NBER WORKING PAPER SERIES

HEALTH INSURANCE MENU DESIGN FOR LARGE EMPLOYERS

Kate Ho
Robin S. Lee

Working Paper 27868
http://www.nber.org/papers/w27868

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 2020

We thank Mike Chernew, Alice Baird, Rita Moore, and many others at Harvard for supporting this project; Chris Kardos for helpful discussion; and David Cutler, Leemore Dafny, Liran Einav, Keith Ericson, Matthew Gentzkow, Jon Kolstad, Neale Mahoney, Victoria Marone, Ariel Pakes, Stephen Ryan, Mike Whinston, and numerous conference and seminar participants for comments and feedback. Jeff Gortmaker, Jacob Dorn, and Ron Yang provided exceptional research assistance. Robin Lee gratefully acknowledges support from the National Science Foundation (SES-1730063). He advised CalPERS on its health plan offerings in 2020. The authors have no other relevant or material financial interests that relate to the research described in this paper. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Health Insurance Menu Design for Large Employers
Kate Ho and Robin S. Lee
NBER Working Paper No. 27868
September 2020
JEL No. I11,I13,L2

ABSTRACT

We explore the challenges faced by a large employer designing a health insurance plan menu for its employees. Using detailed administrative data from Harvard University, we estimate a model of plan choice and utilization, and evaluate the benefits of cost sharing and plan variety. For a single plan with a generous out-of-pocket maximum, we find that a modest coinsurance rate of approximately 30% with a zero deductible maximizes average employee surplus. Gains from offering choice are limited if based solely on financial dimensions, but can be economically significant if paired with other features that appeal to sicker households.

Kate Ho
Princeton University
Department of Economics
237 Julis Romo Rabinowitz Building
Princeton, NJ 08544
and NBER
kate.ho@princeton.edu

Robin S. Lee
Department of Economics
Harvard University
Littauer Center 120
Cambridge, MA 02138
and NBER
robinlee@fas.harvard.edu

Health Insurance Menu Design for Large Employers*

Kate Ho† Robin S. Lee‡

September 2020

Abstract

We explore the challenges faced by a large employer designing a health insurance plan menu for its employees. Using detailed administrative data from Harvard University, we estimate a model of plan choice and utilization, and evaluate the benefits of cost sharing and plan variety. For a single plan with a generous out-of-pocket maximum, we find that a modest coinsurance rate of approximately 30% with a zero deductible maximizes average employee surplus. Gains from offering choice are limited if based solely on financial dimensions, but can be economically significant if paired with other features that appeal to sicker households.

1 Introduction

Employer-sponsored health insurance covers the majority of the non-elderly U.S. population and accounts for a fifth of the nation's $3.6 trillion in annual health care expenditures. In an attempt to curtail well-documented increases in health spending, employers have become more reliant on forms of enrollee cost sharing—including deductibles, copays, or coinsurance—that partially expose individuals to the price of care. For example, the share of employees facing a deductible increased to 82% in 2019 from 63% a decade prior, and the average deductible for single-coverage plans tripled over the same period.[1] Though requiring households to pay some fraction of their health spending is socially desirable when individuals adjust their health care utilization in response to insurance (Arrow, 1963; Pauly, 1968; Zeckhauser, 1970), the optimal level and form of cost sharing varies across the population, reflecting differences in underlying preferences and health needs.

*Project start date: February 2017. We thank Michael Chernew, Alice Baird, Rita Moore, and many others at Harvard for supporting this project; Chris Kardos for helpful discussion; and David Cutler, Leemore Dafny, Liran Einav, Keith Ericson, Matthew Gentzkow, Jon Kolstad, Neale Mahoney, Victoria Marone, Ariel Pakes, Stephen Ryan, Mike Whinston and numerous conference and seminar participants for comments and feedback. Jeff Gortmaker, Jacob Dorn, and Ron Yang provided exceptional research assistance. Lee gratefully acknowledges support from the National Science Foundation (SES-1730063). Ho has no relevant or material financial interests that relate to the research described in this paper. Lee advised CalPERS on its health plan offerings in 2020, and has no other relevant or material financial interests that relate to the research described in this paper.

†Princeton University, NBER and CEPR, kate.ho@princeton.edu.

‡Harvard University and NBER, robinlee@fas.harvard.edu.

[1] Statistics in this paragraph are from the 2018 and 2019 Kaiser Family Foundation Annual Health Benefits Survey, and CMS 2018 National Health Expenditures. Single-coverage deductibles averaged $533 in 2009 and $1655 in 2019.

In this paper, we explore questions faced by a large self-insured employer responsible for providing health benefits to a heterogeneous population of employees and their dependents. For a single insurance plan, how much cost sharing is optimal, and should it take the form of a deductible (full liability until the deductible is reached), coinsurance (a fixed percentage of spending), or some combination of the two? How much can then be gained by introducing a menu of plan options, and how sensitive is this to how enrollee selection across plans is managed? Much has been written about the theoretical tradeoffs of imposing various forms of cost sharing in insurance markets. Less is known about the mapping between real-world variation in underlying health needs and preferences and the optimal cost sharing for a single plan (Manning and Marquis, 1996; Kowalski, 2015); and the extent to which an employer, able to freely set premiums and cross-subsidize plans, can manage adverse selection (Cutler and Reber, 1998) and achieve gains that a competitive marketplace cannot (Rothschild and Stiglitz, 1976; Handel, Hendel and Whinston, 2015).

We address these questions by developing a framework that can be used to optimize the financial characteristics of the health insurance plans that an employer offers. More broadly, our approach is also useful for a market designer or planner that can regulate premiums and choose financial coverage levels for a set of insurance plans. We return—two decades later—to the same employer setting as the foundational work of Cutler and Reber (1998): the health plan offerings of Harvard University. Following several years of offering insurance plans to its employees with effectively no deductibles or coinsurance for in-network care, Harvard introduced positive cost sharing for its non-union employees across all of its plan offerings in 2015. In 2016, partly in response to backlash and pressure from its constituents, Harvard reintroduced a zero-deductible and zero-coinsurance plan alongside its existing offerings. We obtain data comprising health insurance choices and utilization decisions for over eleven thousand households from 2014–16, and observe how coverage and spending decisions change for both non-union employees who are affected by these plan changes, and union employees who are not. We document significant variation in underlying health and medical spending and provide difference-in-difference evidence that households reduce the amount spent on medical care when exposed to positive levels of cost sharing, with a greater impact on spending for higher-spending households.

Central to our exercise is a two-stage model of of insurance plan choice and health care utilization based on Cardon and Hendel (2001) that incorporates both plan choice under uncertainty and risk aversion, and optimal medical spending with non-linear cost sharing. Our implementation of this model, closely following Einav et al. (2013), allows consumers to vary in their responsiveness to insurance coverage by adjusting their medical spending (moral hazard), and select across plans based on this variation as well as their observed and unobserved determinants of health spending. We estimate this model and recover primitives that include consumers' preferences over premiums, non-financial plan characteristics, and increased risk from cost sharing; their health severities; and the sensitivity of medical spending to coverage. Informed by the variation in the data described above, our model predicts that a majority of households in our population would increase health spending by as much as 30% with no cost sharing (full insurance) relative to what they would spend under a plan with a deductible equal to the out-of-pocket maximum observed in our setting. These large effects are consistent with estimates of moral hazard from other health insurance settings (Einav et al., 2013; Brot-Goldberg et al., 2017; Marone and Sabety, 2020).

Having recovered parameters that capture the preferences and health severities in our population, we conduct a series of simulations that use our estimated model to explore the impact of varying the number of plans, and the plan characteristics, offered to employees. When comparing outcomes across simulations, we focus on the household-average level of employee surplus per year under the requirement that premiums cover total medical spending net of out-of-pocket payments, holding fixed the per-household premium contribution from the employer. Our employee surplus measure—which treats medical spending as a cost—will coincide with social surplus when there are no markups on medical care.[2] This measure also has the advantage of implying that the optimal level of cost sharing is invariant to the fixed level of employer premium contributions, allowing us to side-step issues that arise when employers may substitute between premium contributions and wages.[3] Our simulations use the population of Harvard employees in 2016, and maintain the levels of out-of-pocket maximums observed in the data.

Summary of Findings. We conduct four sets of simulations.

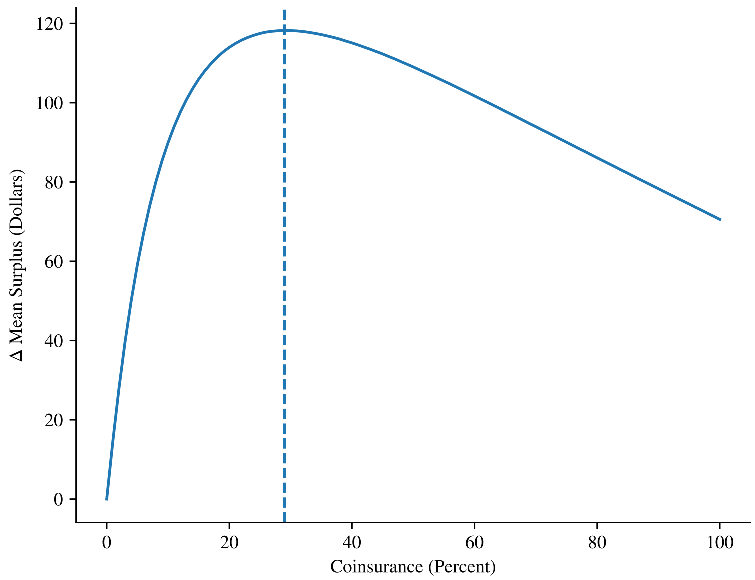
The first set investigates the optimal coverage level when a single plan is offered by the employer. Figure 1 plots our calculated change in average employee surplus as the coinsurance rate for this plan varies with the deductible held fixed at zero. The intuition for the shape of the graph is familiar from the previous literature. From an individual's perspective, increasing the coinsurance rate generates a loss both from an increase in the out-of-pocket cost of care and from increased exposure to financial risk. However, it also reduces the individual's medical spending (moral hazard), and exerts a positive externality across all enrollees through a commensurate reduction in premiums. The steep slope of the curve as coinsurance increases from zero suggests that moral hazard is high relative to risk aversion and other offsetting factors. The optimal coinsurance rate for a single plan with a zero deductible is 29% for our population. The average employee surplus gain from reaching this point, relative to zero coinsurance, is $118 per household per year—equivalent to approximately 10% of annual premium contributions for an individual. Further increases in coinsurance past the optimal level lead to smaller savings from reduced utilization that are outweighed by the cost of increased risk exposure. (We provide details on the construction of this figure, and further discussion, in Section 5.)

Our initial simulations also show that even if an employer could introduce a deductible alongside coinsurance, it would not be optimal to do so. Intuitively, while a positive deductible increases the proportion of households with zero spending, the corresponding optimal coinsurance rate falls,

[2]With positive markups, the socially optimal level of insurance coverage will generally be higher than the one that is optimal for the employer (Ho and Lee, 2020).

[3]In our model based on Cardon and Hendel (2001) and Einav et al. (2013), there are no wealth effects and premiums enter additively into each household's utility; hence employer premium contributions represent only a transfer to employees. Explicitly addressing issues of imperfect substitution between wages and premiums would require a model of the local labor market, and lies outside the scope of this paper.

Figure 1: Change in Average Employee Surplus, Single Plan



Notes: Estimated change in annual household-average level of employee surplus in dollars from offering a single plan with a zero deductible and positive coinsurance rate (horizontal axis), relative to a single plan with a zero deductible and zero coinsurance rate. Out-of-pocket maximums are $2000 per individual and $6000 per household. See Section 5 for details on the construction of this figure.

so high-spending enrollees who spend past the deductible face a lower out-of-pocket price. On net, moving the deductible away from zero with the corresponding optimal coinsurance adjustment leads to small changes in total spending (and in some cases increases spending), resulting in minimal premium reductions whose benefit is more than offset by the loss from increased risk. These findings provide further evidence—using a different model and assumptions on consumer behavior—that offering a high-deductible health plan in isolation may not be desirable for an employer (Brot-Goldberg et al., 2017).

