation assigns high-risk to plan A

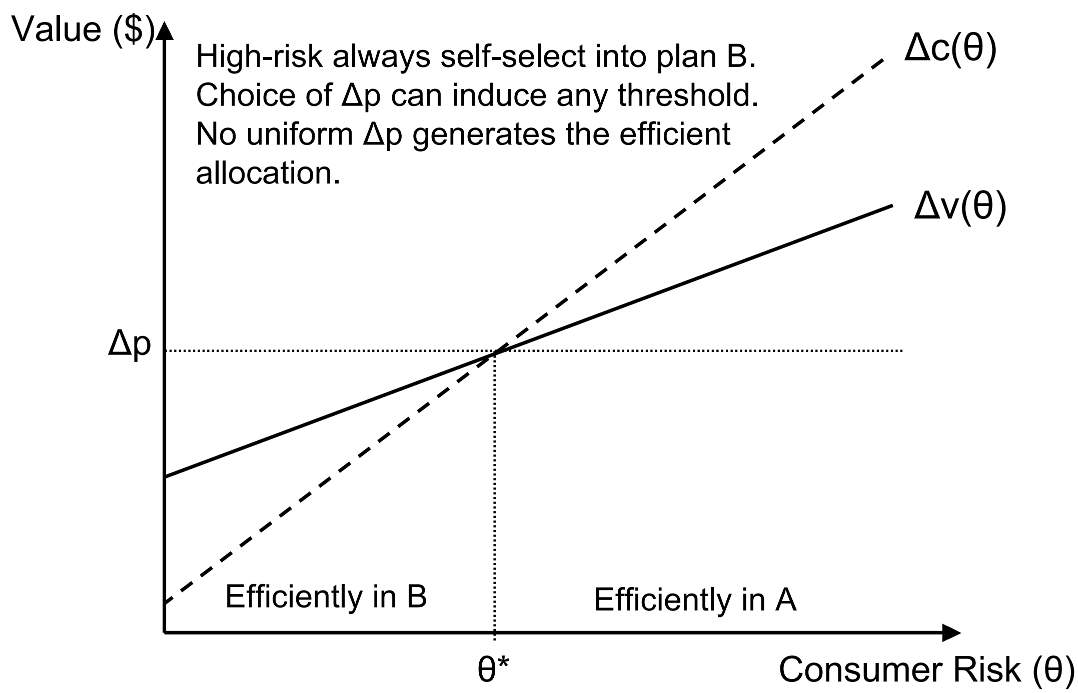