We next investigate the gains that are achievable when an employer can offer multiple plans that vary only in their level of cost sharing. Our analysis here follows two steps. First, we explore what is achievable under what we refer to as *assignment*, where the employer can choose a plan for each household. To obtain a plausible upper-bound on potential gains, we consider the extreme case where each household can be offered a "tailored" plan with a household-specific deductible and coinsurance rate. We document non-trivial heterogeneity in offered coverage levels, reflecting significant underlying heterogeneity in preferences and health needs within the population. The employer can gain 28% more than is achievable with a single plan. If instead the employer is able to assign households to one of only two plans, it can achieve an additional 16% of the average surplus gained under the single-plan optimum, or 58% (≈ 16/28) of the additional gains from household-level tailoring. The optimal coinsurance rates for these two plans imply quite different levels of coverage: 15% and 51%.

While our results under assignment may not correspond to a feasible policy, they do provide a useful benchmark for comparison to our third set of simulations with *selection*, where households are free to choose among the set of offered plans. In these simulations, we allow the employer to stabilize the market, addressing potential adverse selection "death spirals" (Cutler and Reber, 1998) by constraining premiums to reflect only a fraction of the cost difference per enrollee across plans. However, even with this flexibility, an employer is able to achieve only modest gains from offering two plans with selection: only 1% more than the single-plan optimum, or 6% (1/16) of the gain under assignment to one of two optimally designed plans. There are two reasons for these limited gains. The first is a direct result of selection. Households now choose the plan that is privately optimal for them, and not the option that equates their benefits from coverage to their impact on total spending (borne largely by others in the form of higher premiums). Second, the employer optimally chooses a substantial premium subsidy for the more generous plan in order to stabilize the market. This both reduces the extent to which households can be separated across plans, and limits the difference in coverage levels that can be sustained: the optimal pair of coinsurance rates with selection are 20% and 35%.

Last, we consider the potential value from additional firm strategies to help close the gap between the gains with assignment and with selection. Results from our simulations with assignment indicate that larger households and households with members with greater sickness severity are optimally provided with higher coverage in our setting. We find that varying coverage levels based on household size is not very effective: the gains are approximately 2% relative to offering a single plan, only slightly higher than those obtained with two plans under selection. However, our setting suggests a different strategy that takes advantage of underlying plan differentiation based on two salient non-financial characteristics: plan carrier and type. Two carriers are offered, one of which has a denser network of physicians in the suburbs and outlying geographic areas, and is anecdotally viewed as more attractive to families living in those areas, than the other. Point-of-Service (POS) plan types provide access to out-of-network providers, and to specialists without a referral, and hence are typically viewed as more attractive to high-severity consumers than the more restrictive Health Maintenance Organizations (HMOs). Estimates from our plan choice model are consistent with these perceptions. We show that introducing plans that differ on these non-financial dimensions yields significant benefits. When we design an optimal two-plan menu with different carrier and plan-type combinations and different coverage levels, we achieve gains that are statistically indistinguishable from offering two such plans with assignment. That is, we do much better than the 6% of gains achieved in moving from assignment to selection when plans vary only along financial dimensions. In essence, the employer effectively manages selection by pairing more generous coverage with the non-financial characteristics that appeal to larger and sicker households, and hence finds it optimal to sustain greater coverage differences (coinsurance rates of 20% and 60%).[4]

[4]The idea that offering plans differentiated along multiple dimensions can help a firm screen enrollees has, of course, been raised in the previous literature. For example, Veiga and Weyl (2016) note that non-price product features allow an insurer to sort profitable from unprofitable consumers. Our discussion applies the same intuition but holds plan non-financial characteristics fixed. Many of the points in that paper would apply in our setting if insurers could

Broader Implications. Our findings complement those from other papers that have explored the welfare benefits of health insurance plan choice in different settings. Bundorf, Levin and Mahoney (2012) also consider inefficiencies arising from consumer self-selection into plans, this time in the small group market. They point out that uniform pricing, where all individuals face the same menu of prices, may lead to an inefficient allocation of enrollees when consumers with similar health risks have different preferences and different costs of coverage in particular plans. They demonstrate this point in an empirical setting where employees choose between two plans with fixed coverage levels, one with a broad and the other with a narrow network. Marone and Sabety (2020), in a contemporaneous working paper closely related to ours, build on Bundorf, Levin and Mahoney (2012) and examine whether offering "vertical" choice over financial coverage levels, with a uniform pricing scheme, can be welfare-improving in the context of a population of public school employees in Oregon. Using a model based on Einav et al. (2013), and consistent with our findings, Marone and Sabety predict limited gains from offerings solely ranked by financial coverage levels as many consumers select into plans that are not socially optimal.[5] Our contribution relative to these papers is to broaden the set of instruments at the disposal of the employer, essentially maintaining the uniform pricing assumption but allowing a free choice of both coinsurance and deductibles with single and multiple plans (implying that plans that differ only on financial characteristics need not be vertically ranked). We show that the gains from optimizing coverage for a single plan can be substantial, and when paired with other forms of horizontal differentiation, the gains from multiple plans can be meaningfully larger.

Our findings are also related to a literature considering sub-optimal choices made by consumers in the context of health care, which include Handel and Kolstad (2015) on "behavioral hazard," and Abaluck and Gruber (2011) and Ericson and Sydnor (2017) on "choice inconsistencies" and other choice frictions. These papers stress that, if consumers do not make rational decisions, then plan features such as high deductibles and large choice sets may substantially reduce welfare. We view our results as complementary to this literature: we show that positive deductibles (and not just high deductibles) may harm welfare, and that offering even a small choice set of plans raises issues that often limit potential gains from plan variety. The framework that we use can also be extended to account for the possibility that consumers make sub-optimal choices. For example, if behavioral hazard is a concern, the employer's objective can be modified to account for potentially harmful utilization reductions in response to enrollee cost-sharing. Such a change might reduce the optimal coinsurance rate, but would not affect our finding that a zero-dollar deductible is optimal, nor that non-financial plan differentiation can help manage selection.[6]

respond to the employer's coverage choices by changing their networks or other attributes. We also find that other forms of horizontal plan differentiation may be beneficial, separately from helping an employer manage selection (see Section 5.5). This is consistent with Dafny, Ho and Varela (2013), who use data on over 800 large employers and a multinomial logit model of employee choice to show that employee surplus would improve measurably if the menu included a more diverse choice of plan types and carriers.

[5] In their main analysis, Marone and Sabety (2020) focus on four metal tier plans, and find no social gains from offering multiple plan types. In robustness analyses, they expand the choice set to include 40 vertically ranked plans, and find a small gain from offering choice in the neighborhood of the Gold plan.

[6] We show in Section 2 that spending reductions in response to increased coverage are concentrated in relatively

We note several additional caveats. First, Harvard offers fairly generous (low) out-of-pocket maximums; the optimal coinsurance rate would likely be lower in other settings with greater overall risk exposure. Second, we do not consider the extensive margin of selection into insurance, focusing only on households who enroll in coverage (Geruso et al., 2020; Saltzman, forthcoming). Third, while we allow for plan-type and carrier differentiation—and allow households with different health needs to have different preferences for these characteristics—we do not endogenize them. The issue seems particularly relevant for network breadth, and may also generate additional considerations: for example, if carriers respond to the use of network differentiation to manage selection by adding or dropping medical providers, this will affect provider price negotiations (Ho and Lee, 2017, 2019). Relatedly, if cost sharing increases price-responsiveness at the point-of-care, negotiated provider payment rates may decrease (Gowrisankaran, Nevo and Town, 2015; Brown, 2019; Prager, forthcoming). We also do not consider the possibility that an employer could improve its ability to negotiate lower premiums or administrative fees with insurers by offering multiple plans from different carriers (Cutler and Reber, 1998). Last, we do not address the impact of plan design on re-classificiation risk and year-over-year dynamic considerations (Ghili et al., 2020).

Despite these features that lie outside our model, our analyses deliver several clear findings that are still useful for employers and others designing health plan menus. First, gains from optimizing the level of cost sharing for a single plan can be meaningful. A moderate coinsurance rate with no deductible is optimal in our setting. Second, if consumers can choose among multiple plans, the divergence between a household's willingness to pay for insurance and its efficient coverage level (documented also in Marone and Sabety (2020)) may imply only small gains from offering multiple plans that are differentiated solely on financial characteristics, even when an employer is able to cross-subsidize premium levels across plans. Offering plans that are differentiated along both financial and non-financial dimensions may allow the employer to better manage enrollee selection and generate substantially larger gains. In our setting, pairing more financially generous plans with characteristics that attract enrollees for whom more generous plans are socially efficient yields economically significant benefits.

2 Empirical Setting: Harvard University, 2014–16

We examine the health insurance choices and medical utilization of employees and their dependents at Harvard University between 2014 and 2016.

2.1 Background

Plan Offerings. Harvard offers a range of health plans to its employees. In 2014, the plan menu for both union and non-union Harvard employees included four plans in total: HMO and POS plans from both the Harvard University Group Health Plan (HUGHP) and from Harvard Pilgrim (HP).[7]

low-value care, in contrast to Brot-Goldberg et al. (2017) who find that utilization falls significantly across both low-value and high-value care when enrollees are moved into a high-deductible health plan.

[7]There was also a PPO option for non-Massachusetts residents, who are excluded from our sample.

HUGHP is an integrated plan that employs salaried primary care physicians located close to the university in Cambridge, Massachusetts.[8] HP is a non-integrated plan with a more geographically dispersed network of PCPs. Patient cost sharing was limited in all plans in 2014: deductibles for in-network providers were zero, patients paid a small fixed amount per outpatient visit rather than a percentage of the total cost of their care, and they paid nothing for inpatient care.[9]

In 2015, in an effort to control health care spending, Harvard increased deductibles in the HMO and POS plans offered to non-union employees to \$250 per individual (with a \$750 maximum per family) for in-network care, and introduced a 10% coinsurance rate for in-network inpatient care.[10] This adoption of increased cost sharing led to documented dissatisfaction among Harvard faculty and employees.[11] Partly in response, Harvard added a higher-coverage "POS-Plus" plan to its offerings in 2016; this POS-Plus plan had a zero deductible and no out-of-pocket (OOP) payments for hospital care. Union employees' plan menus and cost-sharing terms did not change during our study period. As we discuss in our empirical application, this variation in plan contracts between union and non-union employees is important for recovering the underlying health needs and preferences of our population.

Premiums. Because Harvard is a self-insured employer, there is no competition between insurers over premium levels (as there would be for fully-insured employers or on an insurance exchange). Each year, Harvard sets the premiums for each plan to cover anticipated medical spending for all employees, net of any OOP payments made by its employees due to cost sharing.[12] Harvard treats all employees as belonging to the same risk pool, regardless of what plan they enroll in, and thus does not require that the premiums charged for a given plan cover that particular plan's expected costs. Harvard also provides a health premium subsidy to each household, based on salary tier and household size and equal to a fixed percentage of the lowest-priced plan offering.

2.2 Data and Plan Details

Enrollment and Sample Characteristics. Our sample comprises households (i.e., employees and their dependents) that enroll in a plan offered by Harvard and are present in our data for at least a full calendar year.[13] For each year the employee is present, we observe the menu of health plan

[8]Specialty care in the HUGHP plan is provided by physicians in the larger Blue Cross Blue Shield network.

[9]By comparison, a 2014 survey by the Kaiser Family Foundation and Health Research and Educational Trust found that 80% of covered workers employed by firms with over 200 employees faced a deductible; 32% of such workers had at least a \$1000 deductible; and for those without a deductible, over 77% had patient cost sharing for a hospital admission (Claxton et al., 2014).

[10]Harvard also introduced a high deductible health plan (HDHP) for each carrier for non-union employees in 2015: these plans had lower premiums than the HMO and POS options, but higher deductibles (set at \$1500 per individual or \$4500 per family) and in-network coinsurance rates (15%). These plans were unpopular: fewer than 2% of employees chose the HDHP plan options.