Figure 2A: Efficient allocation assigns high-risk to PPO

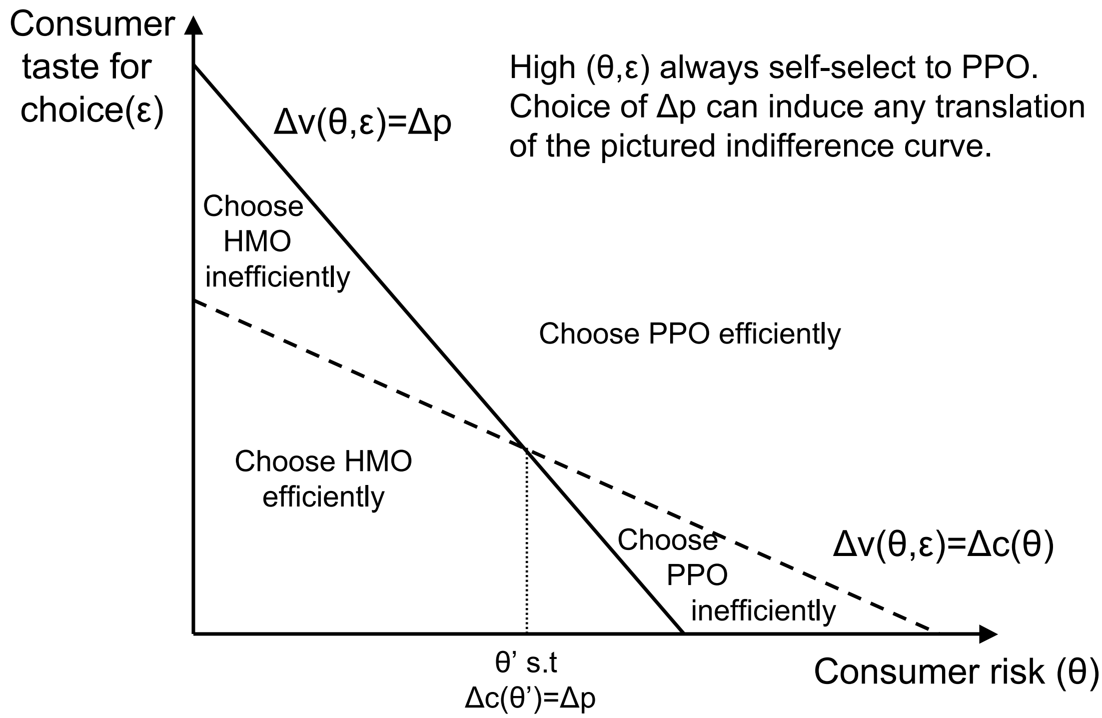