[11]This debate received attention in the national press: see, for example, "Harvard Ideas of Health Care Hit Home, Hard," New York Times, January 5, 2015, and "Harvard Professors are Angry That Their Amazing Health Insurance is Getting Slightly Less Generous," Washington Post, January 5, 2015.

[12]There is also a small administrative markup that we ignore for simplicity of exposition and analysis.

[13]We exclude employees who opted out of Harvard insurance either in 2014 (16% of non-union employees) or later. The proportion did not change meaningfully during our time period as coverage options changed. We also do not

Table 1: Summary Statistics

	Full Sample	Base (2014-5) Sample	Base Sample: Non-Union	Base Sample: Union
# Employees (Households)	11,188	7,707	4,803	2,904
Mean Age	40.5	43.9	45.3	41.6
Mean Tenure	8.3	11.3	11.9	10.2
% Male	47.5	47.8	52.3	40.4
Faculty	15.1	17.0	27.3	0.0
2-Person Family	15.6	17.4	18.5	15.6
> 2-Person Family	28.0	32.4	37.6	23.7
Mean Family Size (N>2)	3.79	3.82	3.84	3.77
Mean 2014 Severity Score		91.1	86.0	99.5
Mean 2014 Spending		$7,756	$8,128	$7,141

Notes: Summary statistics for full sample (used for model estimation) and "base sample" of employees who are present in both 2014 and 2015 (used for most descriptive analyses). Mean Tenure is number of years employed by Harvard; 2-person families include both couples and a single adult plus a child. Severity score is generated using the Diagnostic Cost Group (DCG) risk assessment method. Mean 2014 spending is the average annual spending (both household OOP and insurer contributions) on facilities and professional services. See Section 2.2 for details.

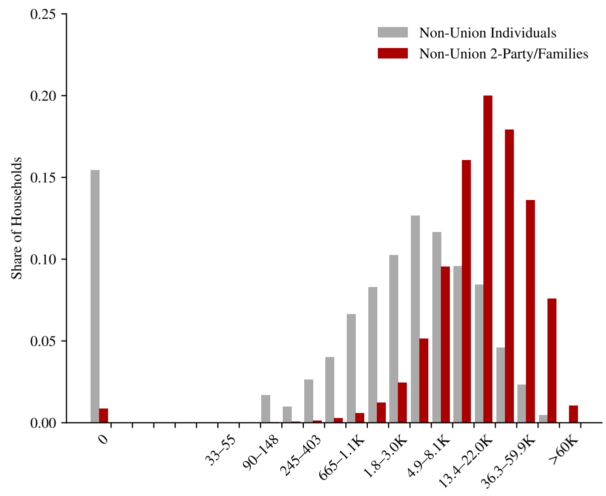
options that are available, the employee's plan choice (made during the November open-enrollment period of the previous year), and claim-level information on the household's medical utilization and spending for the relevant year. We also observe demographic information including age, sex, employment type, salary band, and the number and ages of other family members. Finally we obtained a summary variable used by the employer to predict each individual's health status, utilization and spending in the relevant year. This summary variable, referred to as a severity score, is generated using the Diagnostic Cost Group (DCG) risk assessment method licensed through DxCG, Inc., together with detailed medical claims data. The predictions condition on an individual's age and sex, and take into account the conditions and diseases for which the individual previously received treatments; they also account for variation in how long different medical conditions observed in the claims data are expected to persist.[14]

Column 1 of Table 1 provides summary statistics for our full sample of 11,188 employees. Column 2 provides statistics for what we refer to as the "base sample" for certain descriptive analyses that we conduct in Section 3; this base sample contains 7,707 employees from the full sample who are present in both 2014 and 2015. The two samples are similar. The base sample is 48% male, with an average age of 44; 17% are faculty. Just over half of employees have single coverage, 17% are in two-party households, while the remainder are in households of three or more individuals (and for these employees the household size is 3.8 on average). Annual household

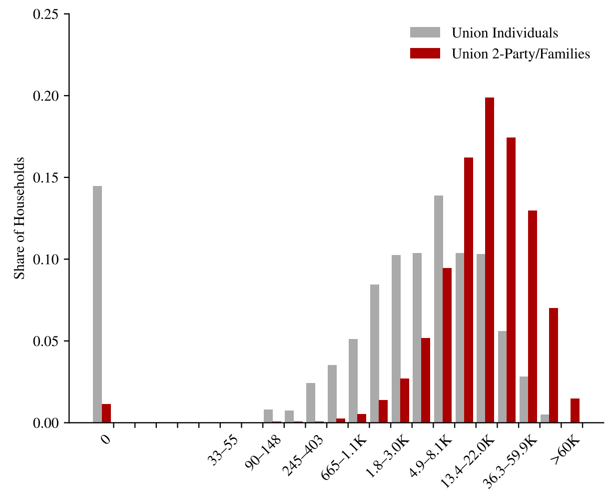
include enrollees with a qualifying event that allowed them to switch plans mid-year, since such events may affect spending in ways we do not model; seasonal employees; retirees; and non-Massachusetts residents. Finally, we exclude 280 employees who chose a high-deductible health plan, and 149 employees who switched insurance carriers during our sample period. HDHP take-up was low and employees occasionally switched plans within-carrier but very rarely switched between HUGHP and HP; we do not model the preferences of these small number of households.

[14]The data provider's prospective DCG tool generates severity scores that correspond to the academic calendar. For calendar year t, we use the score that corresponds to the July t – June $t+1$ academic year which is computed using claims from July $t-1$ to June t.

Figure 2: Medical Spending (2014)



(a) Non-union Households



(b) Union Households

Notes: Histograms of 2014 medical spending in dollars for individuals (light) and 2-party and family households (dark), by union status. These figures use a log scale: each bin $k = 1, \ldots, 22$ corresponds to spending in the range $exp(0.5 \times (k-1))$ – $exp(0.5 \times k)$, with all spending above $exp(0.5 \times 22) \approx 60K$ contained in last bin. The labels on the x-axis show the corresponding dollar amounts for each bin.

spending on facilities and professional services, which includes both household OOP payments and employer contributions, averages \$7,756 per employee. The prospective severity score (defined at the individual level) is normalized so that the expected score for a random draw from a nationally representative population equals 100. We report the average across all households of the mean severity score within each household. On average, households in our sample have an average severity score that is 9% lower than the benchmark.

The remaining columns of Table 1 report summary statistics for the base sample, split by union status. There are 4,803 non-union employees and 2,904 union employees. Union employees tend to be younger; are less likely to be male; and none are faculty. They are much more likely to be single and less likely to be in families of three or more. Reflecting these differences, their average annual family spending is lower (at \$7,141 compared to \$8,128 for non-union employees) despite their average severity score being relatively higher (99.5, compared to 86.0 for non-union employees).

Figure 2 shows the distribution of log medical spending in the first year of our data in more detail. The left and right panels contain the union and non-union samples; pale bars denote individuals while dark bars contain two-party and larger households. In both samples, individuals have a significantly higher probability of zero spending than larger households (15.4% compared to 0.9% for non-union employees) and the spending distribution conditional on positive spending is shifted to the left for individuals. As is typical in such datasets, the spending distribution is highly skewed: for example, mean non-union family spending is almost double the median value (\$12,026 compared to \$7,023). Conditional on household size, union employees are observed to have higher spending than non-union employees, consistent with their higher severity scores. This is reflected

Table 2: Insurance Plan Options

	2014		2015		2016		
	HMO	POS	HMO	POS	HMO	POS	POSP
Individual Coverage							
Share	80%	20%	79%	21%	76%	19%	5%
Employee Premium (HP, T1)	$94	$123	$91	$120	$98	$129	$140
Deductible	$0	$0	$250	$250	$250	$250	$0
Coinsurance	0%	0%	10%	10%	10%	10%	0%
Copay	$20	$20	$20	$20	$30	$30	$30
OOP Max	$2000	$2000	$1500	$1500	$1500	$1500	$2000
Avg. Share OOP	7%	7%	35%	35%	37%	37%	10%
Family Coverage							
Share	75%	25%	75%	25%	70%	20%	10%
Employee Premium (HP, T1)	$256	$337	$250	$329	$264	$348	$379
Deductible (Per Person)	$0	$0	$250	$250	$250	$250	$0
Deductible (Per Family)	$0	$0	$750	$750	$750	$750	$0
Coinsurance	0%	0%	10%	10%	10%	10%	0%
Copay	$20	$20	$20	$20	$30	$30	$30
OOP Max (Per Person)	$2000	$2000	$1500	$1500	$1500	$1500	$2000
OOP Max (Per Family)	$6000	$6000	$4500	$4500	$4500	$4500	$6000
Avg. Share OOP	6%	6%	25%	25%	27%	27%	8%

Notes: Insurance plan characteristics for non-union enrollees by plan type, 2014-16. In each year, each plan type was offered by both HUGHP and Harvard Pilgrim. Premiums are monthly employee contributions (and do not include employer contributions), and reported for Harvard Pilgrim plans for a household in salary tier 1. Copays apply to physician office visits only. "Avg. share OOP" is the simulated average share of spending paid OOP across the whole sample (excluding premium); see main text for details.

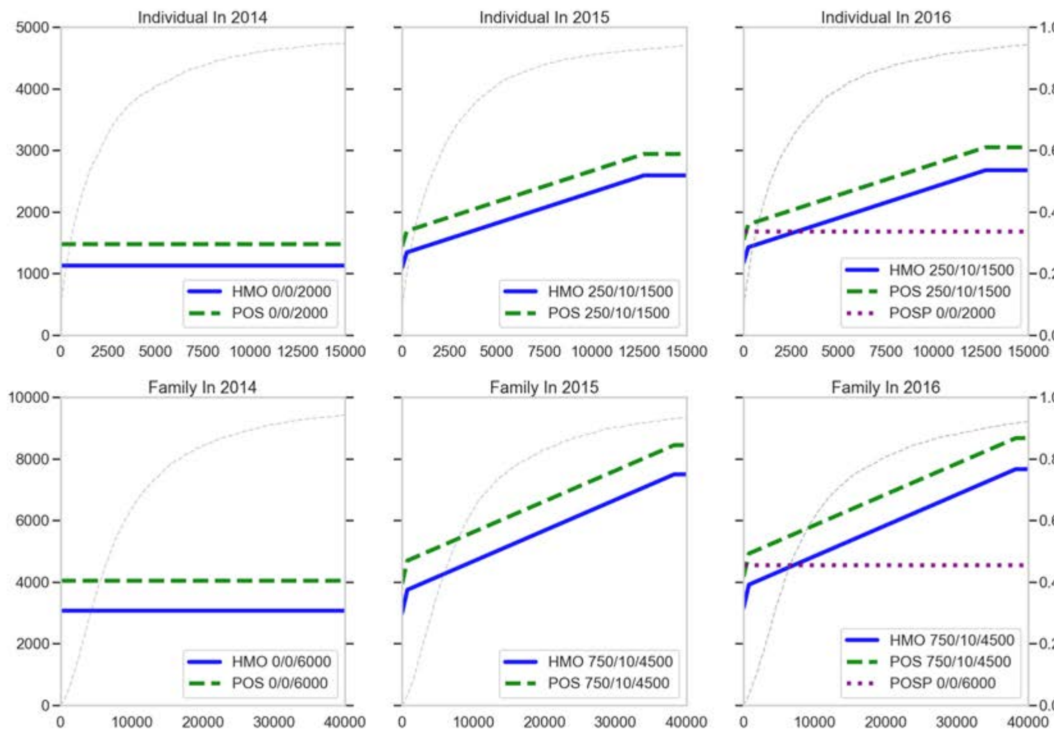
in both the probability of zero spending and the mean of the log spending distribution.

Health Plan Options. Table 2 summarizes the financial characteristics of the plan-type options offered in each year, and the fraction of employees who chose each option in our full sample.[15] In each year, each plan type was offered by both HUGHP and HP. Each enrollee pays the full cost of their medical utilization until they meet the deductible. From that point they pay a fixed percentage of their costs, determined by the coinsurance rate, until their total OOP spending reaches the OOP maximum. In 2014, deductibles for in-network providers were zero in all plans, and non-preventive office visits and outpatient events were priced at a $20 copay, with a zero coinsurance rate for inpatient or outpatient in-network care. In 2015, the HMO and POS plan deductibles and coinsurance rates increased, while the OOP maximum was reduced from $2000 to $1500 per person. In 2016, the POS-Plus plan was added to the choice set; its financial characteristics essentially mirrored those of the HMO and POS plans in 2014, with a zero deductible and no coinsurance (though a higher copay for physician office visits).