Figure 2B: Efficient allocation assigns high-risk to HMO

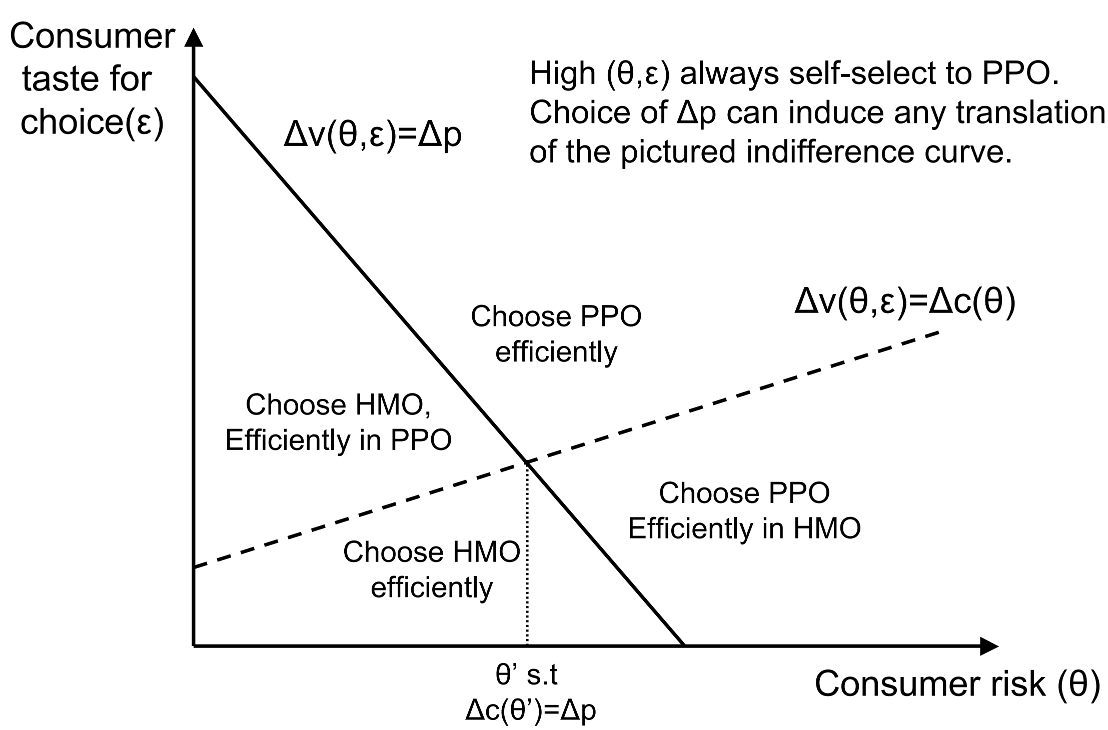