We summarize the variation in consumer cost-sharing requirements across the different plans

[15] Employees can choose between individual and family (i.e., two or more individuals) coverage; for each there are three premium levels based on the employee's salary tier. Our analysis treats family size and income as exogenous. We distinguish between two-party households and households containing three or more members since these are likely to have different spending levels, but in practice (and in the model) they are offered the same set of "family plan" options.

Figure 3: Plan Generosity



Notes: Straight and dashed-lines denote total annual OOP costs including premiums (left vertical axis) as a function of in-network spending (horizontal axis) for non-union employees under different plan types, by individual or family coverage and year. Only premiums, deductibles, coinsurance, and OOP maximums are considered; copays, which depend on the number of office visits, are not included. Premiums are for Harvard Pilgrim plans for a household in salary tier 1. Dotted grey line represents the cumulative distribution (right vertical axis) of annual in-network spending for individual or family households, by year.

and years in two ways. First, following Einav et al. (2013), we use each plan's rules to simulate the average share of medical spending that would be paid out-of-pocket, excluding premiums, under every plan for every individual in the data (not just those enrolled in the relevant plan), given their observed medical utilization in a particular year. We report the average of this share across all employees in the data in the final row of Table 2. The change in cost sharing between 2014 and 2015 is apparent: the average share of medical spending paid by an employee through some form of cost sharing in a family plan in 2014 was 6%, increasing to 25% in 2015. This remained approximately the same for HMO and POS plans in 2016; however, the POS-Plus plan introduced that year had more generous coverage, with the average share of spending covered by the employee being 8%.

Second, in Figure 3, we graph the annual OOP costs including premiums as a function of total in-network medical spending under each plan, by individual or family coverage and by year. In 2014, with essentially no cost-sharing (except for copays, not graphed), annual OOP costs are flat with respect to medical spending, consisting only of employee premium contributions. In subsequent years, the cost sharing schedule exhibits non-linearities for HMO and POS plans, both when a household's medical spending exceeds the deductible, and when it passes the OOP maximum. For any amount of medical spending, the HMO plan results in lower OOP costs than the POS plan; however, as noted above, the POS plan has other non-financial characteristics (e.g., policies for seeking out-of-network or specialty care) that are not represented in these graphs, but will be accounted for in our subsequent analysis. Also, the POS-Plus plan in 2016 neither financially dominates nor is dominated by either HMO or POS plan. For example, due to its higher premium, the POS-Plus plan results in higher OOP expenditures for low amounts of medical spending, but its high coverage implies it has lower OOP expenditures for higher amounts of spending. Overall, however, the POS-Plan—like the 2014 plans it emulated—can again be viewed as financially generous: using the plotted cumulative distribution of annual spending for individuals and households in the graphs (dotted grey lines), we see that the POS-Plus plan would minimize OOP expenditures for approximately 30% of individual and 55% of family households in 2016.

3 Descriptive Evidence

We now provide descriptive evidence for selection on health status, moral hazard, and risk aversion in our population.

3.1 Adverse Selection and Moral Hazard

Households choosing high-coverage plans. We begin with a variant of the positive correlation test proposed in Chiappori and Salanie (2000), and look for the presence of adverse selection and moral hazard in the population by examining whether individuals who choose more comprehensive insurance coverage have higher health severity scores and medical spending.

We consider only non-union employees for this analysis, and add data on spending changes over time to help distinguish between adverse selection and moral hazard. We compare the severity scores and spending of the subsample of 1,576 non-union employees who are in the data in both 2015 and 2016; who are enrolled in a POS plan in 2015; and who choose either a POS or a (higher-coverage) POS-Plus plan in the following year. The intuition is that if selection on health status is important, then the households who choose the POS-Plus options when they become available in 2016 are likely to be relatively sick: they may have high severity scores and high realized spending in previous years relative to the full sample. Differential spending *trends* for households choosing the POS-Plus options relative to other similar households would be consistent with consumers adjusting their medical spending in response to cost sharing. Following Einav et al. (2013), we refer to this behavior as moral hazard.

Table 3 describes the sample of households who enroll in a POS plan in 2015, and also enroll in either a POS or POS-Plus plan in 2016. Consistent with selection on health status, the average 2015 severity score for the 1,072 households remaining in a POS plan in 2016 is 103.4, while that for those who choose a POS-Plus plan is substantially higher at 119.5. Average spending for the two populations in 2015 (prior to any plan switching) follows the same pattern: households who

Table 3: Selection Into Plan Coverage

	2015 POS Sample	POS (2016)	POSP (2016)
# Employees	1,576	1,072	504
% Two-Party	20.7%	20.7%	20.8%
% Family	41.6%	38.6%	47.8%
2015 Avg. Severity Score	108.6	103.4	119.5
	(2.8)	(3.5)	(4.4)
2015 Spending	$11,827	$11,039	$13,503
	($461)	($559)	($809)
2016 Spending	$14,106	$11,970	$18,648
	($573)	($595)	($1,247)

Notes: Comparison of DCG severity scores and annual family spending for the subsample of non-union enrollees who are in the data in both 2015 and 2016; are enrolled in a POS plan in 2015; and who choose a POS or POS-Plus plan in 2016. Standard error of the mean reported in parentheses.

enroll in a POS-Plus plan in 2016 spend approximately $2,500 more in 2015 than those who remain in a POS plan.

The data also provide suggestive evidence of moral hazard. The average spending of POS-Plus enrollees increases from $13,503 in 2015 to $18,648 in 2016 (a 38% increase), while spending for those who remain in a POS plan increases from $11,039 to $11,970 (8%). While these summary statistics do not control for factors such as underlying population spending trends and differences in demographics across samples, all of which will be addressed by the model set out below, they still provide useful suggestive evidence that both selection and moral hazard may be important components of consumer behavior.

Moral Hazard. Following Einav et al. (2013), we provide additional evidence for the presence of moral hazard in our population by conducting a difference-in-differences analysis comparing spending across groups of employees for whom insurance coverage did and did not change. In our setting, we compare the mean, median and percentiles of spending in 2014 and 2015 for non-union employees (our treatment group, whose coverage was reduced during this period) to union employees (the control group, whose coverage did not change).[16]

Table 4 presents spending statistics for individuals (top panel), and for two-party households and families (bottom panel). Conditional on household type, union employees have higher mean and median spending than their non-union counterparts in the first year of the sample (see Figure 2). As the table shows, this difference persists throughout the distribution. Furthermore, the spending statistics show patterns consistent with moral hazard: spending increases for the control group (union employees) while it increases much less, or even falls over time, for the treatment group exposed to greater cost sharing. The relative reduction in mean spending is quite large, at $576 (or 15.1%) for individuals and $1,001 (or 7.7%) per year for families.

[16]Non-union employees' coverage was reduced by the same amount between 2014-15 whether or not they switched plans, while union employees' coverage was unchanged and equal across plans. Less than 2% of non-union employees and 1.5% of union employees switched plans between the two years.

Table 4: Difference-in-Difference Results

	Observations	Mean	Fraction with zero spending	Percentile of Spending Distribution 10th	25th	50th	75th	90th
Individuals								
Control (Union)								
2014 Spending	1,763	3,644	0.143	0	349	1,364	3,645	8,591
2015 Spending	1,763	4,465	0.113	0	413	1,548	4,217	10,347
Treated (Non-Union)								
2014 Spending	2,108	3,276	0.154	0	256	1,093	3,065	7,586
2015 Spending	2,108	3,520	0.142	0	257	1,124	3,285	7,698
Treated-Control Differences (Levels, Non-Union - Union)								
2014 Difference		-369	0.011	0	-92	-271	-579	-1,005
2015 Difference		-945	0.028	0	-156	-424	-932	-2,649
2015-2014 Differences (Levels)								
Control (Union)		820	-0.029	0	64	184	572	1,756
Treated		244	-0.012	0	1	30	220	112
Difference-in-Differences (Levels)		-576	0.017	0	-64	-154	-353	-1,643
Difference (Percentages)								
Control		22.5%	-20.6%	-	18.5%	13.5%	15.7%	20.4%
Treated		7.4%	-8.0%	-	0.3%	2.8%	7.2%	1.5%
Difference-in-Differences (Percentages)		-15.1%	12.6%	-	-18.2%	-10.7%	-8.5%	-19.0%
Two-Party and Families								
Control (Union)								
2014 Spending	1,141	12,544	0.011	1,623	3,525	6,796	13,990	28,643
2015 Spending	1,141	14,259	0.01	1,836	3,487	7,072	14,929	32,978
Treated (Non-Union)								
2014 Spending	2,695	11,923	0.009	1,705	3,518	6,924	13,719	27,244
2015 Spending	2,695	12,637	0.009	1,611	3,384	6,850	13,772	27,749
Treated-Control Differences (Levels, Non-Union - Union)								
2014 Difference		-621	-0.003	82	-6	128	-271	-1,399
2015 Difference		-1,622	-0.001	-225	-102	-222	-1,157	-5,229
2015-2014 Differences (Levels)								
Control (Union)		1,715	-0.002	213	-38	276	940	4,335
Treated		714	0.0	-94	-134	-73	53	504
Difference-in-Differences (Levels)		-1,001	0.002	-307	-96	-349	-886	-3,830
Difference (Percentages)								
Control		13.7%	-15.4%	13.1%	-1.1%	4.1%	6.7%	15.1%
Treated		6.0%	4.3%	-5.5%	-3.8%	-1.1%	0.4%	1.9%
Difference-in-Differences (Percentages)		-7.7%	19.7%	-18.6%	-2.7%	-5.1%	-6.3%	-13.3%

Notes: Basic difference-in-difference results for individuals (top panel) and two-party households and families (bottom panel) that comprise the 'base (2014-5) sample' summarized in column 2 of Table 1.

Table 5 presents estimates from a series of OLS difference-in-difference regressions. We perform the regressions both when the dependent variable is measured in levels and when it is measured in logs, and conduct them separately for individuals and families.[17] For individuals, both specifications

[17] We cannot estimate the model in simple logs because of the high proportion of enrollees with zero spending. Instead we follow Einav et al. (2013) by measuring spending in this specification as $log(1 + \text{spending})$.

Table 5: Difference-in-Difference Results, OLS Regressions

	Individual		Two-Party/Family	
	Levels	Logs	Levels	Logs
	(1)	(2)	(3)	(4)
Estimated treatment effect	-576.50^{**}	-0.20^{**}	$-1,001.15$	-0.10^{*}
	(264.83)	(0.09)	(667.14)	(0.05)
2015 fixed effect	820.47^{***}	0.30^{***}	$1,714.86^{***}$	0.08^{**}
	(195.50)	(0.06)	(578.46)	(0.04)
Union fixed effect	368.84^{*}	0.23^{**}	620.75	-0.02
	(220.57)	(0.09)	(574.94)	(0.05)
Constant	$3,275.56^{***}$	6.15^{***}	$11,923.33^{***}$	8.76^{***}
	(148.83)	(0.06)	(296.43)	(0.03)
Mean dependent variable	3,697	6.35	12,614	8.77
Observations	7,742	7,742	7,672	7,672

Notes: Difference-in-difference estimate of the spending reduction associated with the decline in coverage experienced by non-union employees between 2014-15, relative to union employees. Dependent variable is the total annual medical spending for the household (or log of 1 + total spending for log specifications). The estimated treatment effect is an interaction between the 2015 fixed effect and the Union fixed effect. *p<0.1; **p<0.05; ***p<0.01

show that the reduction in coverage generated a statistically significant reduction in spending relative to the control group of approximately $580 in the linear specification, or 20% in the log specification. The estimated effects for two-party and family households are smaller in percentage terms and less precisely estimated. In levels we estimate a statistically insignificant relative spending reduction of $1000 per family; in logs we estimate a 10% relative reduction that is significant at p=0.10.