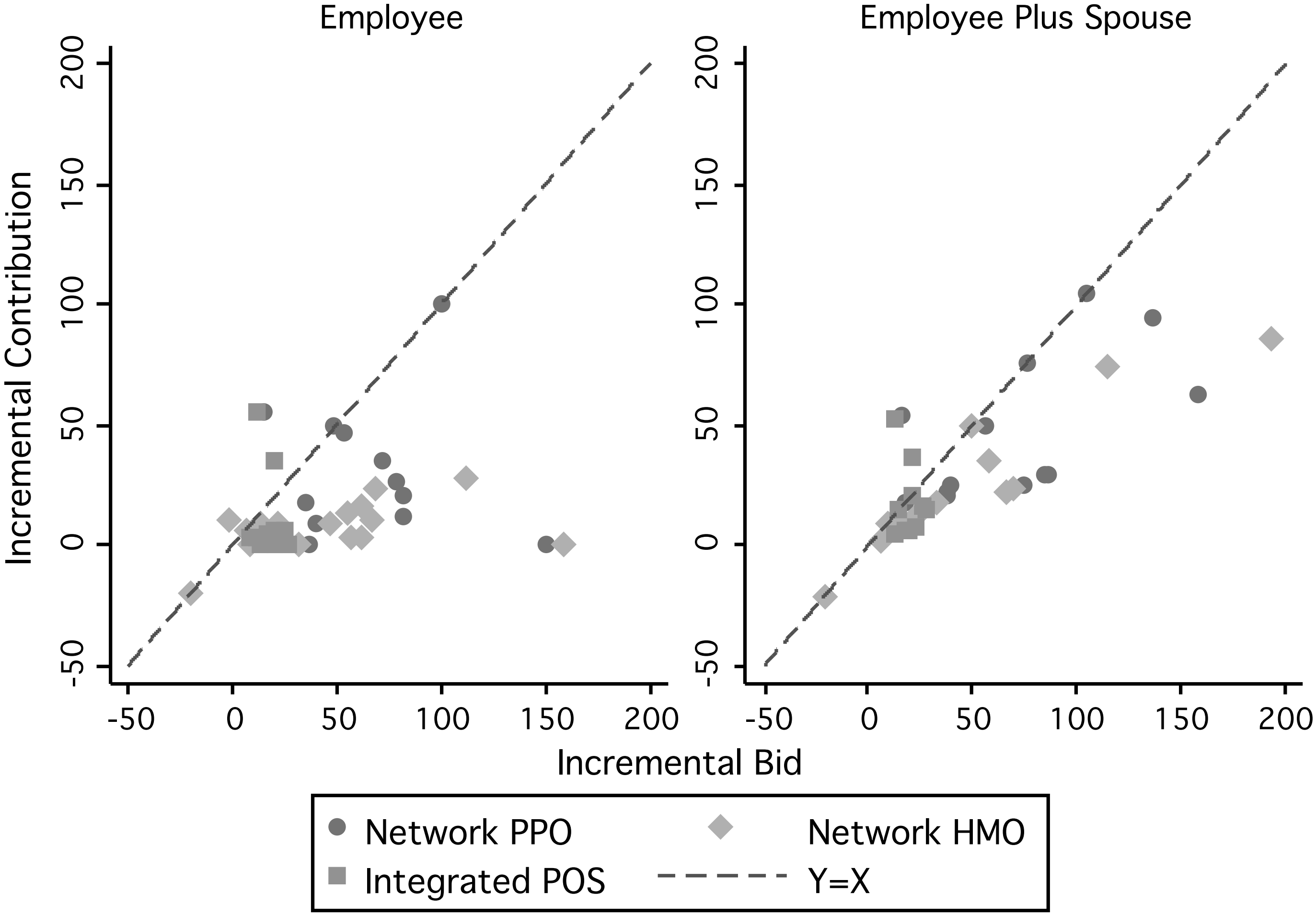
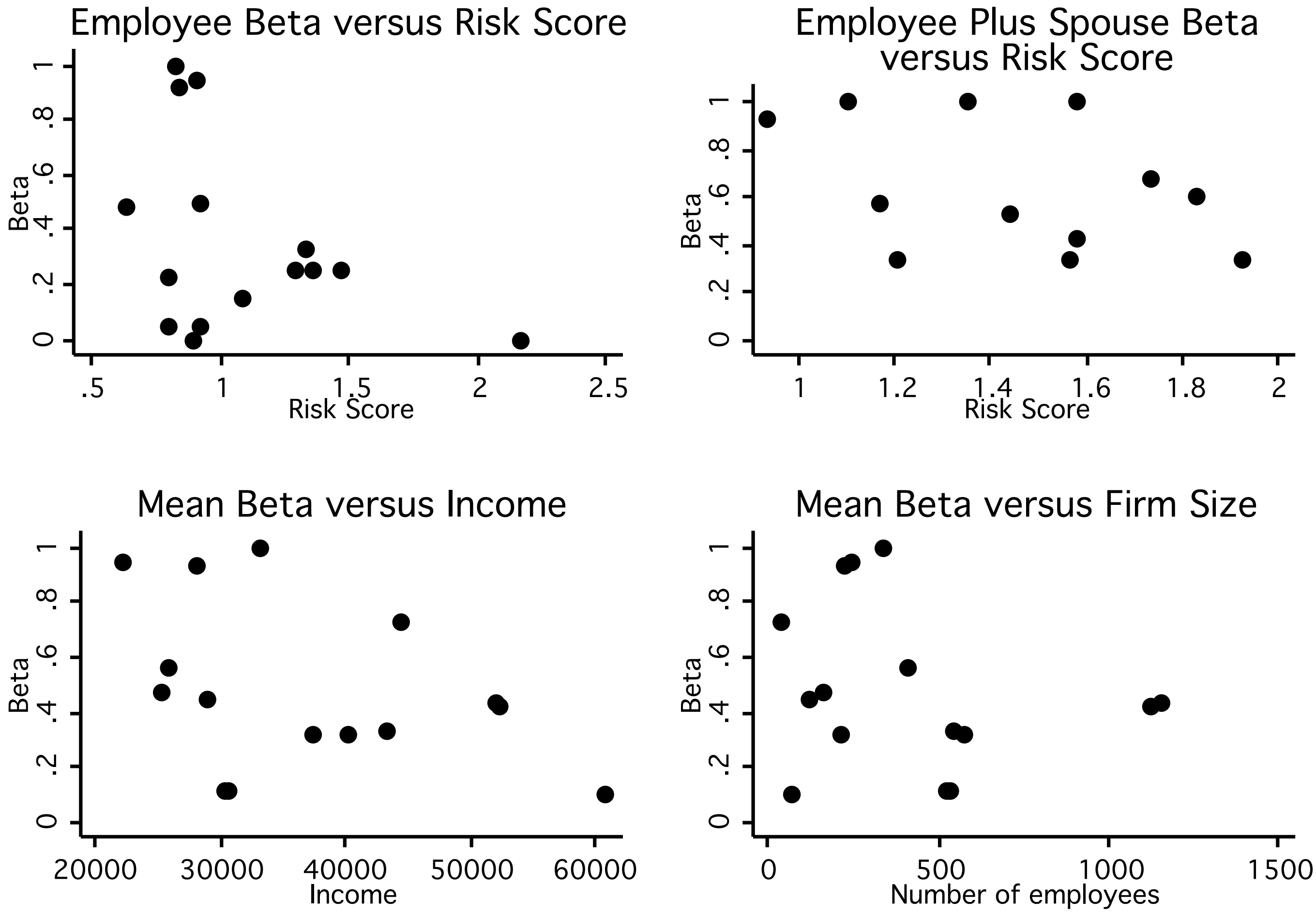