The greater significance of results for the log specification is consistent with changes in spending in response to cost sharing being a function of the level of underlying spending. In other words, rather than all households increasing or decreasing their spending by the same amount in response to cost sharing (which would be consistent with the specification measuring the dependent variable in levels), households who spend more adjust their spending by more (which is consistent with the log specification). We investigate this further in Table B1 by repeating the difference-in-difference comparisons from Table 4, for individuals and families, focusing just on households (in treatment and control groups) whose mean health severity DCG score is in the highest quartile of the distribution for their family type and year. The top panel of Table B1 reports results for individuals. While the mean relative spending reduction for the full sample was $576, that for the highest severity quartile is three times larger at $1,575. The comparison for families in the bottom panel of Table B1 is also notable: the $1,001 average relative spending reduction for the full sample increases to $1,516 for the highest-severity individuals. The differences in both tables continue throughout the spending distribution. These findings suggest that the amount a household adjusts its spending in response to cost sharing may be greater for sicker households, perhaps because these families have more contact with the medical system, and hence more opportunities to make decisions (on the intensive or extensive margin) that increase spending.

In Appendix Tables B2-B3 we investigate the sources of the relative spending reductions. We repeat the difference-in-differences analysis from Table 4 separately for categories of spending defined by place of care: inpatient, outpatient, and physician office visits respectively.[18] Inpatient episodes are uncommon, with over 87% of households having zero spending in this category in any year, and differential changes between treatment and control groups are noisy and driven by outliers. Most of the relative spending reductions come from plausibly elective care: physician office visits and outpatient spending; results for these two categories are in the tables. This is consistent with the findings of Einav et al. (2013).

We also assess the extent of "behavioral hazard" (Handel and Kolstad, 2015; Brot-Goldberg et al., 2017) in our data by investigating whether enrollees cut back on high-value utilization in response to relative coverage reductions. If consumers were observed to respond to the introduction of deductibles and coinsurance rates by reducing high-value or necessary care, this would suggest that their utilization decisions were not fully informed or fully rational, and cast doubt on models that assume away such frictions. Following Brot-Goldberg et al. (2017) and those authors' ongoing work, we categorize four types of high-value care: HPV/Hepatitis vaccines; flu vaccines; preventive exams and counseling; and STI screening.[19] Brot-Goldberg et al. (2017) found that spending on similar categories of care declined significantly in their setting when enrollees were moved en masse into a high-deductible plan, despite the fact that preventive care visits remained free at the point of service and could potentially prevent substantially increased spending and worse health in the future. Appendix Table B4 shows much less clear results in our setting. The relative change in this high-value spending was a very small ($48) average relative spending reduction per year for families and an even smaller ($2) relative *increase* for individuals. We conclude that, perhaps not surprisingly given the smaller changes in OOP prices and very limited deductibles in our sample, the estimates do not suggest the presence of measurable behavioral hazard in this setting.

3.2 Risk aversion

Last, we use the data on 2016 enrollment in the new high-coverage POS-Plus plan to assess the potential importance of risk aversion in our population. Appendix Tables B5 and B6 consider the same subsample as Table 3: the non-union employees who are in the data in 2015-16; are enrolled in a POS plan in 2015; and choose POS or POS-Plus plans in 2016. We examine whether sicker or larger households—that is, households likely to have a relatively high variance of spending and face greater financial risk—are more likely to select into the higher-coverage option. We find that they are: Table B5 shows that the proportion of households choosing the POS-Plus option is monotonically increasing as we move up the DCG severity score quartiles for continuing employees.

[18] A fourth category, emergency/urgent and other utilization, is too small and heterogeneous to be informative.

[19] The definition of high-value care follows preventative care guidelines from the Affordable Care Act (2010). Thanks to Jonathan Kolstad for suggesting this analysis and providing details of high-value care categorizations. STI screening is noisy but relatively small; results are essentially unchanged when we remove it.

We also consider the set of new employees who enter the data in 2016; their choices may be more informative because they face no switching costs when the new plan is introduced. The sample sizes are smaller but with the exception of the smallest groups (generally containing fewer than 20 employees) the positive correlation between severity and POS-Plus choice remains. Table B6 conducts the same analysis looking at family size. The correlations are smaller but still positive: larger families, facing a higher variance in utilization requirements, are more likely to choose the POS-plus option. The relationship is monotonic for both continuing and new employees. Though these findings are consistent with households being risk averse, we note that both sicker families facing a high level of health care spending and households facing greater risk from a higher variance in medical spending may rationally choose the more financially generous POS-Plus plan.

4 Model

In this section, we present and detail the estimation of our stylized two-stage model of household insurance choice and health care utilization. The identification of our model relies on the variation in plan offerings and the observed changes in household plan choices and spending documented in the previous sections.

Following the discrete-continuous setup of Cardon and Hendel (2001), we model households as choosing among a set of available health insurance plans at the beginning of each year given their beliefs over health care needs (discrete); and then deciding the amount of spending during the course of the year once health needs are realized (continuous). We follow the implementation presented in Einav et al. (2013) (hereafter, EFRSC) in which households make plan and health consumption choices based on their price sensitivity for medical care (interpreted as moral hazard), underlying health risk, and risk preferences. To match our empirical setting, we also incorporate features from other key papers on health plan choice, including non-linear cost sharing (as in Cardon and Hendel, 2001) and horizontal plan differentiation and switching costs (as in Handel, 2013; Handel, Hendel and Whinston, 2015). We also introduce two new components: hassle costs of incurring positive medical spending; and preferences over horizontal plan characteristics that differ by sickness severity.

We first present the baseline model and its econometric parameterization. We then detail how the parameters governing the underlying distribution of household preferences and characteristics can be estimated in order to rationalize observed coverage and utilization decisions. Our approach to estimation differs from prior work, as we are among the first to utilize both coverage and spending choices for estimation in the presence of state dependence and moral hazard (see also Marone and Sabety, 2020); doing so requires modeling the sequence of household decisions over multiple years for both medical care utilization and insurance plan choice.

4.1 Medical Care Utilization and Insurance Plan Choice

At the beginning of each year, a household engages in the following sequential choice model. In Stage 1 the household, internalizing the needs and preferences of its members, chooses an insurance plan from a set of plans offered by its employer.[20] In Stage 2, conditional on plan choice and the realization of health care needs, the household chooses the amount of medical care to consume.

Following the notation and structure of EFRSC, a household is characterized by three objects: $(F_{\lambda,t}(\cdot), \omega, \psi)$ (where, for clarity, we omit household-specific subscripts until later). $F_{\lambda,t}(\cdot)$ represents the household's expectation over its uncertain *health needs* $\lambda \geq 0$ in period (year) t. The value of λ in a given period is realized after a household chooses its plan, and a higher value of λ corresponds to a household with greater health care needs. The second object, ω, represents the household's price sensitivity for medical care, and can be interpreted as the household's level of moral hazard (explained below). Last, ψ represents the household's coefficient of absolute risk aversion.

We present our model in reverse order, beginning with Stage 2.

Stage 2: Medical Care Utilization. At the beginning of Stage 2 for a given period t, a household is enrolled in an insurance plan j, and realizes health needs λ. The household chooses its optimal level of medical spending m for that period to maximize its ex-post realized utility for that period (measured in dollars), given by:

$$\begin{aligned} u_{j,t}(m; \lambda, \omega, j_{t-1}) = \; & \underbrace{h(\lambda, m; \omega)}_{\text{health and spending}} + \underbrace{(y_t - c_{jt}(m) - p_{jt})}_{\text{income and expenditures}} \\ & + \underbrace{\begin{bmatrix} \boldsymbol{x}_{jt} \\ \lambda \cdot \boldsymbol{x}_{jt} \end{bmatrix}' \boldsymbol{\beta}_x + \delta \times \mathbf{1}(j_t \neq j_{t-1})}_{\text{plan characteristics and switching costs}} , \end{aligned} \tag{1}$$

where $h(\cdot)$ is a function of the household's health needs (λ) and medical spending (m), y_t is the household's annual income, $c_j(m)$ are out-of-pocket (OOP) payments made by the household for its medical spending (which depend on the plan's coverage characteristics—i.e., coinsurance rates, deductibles, and OOP limits), p_j is the annual plan premium, $\boldsymbol{x}_j$ are other plan characteristics (interacted with health needs), and δ represents any switching costs incurred when a household enrolls in a different plan than the one from the previous period (where $\mathbf{1}(j_t \neq j_{t-1})$ is the indicator function for this event).[21] Plan characteristics that we condition on in our empirical analysis include: carrier-by-year fixed effects, a POS-plan indicator (which captures any benefits from more easily accessible out-of-network care), and whether the plan is offered by Harvard Pilgrim interacted with the POS-plan indicator and with whether the household is located outside of Cambridge. The

[20] In reality, most employees choose their household's insurance plan for the subsequent calendar year during an "open enrollment period" that typically takes place the previous fall. For ease of exposition, we will refer the plan choice decision as if it occurs "at the beginning" of each calendar year.

[21] Premiums and out-of-pocket payments for each plan differ depending on whether the household comprises a single or multiple individuals; for brevity, we omit this in notation but account for it in our application.

inclusion of additively separable plan characteristics and switching costs is shared with the plan-choice models used in Handel (2013) and Handel, Hendel and Whinston (2015). The interaction of household health needs λ with plan characteristics, to our knowledge, is new to this literature. It is useful in helping match the model's predictions to observed plan shares and spending data, and is an important feature that allows us to account for differential selection of households with different sickness severities in our counterfactual simulations.

Stage 1: Insurance Plan Choice. In Stage 1, each household chooses an insurance plan from a choice set of plans, denoted $\mathcal{J}_t$, to maximize its expected utility for the period. The household anticipates that its health needs are drawn from the distribution $F_{\lambda,t}$, and that its medical spending given these needs are governed by optimal Stage-2 medical spending, denoted by $m^*_{j,t}(\lambda)$. The household has constant absolute risk aversion (CARA) preferences over ex-post utilities given optimal medical spending, denoted $u^*_{j,t}(\lambda, \omega, j_{t-1}) \equiv u_{j,t}(m^*_{j,t}(\lambda); \lambda, \omega, j_{t-1})$. Hence, the expected utility that a household anticipates from enrolling in plan j at the beginning of period t is:

$$v_{j,t}(F_{\lambda,t}, \omega, \psi, j_{t-1}) = -\int exp(-\psi \times u^*_{j,t}(\lambda, \omega, j_{t-1}))dF_{\lambda,t}(\lambda) \ , \tag{2}$$

and the household's optimal choice of plan is:

$$j^* = \arg\max_{j \in \mathcal{J}_t} v_{j,t}(\cdot) \ . \tag{3}$$

Similar to much of the existing literature, our plan choice model assumes that a household anticipates expected medical spending for the period in which its choice is made, but does not condition its choice on realizations in future periods. It also rules out the possibility that a household anticipates lock-in and the impact of its current choice on future switching costs.

4.2 Econometric Parameterization

We now provide the econometric parameterization of our model.

Stage-2 Utility and Medical Spending. The function $h(\lambda, m; \omega)$ in each household's Stage-2 utility takes the following form:

$$h(\lambda, m; \omega) = (m - \lambda) - \frac{1}{2\omega\lambda}(m - \lambda)^2 - \zeta \times \mathbf{1}(m > 0) \ , \tag{4}$$

where $\zeta \geq 0$ represents a "hassle-cost" of consuming strictly positive medical care (described below).

The first two terms on the right-hand-side of (4) follow the "multiplicative moral hazard" specification from EFRSC.[22] Conditional on consuming positive medical care, a household that

[22] Also similar to EFRSC, we do not explicitly model the timing of individual claims within a given calendar year; rather, we model only total medical spending for a given year. See Einav, Finkelstein and Schrimpf (2015), Dalton, Gowrisankaran and Town (2020) for work on endogenizing the timing of medical spending.

is enrolled in plan j at period t will spend the optimal amount $m^*_{>0}$ that satisfies the following first-order condition (derived from (1)):

$$m^*_{>0} = \lambda \times (1 + \omega(1 - c'_{j,t}(m^*))) \ . \tag{5}$$

That is, conditional on consuming a positive amount of medical care, a household spends λ plus an additional amount $\lambda \times \omega(1 - c'_{j,t}(m^*))$. This additional amount represents *moral hazard*, and will be strictly positive whenever there is incomplete cost sharing ($c'_{j,t}(\cdot) < 1$). Note also that (5) implies that moral hazard is increasing in health needs λ. This is consistent with the finding discussed in the previous section: higher-spending households had greater observed reductions in medical spending when exposed to increasing levels of cost sharing.

The third term on the right-hand-side of (4), representing the hassle-cost of positive medical spending, is novel to our model, and allows the model to rationalize the mass of enrollees with zero medical spending in the observed data. If $\zeta > 0$, a household will optimally choose not to consume any health care even when $\lambda > 0$ if λ is sufficiently small. Given our specification of each household's second-period utility ((1) and (4)), each household's optimal medical spending is:

$$m^* = \begin{cases} m^*_{>0} & \text{if } h(\lambda, m^*_{>0}; \omega) \geq c_{j,t}(m^*_{>0}) + h(\lambda, 0; \omega); \\ 0 & \text{otherwise.} \end{cases} \tag{6}$$

Thus, a household spends the optimal positive amount on medical care $m^*_{>0}$ only when the gains from doing so ($h(\lambda, m^*_{>0}; \omega)$) exceed the incurred financial and hassle costs from such care, and the opportunity cost of spending nothing.

Household Heterogeneity. We index each household by i and observe its plan choice $j_{i,t}$ and level of medical spending $m_{i,t}$ in each period t. Denote the "family type" of each household by $\tilde{\tau}(i)$, which represents if the household has one (single or individual), two (two-party), or three or more (family) members; and the "household type" of each household by $\tau(i)$, which represents both the family type and union status of the household. Each household conditions its choices based on the following objects in our model: its time-varying distribution over health $F_{i,\lambda,t}$; its moral hazard parameter ω_i; its risk aversion ψ_i; and time-varying hassle-costs $\zeta_{i,t}$.

We parameterize these objects as follows. Following EFRSC, each household's health needs $\lambda_{i,t}$ in a given year is log-normally distributed so that $log(\lambda_{i,t}) \sim N(\mu_{\lambda,i,t}, \sigma^2_{\lambda,\tilde{\tau}(i)})$, where the variance of realized health needs in any given year is family-type specific. We assume that $\mu_{\lambda,i,t} = \boldsymbol{x}'_{\lambda,i,t}\boldsymbol{\beta}_\lambda + \nu_{\lambda,i}$, where $\boldsymbol{x}_{\lambda,i,t}$ denotes potentially time-varying observable household characteristics that include household type, salary tier, age tier, and DCG severity quartile; and $\nu_{\lambda,i} \sim N(0, \sigma^2_\mu)$ is an (unobserved) household-specific time-invariant mean shifter.[23] Note that the log-normal distributional assumption for $\lambda_{i,t}$, consistent with previous literature, implies that a household

[23] For salary tier, we use an indicator for each of three salary tiers used by Harvard for the primary member. For age tier, we use a 40+ and 50+ indicator for the primary member of the household. For DCQ severity quartile, we take the mean DCG score for each household, and then compute quartiles within each family type and year.

with a relatively high value of $\mu_{\lambda,i,t}$ has both higher expected (mean) health needs and a larger variance of the $\lambda_{i,t}$ distribution: that is, the household also faces greater risk.

Each household takes on one of K potential values for its moral hazard parameter ω, with each potential value $\omega_k \in (\omega_1, \ldots, \omega_K)$ occurring with probability $\exp(\beta_{\omega,k})/(\sum_{l=1,\ldots,K} \exp(\beta_{\omega,l}))$. All households share a common CARA coefficient, where $\psi = \exp(\beta_\psi)$.[24] We parameterize hassle-costs $\zeta_{i,t} = \exp(\beta_{\zeta,1,\tau(i)} + \beta_{\zeta,2,t})$ so that hassle costs are positive, and vary by household type ($\beta_{\zeta,1,\tau(i)}$) and by year ($\beta_{\zeta,2,t}$).

Cost Sharing. We assume that each plan's financial characteristics can be completely summarized by three objects: a deductible ded_{jt}, the amount below which a household is responsible for the entire cost of medical care; an OOP maximum $oopmax_{jt}$, an amount above which a household is not responsible for any cost of care; and a coinsurance rate $coins_{jt}$, representing the fraction of medical expenses between the deductible and OOP maximum that the household must pay. Hence, the OOP responsibility for each household given medical spending m is:

$$c_{j,t}(m) = \begin{cases} m & \text{if } m < ded_{jt} \ , \\ ded_{jt} + coins_{jt} \times (m - ded_{jt}) & \text{if } m \in [ded_{jt}, medmax_{jt}) \ , \\ oopmax_{jt} & \text{if } m \geq medmax_{jt} \ , \end{cases} \tag{7}$$

where $medmax_{jt} = ded_{jt} + (oopmax_{jt} - ded_{jt})/coins_{jt}$ is the amount of medical spending at which a household reaches the OOP maximum for the plan.

Switching Costs. We allow households to choose any insurance plan in their first year of employment, but because carrier switching is extremely rare in our data, we restrict them to choose only plans offered by their current insurance carrier in subsequent years.[25] For example, a household enrolled in a plan offered by HUGHP in 2014 is only allowed to choose among HUGHP's HMO and POS offerings in 2015. This implies that our switching cost parameter, represented by δ, captures the cost of switching plan types rather than carriers.

Model Parameters. The parameters to estimate are $\boldsymbol{\theta}_1 \equiv (\boldsymbol{\beta}_x, \delta)$, household preferences over plan characteristics; and $\boldsymbol{\theta}_2 \equiv (\boldsymbol{\beta}_\lambda, \boldsymbol{\sigma}_\lambda, \sigma_\mu, \boldsymbol{\omega}, \boldsymbol{\beta}_\omega, \beta_\psi, \boldsymbol{\beta}_\zeta)$, in which the first six objects govern the distribution of household health needs, moral hazard, and risk aversion, and the last term govern hassle costs of positive medical spending.[26]

[24]We also estimated a different specification that explicitly allowed for correlation between risk aversion and moral hazard, where both ω and ψ were permitted to differ only by family type. However, there was no significant or economically meaningful difference in risk aversion across the three groups, so we did not pursue this further.

[25]We restrict our sample to households that do not switch carriers; see footnote 13.

[26]We denote $\boldsymbol{\sigma}_\lambda \equiv (\sigma_{\lambda,\tilde{\tau}})_{\tilde{\tau}}$, $\boldsymbol{\omega} \equiv (\omega_k)_k$, $\boldsymbol{\beta}_\omega \equiv (\beta_{\omega,k})_k$, and $\boldsymbol{\beta}_\zeta \equiv ((\beta_{\zeta,1,\tau})_\tau, (\beta_{\zeta,2,t})_t)$.

4.3 Estimation

We estimate our model using simulated maximum likelihood. Each household contributes two components to the likelihood. The first component is the probability of observing the sequence of plan choices in each year the household is in the sample; due to the presence of switching costs and persistent household-level heterogeneity, yearly plan choices are interdependent. The second component is the density of observed medical spending each year. For each parameter evaluation, we compute each household's contribution to the likelihood by taking multiple draws for each household's unobserved values of $((F_{\lambda,i,t})_t, \omega_i)$, calculating the probabilities of observing the sequence of plan choices and medical spending given these values and parameters, and averaging over draws. The routine accounts for the discrete (plan choice) and continuous (amount of medical spending) nature of the environment (as in Dubin and McFadden, 1984), and for the selection of heterogeneous households with different health risks and price sensitivities across plans and over time (in a manner similar to Pakes, 1986).

There are two complications that we address in estimation. The first arises from non-linearities in the out-of-pocket responsibility for each household ($c_{j,t}(\cdot)$), given by (7). Since the marginal out-of-pocket cost of additional medical spending ($c'_{j,t}(\cdot)$) is discontinuous, there are levels of medical spending that are not rationalizable by any realization of health needs λ. To account for this, we allow for potential measurement error in the observed amount of total medical spending; this ensures that any observed spending value can be rationalized by the model. In estimation, the probability density of observed medical spending each year can be expressed as a function of the density of the implied measurement error given realized health needs λ and the predicted optimal level of medical spending $m^*_{j,t}(\cdot)$.

The second complication is a variant of the standard initial conditions problem (Heckman, 1981): due to switching costs, the plan choice probabilities of employees in 2014 who were employed in prior years depends on their prior enrollment decisions, which are not observed. However, we do observe how many years the primary member has been employed at Harvard prior to 2014.[27] We use this information to approximate the distribution of 2013 plan choices for each household in the following procedure. We assume that each household in their first year at Harvard (year $y_0 = 2014 - T$, where T denotes the number of years prior to 2014 the primary member was employed by Harvard) faced the same set of health plans, and had the same household characteristics and preferences as in 2014. We then predict plan choice probabilities in year y_0 for the household assuming zero switching costs, and predict choice probabilities for each subsequent year $y = y_0, \ldots, 2013$ given simulated choices in the prior year to obtain a probability distribution over the household's 2013 plan choices. While clearly imperfect, this approach allows us to use the employee's tenure at Harvard, in addition to the observed household characteristics, to help predict the household's initial state in 2014. For employees who begin employment at Harvard during our sample period (i.e., those who enter in 2014 or later), we assume switching costs are zero in their first year at Harvard.

[27] In 2014, 41.2% of enrollees had been at Harvard for five years or fewer; 61.6% had been employed for under ten years.

We provided further estimation details in the Appendix.

4.4 Identification

With an "ideal" dataset comprising (i) a menu of plan options that is continuous in both coinsurance rates and premiums faced by employees, (ii) an exogenous change in plan coverage levels, and (iii) a sufficient number of periods before and after this exogenous change for which plan choice and medical utilization is observed, EFRSC discuss how the joint distribution of $(F_\lambda(\cdot), \omega, \psi)$ for each household (when stable over time) is non-parametrically identified in this category of models.

Our dataset does not satisfy these conditions—in particular, we observe at most three years for each household, and households have at most four plan offerings to choose from—so we rely on additional parametric assumptions to identify our model. Here, we provide an informal description of how our model parameters are informed by the data. For the following discussion, we condition on each household's observable characteristics, including its household type.

First, note that while the variation in our data is somewhat limited, it has the advantage of being clean and intuitive. We observe first a change in plan coverage levels for a plausibly exogenous defined subsection of the population (non-union enrollees), followed by the introduction of a new high-coverage option that likely appealed to particular types of consumers. The intuition therefore proceeds in several steps. In 2014, both union and non-union households faced the same menu of plans with the same financial coverage. The observed 2014 distribution of health spending can be used to recover parameters governing that distribution for a given level of cost sharing (e.g., $\boldsymbol{\beta}_\lambda, \boldsymbol{\sigma}_\lambda$); this includes also hassle cost parameters ($\boldsymbol{\beta}_\zeta$), which enable the model to rationalize the proportion of enrollees with zero spending in the data.

Next, observed spending changes over time within-plan are useful for two reasons. Recall that coverage for the union population did not change during our time period. Under the assumption of common yearly shifts in the distribution of non-discretionary spending $F_{\lambda,t}(\cdot)$ for union and non-union households, we can use observed spending changes for the union population to estimate these shifts. The average level of moral hazard ω comes from the change in the distribution of medical spending from 2014-2015 for non-union employees relative to union populations, as non-union coverage changed. This follows directly from the difference-in-difference arguments laid out in Section 3.