Figure 3: Contributions and Bids Relative to Integrated HMO



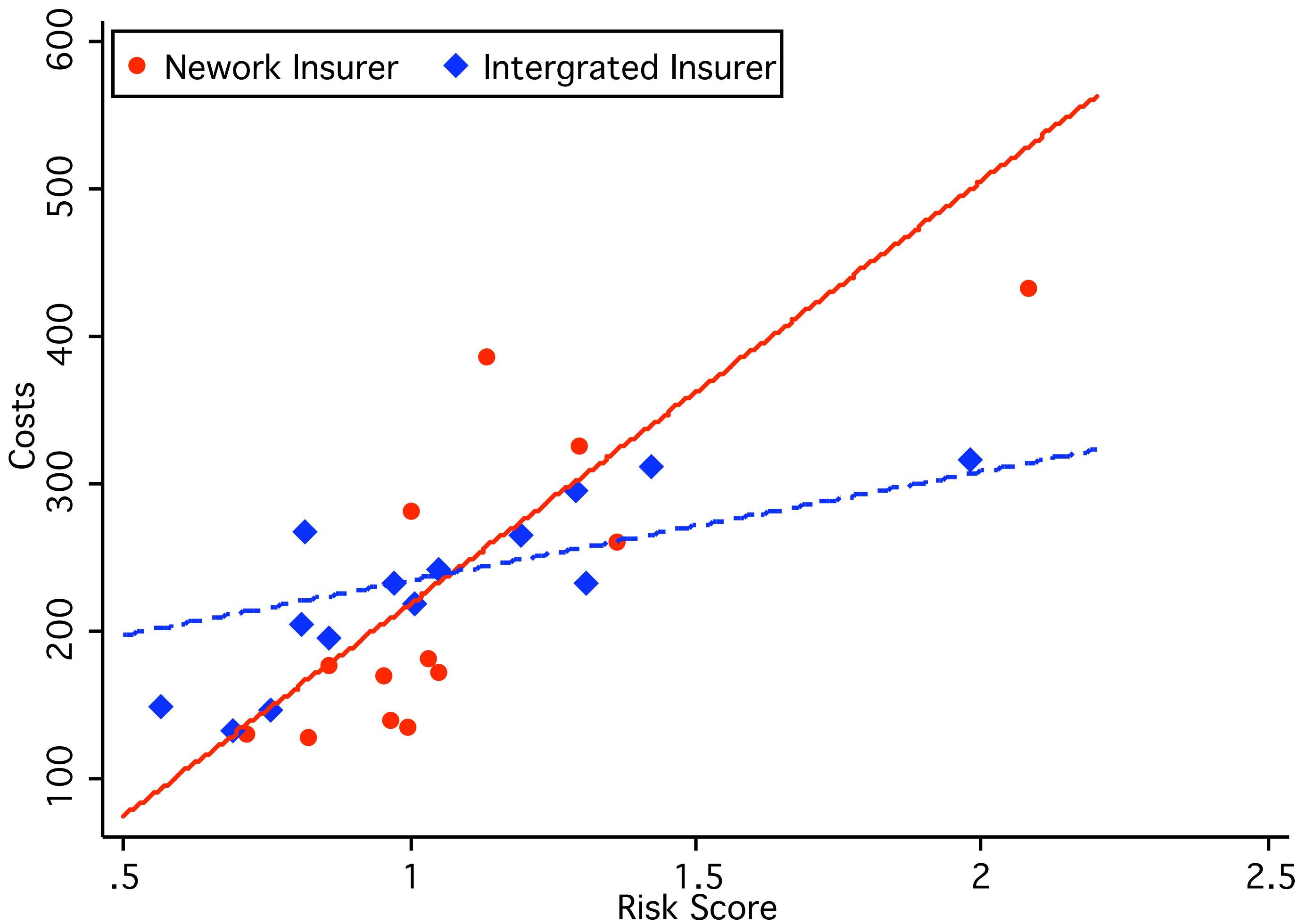
Notes: Incremental Contribution and Incremental Bid are relative to Integrated HMO. In Employee Plus Spouse, numbers are divided by two for comparability.

Figure 4: Employer Contributions and Employee Characteristics



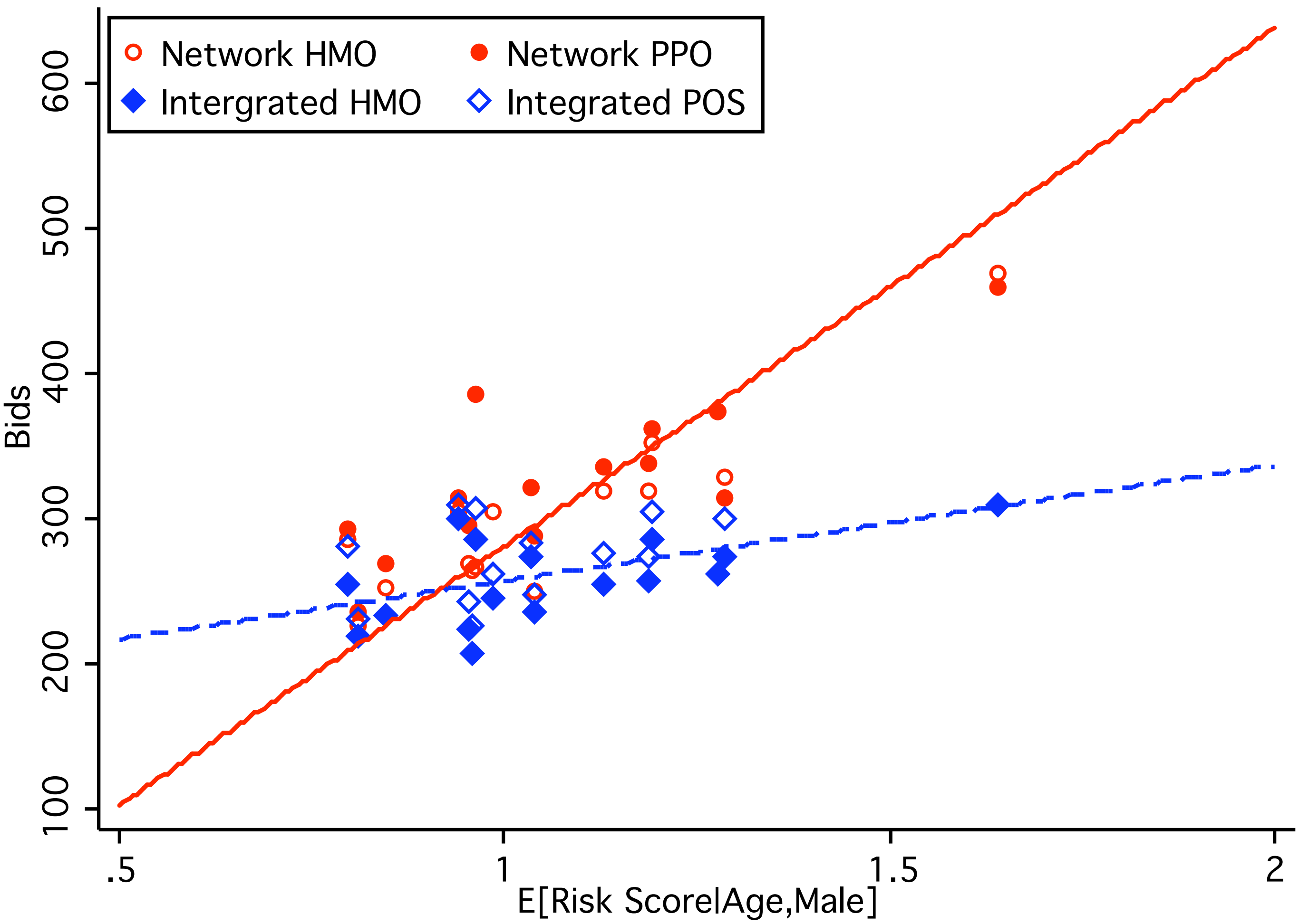
Notes: Each point represents a firm-year. Beta is the incremental pass-through of bids. See text for details. Employee Beta versus Risk Score plots the beta for employees against their mean risk score. Employee Plus Spouse Beta versus Risk Score plots the beta for those in employee plus spouse plans 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-employee-year. Risk Score is the average enrollee risk score in a firm-year. Costs is the insurers monthly cost per enrollee. Fitted lines for the Network HMO and Integrated HMO.