Unobserved heterogeneity across households in moral hazard (governed by parameters $\boldsymbol{\beta}_\omega$, $\boldsymbol{\omega}$) and average health needs (σ_μ) are informed by both plan choice and health spending data, together with the exogenous plan menu change in 2016 when the lowest cost-sharing POS-Plus plans were introduced alongside the HMO and POS products. Unobserved heterogeneity in moral hazard is informed by differences in spending adjustments for households choosing more versus less financially generous plans (i.e., selection on moral hazard as in EFRSC). Unobserved heterogeneity in health needs assists our model in rationalizing plan choice probabilities across HMO and POS plans and between HUGHP and Harvard Pilgrim plans (as we allow health needs λ to interact with these plan characteristics in (1)), as well as any differences in the distribution of realized health spending across these plan types. Finally, the observed plan choices provide information on the remaining model parameters. The household risk preference ψ allows our model to rationalize within-year plan choice probabilities across the POS and POS-Plus plans in 2016 that differ only on financial terms. Identification of parameters in $\boldsymbol{\theta_1} \equiv (\boldsymbol{\beta}_x, \delta)$ that govern preferences over horizontal plan characteristics then follows standard arguments: preferences $\boldsymbol{\beta}_x$ are informed by plan choice probabilities, and any persistence in choice probabilities for households who were previously employed compared to those for new households (controlling for the distribution of household heterogeneity across plans, as parameterized by our model) informs the level of switching costs δ.

4.5 Estimates

In Table 6, we report quantities of interest that are derived from our parameter estimates. We report the underlying parameter estimates, which are less intuitive to interpret and discuss, in Appendix Table B7.

The top panel of Table 6 provides estimates for households' valuations over plan characteristics (measured in thousands of dollars per year). Households with zero health needs value POS-type plans approximately \$1,400 less than HMO-type plans, but this value is increasing for those households with higher realized health (\$180 per \$1,000 increase in health needs). Consistent with correlations between household characteristics and plan carrier choice, our model predicts that Harvard Pilgrim is valued more by households with greater health needs, and by households who live outside of Cambridge. We also estimate substantial switching costs: households incur a \$3,800 cost to switch plan types, which is similar in magnitude to the annual premiums for a family household in our sample.

The next panel provides estimates for the distribution of health needs, governed by $F_\lambda(\cdot)$. We estimate that expected health needs are higher for larger and union households, and are increasing in salary tier and DCG severity scores. DCG scores have the greatest impact of any covariate on expected health needs, with households in the third and fourth quartiles of the sample predicted to have 3.4 and 6.2 times greater expected needs compared to households in the first quartile. Across households, the average expected health need ($E[\lambda]$) is approximately \$6,500 in 2014. Consistent with EFRSC, we estimate that there is substantial unobserved heterogeneity in expected health needs, with a standard deviation of $E[\lambda]$ across households of approximately \$4,900 (coefficient of variation of 0.75); we also find that this is less than the standard deviation in realized health needs, which averages \$9,780 across households (not shown in Table 6).

To rationalize the observed number of households with zero medical spending in a year, our model estimates that households face significant hassle costs for non-zero spending, and that this cost is increasing in the size of the household. Though zero observed medical spending is more likely for single-coverage households (approximately 15% in 2014) than non-single households (approximately 1%), the expected health needs of non-single households are higher; hence, this hassle costs are predicted to be higher as well for the larger households.

We estimate moral hazard to be significant for a substantial proportion of the population. For

Table 6: Implied Quantities from Model

			Estimate	95% CI
Plan Choice	Value at \$0 health needs	POS(P)	-1.39	[-1.48, -1.20]
		HP	-0.44	[-0.53, -0.35]
		HP x POS(P)	0.83	[0.67, 0.95]
		HP x Cambridge	-0.89	[-1.02, -0.70]
	Increase in value per \$1000 increase in health needs	POS(P)	0.18	[0.15, 0.20]
		HP	0.25	[0.21, 0.27]
		HP x POS(P)	-0.13	[-0.15, -0.10]
		HP x Cambridge	-0.06	[-0.09, -0.04]
	Switching Cost (\$000)		3.79	[3.47, 4.04]
Health Needs ($F_{\lambda,t}(\cdot)$)	Mean (\$000)	Single x Union	0.29	[0.27, 0.32]
		2-Party x Union	1.31	[1.13, 1.41]
		Family x Union	1.82	[1.76, 2.11]
		Single x Non-Union	0.22	[0.21, 0.24]
		2-Party x Non-Union	0.96	[0.87, 1.06]
		Family x Non-Union	1.59	[1.43, 1.70]
	Mean Multiplier	Tier 2	1.33	[1.24, 1.36]
		Tier 3	1.50	[1.40, 1.59]
		Age 40+	1.14	[1.12, 1.24]
		Age 50+	1.03	[0.96, 1.08]
		DCG Q2	2.07	[2.00, 2.24]
		DCG Q3	3.41	[3.26, 3.68]
		DCG Q4	6.24	[5.80, 6.71]
		Single x 2015	1.02	[0.98, 1.08]
		2-Party x 2015	1.03	[0.92, 1.09]
		Family x 2015	1.04	[0.97, 1.06]
		Single x 2016	1.12	[1.06, 1.16]
		2-Party x 2016	1.08	[1.02, 1.21]
		Family x 2016	1.07	[0.98, 1.11]
	Implied $E[\lambda]$ (\$000) (all households)	Average	6.49	[6.26, 7.22]
		SD	4.89	[4.42, 5.50]
Hassle Costs (ζ)	Mean (\$000)	Single x Union	0.61	[0.56, 1.14]
		2-Party x Union	0.85	[0.69, 1.55]
		Family x Union	1.71	[1.37, 3.10]
		Single x Non-Union	0.55	[0.51, 1.05]
		2-Party x Non-Union	0.61	[0.46, 1.35]
		Family x Non-Union	0.84	[0.71, 1.83]
Moral Hazard (ω)	Percent increase in spending from 100% to 0% coinsurance	Low	3.51	[1.86, 3.82]
		High	26.3	[12.4, 30.0]
	Probability of type	Low	0.36	[0.33, 0.40]
		High	0.64	[0.60, 0.67]
Risk Aversion (ψ)	CARA Coefficient (10^{-6})		3.07	[2.90, 3.40]

Notes: Implied quantities of interest implied by parameter estimates (reported in Appendix Table B7). 95% confidence intervals obtained from re-estimating the model over 200 bootstrap samples of households.

nearly two-thirds of households, we predict that moving from a 100% coinsurance (with the same out-of-pocket maximum as observed in the data) to full insurance (no liability) would increase medical spending by 26%; for the other third of households, the increase would be only 3%. Our estimated magnitudes are in line with those obtained by EFRSC and Marone and Sabety (2020).[28]

[28]Our estimates are derived from a model based on EFRSC's "multiplicative" moral hazard specification, where additional spending due to moral hazard increases in realized health needs (λ); EFRSC's main specification and

Last, our estimated CARA coefficient is 3.1×10^{-6}. Following Cohen and Einav (2007), we note that this implies a household would be indifferent between receiving nothing, and participating in a 50–50 lottery in which it earns \$100 or loses \$99.97; or participating in a 50–50 lottery in which it earns \$10,000 or loses \$9,700. This is closer to risk-neutral behavior than estimated in other empirical studies of insurance choice (e.g., Handel (2013), Handel, Hendel and Whinston (2015)). There are at least two possible reasons for these differences. First, as in EFRSC and discussed therein, realized utility in this model given by (1) is a function of both health and financial risk, rather than just financial risk as in much of the previous empirical insurance literature. Risk aversion estimates obtained here hence may not be directly comparable to those obtained in papers where risk is solely financial.[29] Second, our population faces limited exposure to financial risk due to generous cost sharing and low OOP maximums. Even the reduced-coverage HMO and POS plans of 2015 had realized actuarial values of around 70% (Table 2), the same as the mandated level for silver plans on the federal health insurance exchanges. Out-of-pocket maximums are at most \$6,000 per household in our data, compared to a maximum of \$13,200 per household in federal exchange plans in 2015.[30] Nonetheless, although our estimates may be appropriate for our setting, in Section 5.6 we repeat our main simulations using risk aversion estimates closer to those obtained in prior work as a robustness exercise.

4.6 Model Fit

Table 7 presents the observed and predicted plan choice probabilities across carriers (HUGHP, HP) and plan types (HMO, POS) by family type in 2014. Recall that our utility specification given by (1) only allows household preferences for carrier and plan type to differ through the expected health needs (λ), and not by family type. Even so, our model fits predicted choice probabilites well. For example, it captures the larger share of family households choosing HP plans over HUGHP relative to individual households, and the larger share of HP enrollees choosing the POS plan compared to HUGHP enrollees. Figure 4 presents the distribution of observed and predicted medical spending. Generally, the fit is reasonable. For individuals, we slightly under-predict the share of households that have zero medical spending (13% as opposed to 17%); for couple and family households, we

Marone and Sabety (2020) employ an "additive" specification in which additional spending is invariant to λ. EFRSC "estimate an average moral hazard parameter (ω) that is about 30 percent of the average health risk" (p. 204) in their sample (where this corresponds to the increase in spending when moving from no insurance to full insurance); Marone and Sabety (2020) estimate this quantity to be 24% for individuals and 14% for families.

[29] In our model, uncertainty over health needs λ affects household's utility via: (i) the function $h(\cdot)$; (ii) the term $\lambda \cdot \boldsymbol{x}_{jt}$; and (iii) out-of-pocket spending $c_{jt}(\cdot)$. As noted by EFRSC (p204): "...one could add a separable health-related component to utility that is affected only by λ to change the risk aversion estimates, without altering anything else in the model." Indeed, one reason why our estimated level of risk aversion differs from EFRSC is that we include such a "separable health-related component to utility" (the second-to-last term in (1)).

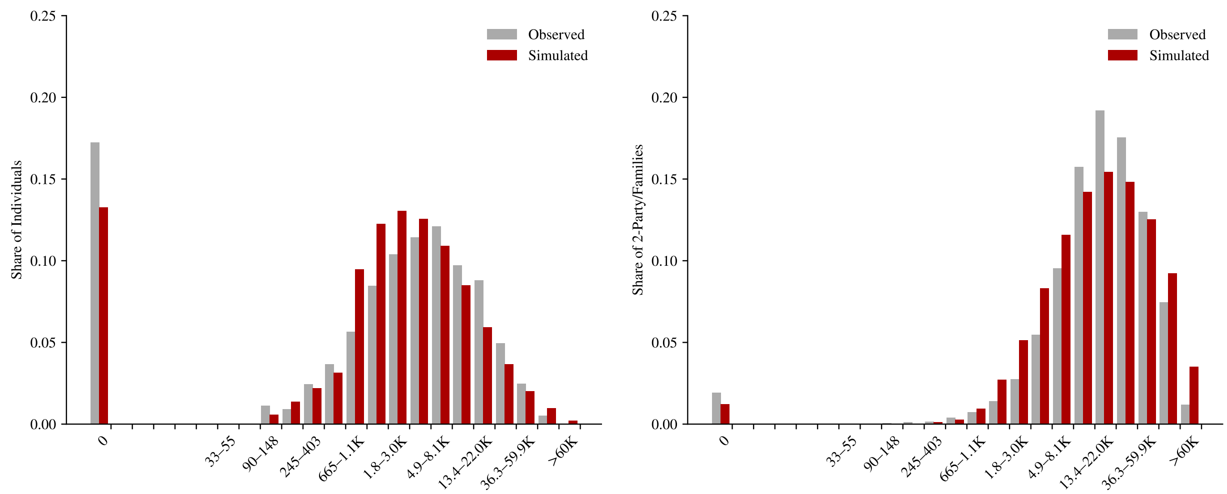
[30] Additionally, our risk aversion parameter is informed by observed differences in 2016 plan choice probabilities for POS and POS-Plus plans (as these are the sole pairs of plans that differ only on financial coverage levels in any given year); insofar as there are other idiosyncratic or unobserved features that affect plan choice, this will affect our estimates.

Table 7: Observed and Simulated Plan Shares (2014)

		Individual		Two-Party		Family (3+)	
Carrier	Plan Type	Observed	Simulated	Observed	Simulated	Observed	Simulated
HUGHP	HMO	0.46	0.45	0.30	0.33	0.23	0.27
	POS	0.09	0.10	0.07	0.07	0.05	0.06
HP	HMO	0.33	0.33	0.45	0.46	0.53	0.48
	POS	0.11	0.13	0.18	0.15	0.19	0.19

Notes: Observed and predicted plan market shares by family type in 2014.