Figure 6: Bids by Risk Score



Notes: Each point represents a plan-employee-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 for the Network HMO and Integrated HMO.

Table A1: Alternative Demand Model Specifications

		Family Structure		Alternative Instruments	
	Baseline	Employees	Family Interactions	Constrained	Predicted
	(1)	(2)	(3)	(4)	(5)
Rescaled Coefficients					
Contribution	-1.00	-1.00	-1.00	-1.00	-1.00
	(.28)	(.36)	(.26)	(.95)	(1.11)
Coinsurance (%)	1.91	0.95	1.73	2.48	1.82
	(.49)	(.47)	(.44)	(1.81)	(.74)
Deductible	-0.01	0.01	-0.01	-0.02	0.00
	(.02)	(.02)	(.01)	(.04)	(.04)
NHMO					
X Risk Score	-1.24	-1.40	-1.63	-1.64	-0.83
	(3.4)	(3.05)	(3.08)	(3.42)	(4.72)
X Age	1.75	1.35	1.49	2.31	1.34
	(.27)	(.27)	(.29)	(.3)	(.16)
X Male	4.93	5.67	1.54	5.64	15.13
	(9.18)	(7.49)	(8.2)	(11.48)	(18.47)
X High Risk	-21.27	-15.06	-14.80	-28.53	-13.92
	(12.76)	(13.18)	(11.62)	(12.32)	(16.02)
X Spouse			-15.35		
			(8.57)		
X Child			-0.65		
			(10.43)		
NPPO					
X Risk Score	-11.07	-1.04	-8.43	-14.27	-10.61
	(6.76)	(4.84)	(5.86)	(7.45)	(7.27)
X Age	0.75	-0.02	0.62	1.00	0.41
	(.45)	(.46)	(.47)	(.65)	(.37)
X Male	-34.64	-15.70	-27.14	-44.26	-31.96
	(14.36)	(12.14)	(13.43)	(14.8)	(11.78)
X High Risk	49.38	7.45	36.23	64.09	46.29
	(19.87)	(21.35)	(17.83)	(19.53)	(23.93)
X Spouse			31.86		
			(13.45)		
X Child			-13.34		
			(16.8)		
IPOS					
X Risk Score	-6.10	-4.83	-6.25	-8.03	-4.70
	(5.41)	(5.05)	(4.94)	(5.09)	(2.8)
X Age	1.58	0.91	1.20	2.09	1.19
	(.39)	(.4)	(.41)	(.38)	(.26)
X Male	-35.24	-25.67	-33.33	-46.64	-23.94
	(12.85)	(10.13)	(11.34)	(14.08)	(7.65)
X High Risk	16.40	24.60	20.97	21.28	15.68
	(17.28)	(19.16)	(15.93)	(16.68)	(11.65)
X Spouse			-13.01		
			(11.43)		
X Child			-9.49		
			(14.33)		
σ_ϵ	109.29	81.66	97.46	145.13	119.91
	-	-	-	-	-
N	3683	2252	3683	3683	3683
Semi-Elasticities					
NHMO	-0.09	-0.08	-0.09	-0.05	-0.06
NPPO	-0.10	-0.10	-0.10	-0.05	-0.05
IHMO	-0.05	-0.05	-0.05	-0.08	-0.09
IPOS	-0.09	-0.09	-0.09	-0.07	-0.06
Total	-0.07	-0.07	-0.70	-0.04	-0.07

Baseline (1) repeats the non-IV specification of Table 4. Family structure is examined in specifications (2) and (3). In Employees (2), the baseline model is estimated on the subsample of employees who enrollee as individuals. In Family Interactions (3), spouse and child by plan interactions are added to the baseline specification. Alternative instruments are examined in specifications (4) and (5) . In (5), we instrument using predicted plan contributions but constrain the parameters in the contribution model to be constant across firms. In (5), the predicted contribution equation is estimated using predicted bids. See Section 4.2 for details. Standardized coefficients are normalized by the coefficient on monthly contributions. 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 $(100 X \text{MarginalEffect})/(12 X \text{MarketShare})$ in percent.