Figure 4: Observed and Simulated Distribution of Spending (2014)



(a) Individual Households

(b) Family Households

Notes: Observed (light) and simulated (dark) medical spending in dollars for households in 2016. These figures use a log scale: each bin $k = 1, \ldots, 22$ corresponds to spending in the range $exp(0.5 \times (k-1)) - exp(0.5 \times k)$, with all spending above $exp(0.5 \times 22) \approx 60K$ contained in last bin. The labels on the x-axis show the corresponding dollar amounts for each bin.

slightly over-predict dispersion in the amount of positive spending.[31]

5 Optimal Plan Menu Design

We now use our estimated model to explore the trade-offs faced by a large self-insured employer designing its insurance plan menu.

We begin with the simplest case in Section 5.2, examining what level of coverage maximizes average employee surplus when a single plan is offered. This initial exploration builds on the theoretical literature that dates back to Arrow (1963) and Zeckhauser (1970), and empirical papers such as Buchanan et al. (1991) and Manning and Marquis (1996). We ask essentially the same question: how should a firm choose a single plan's coverage level to manage the trade-off between

[31] Observed/predicted means and medians are \$2.7K/\$2.7K and \$1.1K/\$1.5K for individuals and \$10.8K/\$13.3K and \$6.7K/\$8.0K for non-single households.

risk coverage and moral hazard for a particular population?

We next turn to measuring the additional gains achievable when an employer can offer multiple plans, focusing first on plans that differ only in their financial characteristics—in particular, the amount and form of cost sharing. Our analysis here follows two steps. First, in Section 5.3, we explore what is achievable under what we refer to as *assignment*, whereby the employer can choose the plan each household enrolls in. To obtain an upper-bound on gains from this setting, we consider the extreme possibility that each household can be assigned to a "tailored" plan with a household-specific deductible and coinsurance rate. We also examine what is feasible if a firm can assign each household to one of only two particular plans.

While the results obtained under assignment are unlikely to correspond to any feasible, implementable policy—due in part to non-discrimination requirements contained in US employment regulations—they nonetheless provide a useful benchmark for comparison to our next set of analyses: the design of an insurance plan menu with *selection*. These analyses, described in Section 5.4, allow households to choose which plan to enroll in, as they do in reality. We show that, analogous to a move from first- to second-degree price discrimination, the achievable gains from providing multiple options are reduced when households can freely choose among plans. Finally we consider the potential for an employer to adopt other strategies to help manage selection. We explore the effectiveness of two such strategies—offering different coverage levels based on household size, or offering plans that are also differentiated on non-financial dimensions—in Section 5.5.

5.1 Simulation Details

The results in this section are obtained by simulating our model of household insurance choice and health care utilization using the estimated parameters presented in the previous section. These parameters govern the heterogeneous preferences and health needs of our population. We use the characteristics of households and parameter estimates from 2016, the final year of our sample, and simulate market outcomes under different plan offerings for a single year. We do not incorporate any dynamic considerations including switching costs (which are set to zero in our simulations).

5.1.1 Average Employee Surplus

The primary objective that we maximize when designing plan menus is *average employee surplus*. We define this object as follows. Each household i is characterized by a vector of parameters $\boldsymbol{\gamma}_i \equiv (F_{\lambda,i}(.), \omega_i, \psi_i)$ which govern its distribution of health needs and levels of moral hazard and risk aversion. Household i's certainty equivalent of choosing plan j, denoted $e_{i,j}$, satisfies $-exp(-\psi_i \times e_{i,j}) = v_j(\boldsymbol{\gamma}_i)$, where $v_j(\cdot)$ is defined in (2). This implies that:

$$e_{i,j}(\boldsymbol{\gamma}_i) = -\frac{1}{\psi_i} ln\Big(\int exp(-\psi_i \times u^*_{i,j}(\lambda, \omega_i)) dF_{\lambda,i}(\lambda) \Big) \; ,$$

where we omit the period t and each household's last period plan choice from notation. Given our CARA utility specification with additively separable income and health, there are no income effects, and we can re-write the previous expression as:

$$e_{i,j}(\boldsymbol{\gamma}_i) = -\frac{1}{\psi_i} ln\Big(\int exp(-\psi_i \times \tilde{u}_j^*(\lambda, \omega_i)) dF_{\lambda,i}(\lambda)\Big) + (y_i - p_j) \ , \tag{8}$$

where

$$\tilde{u}_j^*(\lambda, \omega_i) = h(\lambda, m_j^*(\lambda); \omega) - c_{j,t}(m_j^*(\lambda)) + \begin{bmatrix} \boldsymbol{x}_{jt} \\ \lambda \cdot \boldsymbol{x}_{jt} \end{bmatrix}' \boldsymbol{\beta}_x \ ,$$

with premiums for each plan (denoted p_j) determined in a manner described below. We define the average employee surplus, denoted $\overline{ES}$, to be the average certainty equivalent for the plan that maximizes each household's expected utility:

$$\overline{ES} = \frac{1}{N}\Big(\sum_i \sum_j e_{i,j,t}(\boldsymbol{\gamma}_i) \times \mathbf{1}\{j = j_i^*\}\Big) \ , \tag{9}$$

where N is the number of households, $\mathbf{1}\{\cdot\}$ is the indicator function, and j_i^* represents household i's optimal plan choice (given by (3)).

Remarks. Equation (8) highlights a convenient implication of our utility specification: any employer contribution towards employee premiums that takes the form of a fixed dollar amount is a transfer that affects the level of average employee surplus, but not the features of plans that maximizes average employee surplus. This follows since a fixed premium contribution does not affect a household's choice of plan, medical spending, or the amount of risk that is faced. Hence, we can abstract away from the firm's choice of premium contribution, a non-trivial decision which is likely to depend on employees' perceptions regarding substitutability of health benefits with wages, local labor market conditions, and other related factors that are outside the scope of our data and model. In our setup, the design of an insurance plan menu and the choice of how much to contribute to insurance premiums are separable problems for the employer.

Designing a plan menu to maximize average consumer surplus will also have distributional consequences across households. We do not explicitly address such issues here, instead noting only that in many cases it can be straightforward to redistribute surplus through variable premium contributions. For example, if single-coverage households face a loss in consumer surplus relative to families under the design that maximizes average surplus, the employer can reallocate the gains by making a larger fixed-dollar contribution for single-coverage plans. Differential premium contributions that condition on observables such as family size or salary tier are in fact used frequently by employers, both in our setting and by other large firms.

5.1.2 Plan Characteristics

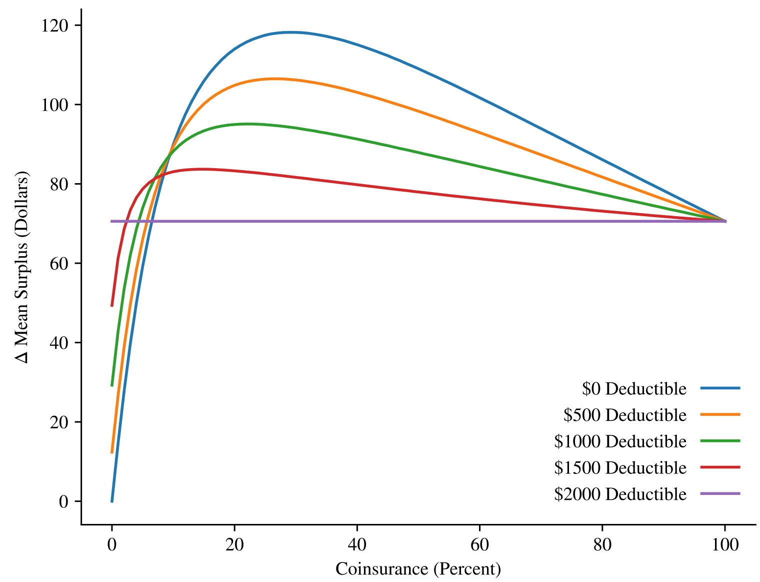
In our simulations in Sections 5.2-5.4, we assume that only HUGHP-HMO plans are offered. In our last set of simulations in Section 5.5, we allow for horizontal differentiation in the form of both HUGHP-HMO and HP-POS plans. Our simulations alter the financial generosity of plans only through coinsurance and deductibles; we hold the out-of-pocket maximums fixed at the levels offered by Harvard in 2014 (\$2000 per person per year, with a maximum of \$6000 per family per year). We do not adjust any non-financial attributes of plans.

Premium Setting. We require that insurance premiums cover employees' total expected medical spending, net of out-of-pocket (OOP) payments that they make in the form of deductibles or coinsurance. When there is only a single plan, or when households are assigned to particular plans, the details of how premiums are determined do not affect the value of average employee surplus given by (9), as long as they cover total medical spending net of OOP payments. This follows, again, since both medical spending and risk in each plan are unaffected by the level of premiums.

When households are able to freely choose among multiple plans, average employee surplus will no longer be invariant to premium levels. Premium differences will affect households' plan choices, and if plans vary in their amount of coverage and cost sharing, total medical spending and risk will adjust. In addition, as is well known, if each plan's premium is chosen so that it covers only its own enrollees' average costs, adverse selection can lead to a "death spiral." That is, a higher-coverage plan that attracts the least healthy may require high premiums to cover expected costs, which in turn—if the average medical spending of households on the plan continues to exceed the marginal households' willingness to pay—leads to higher and higher premiums until it no longer attracts any enrollees. Indeed, such a death spiral was documented by Cutler and Reber (1998) among the set of insurance plans offered by Harvard University in the 1990s. Following the reduction of the premium subsidy for the most generous plan option, that plan's out-of-pocket costs including premiums more than tripled, its enrollment fell (disproportionately among those who were younger and spent less on medical care), and—within three years of the change—the plan was eliminated completely from Harvard's plan menu.

To address such adverse selection concerns, we provide the employer with an additional tool to promote plan stability. Rather than require that the employer set premiums separately for each plan so that each option's premiums cover its own expected costs (i.e., expected medical spending net of OOP payments for its enrollees), we allow the employer to cross-subsidize plans, flexibly choosing the extent to which underlying differences in plans' expected costs are reflected in premiums. Specifically, when two plans are offered by the employer and employees can choose between them, we allow the firm to choose a *subsidy level* denoted $\kappa \in [0, 1]$, where κ represents the fraction of the two plans' (expected) average cost difference per enrollee that is reflected in the premium difference between plans. For example, assume that the average cost difference per enrollee between two different plans is \$X. A value of $\kappa = 1$ would force premiums on each plan to cover their own average costs, and hence the premium difference between the two plans would be \$X, while a value of $\kappa = .5$ would restrict the premium difference to be only \$X/2. In both cases, premiums collected in aggregate would still be required to cover total medical spending. This premium setting process captures the spirit of those used by Harvard both during the period of our study and also prior to the 1995 change studied in Cutler and Reber (1998).

Figure 5: Simulation A.II (Change in Average Employee Surplus Relative to Full Insurance)



Notes: Each line corresponds to the change in estimated annual household-average level of employee surplus in dollars from offering a single HUGHP-HMO plan with a fixed deductible and positive coinsurance rate (horizontal axis), relative to a single HUGHP-HMO plan with a zero deductible and zero coinsurance rate (Simulation A.I). Deductibles listed correspond to individual amounts, and out-of-pocket maximums are fixed at $2000 per individual and $6000 per household.

5.2 The Optimal Single Plan

We begin our analysis by determining the deductible and coinsurance rate that maximizes average employee surplus when the employer can only offer a single plan. We allow the level of coinsurance to vary from 0–100%, and adjust the individual deductible in $250 increments between $0 and $2000, with the upper limit corresponding to the individual out-of-pocket maximum in our setting. Consistent with our empirical setting, two-party and family deductibles and OOP maximums are twice and three times the individual maximum. We emphasize that given these OOP maximums, an insurance plan with 100% coinsurance still significantly limits an enrollee's total amount of risk exposure.

We find that the *optimal* (i.e., average employee surplus maximizing) single plan has a zero deductible and a coinsurance rate of 29%: it generates a $118 higher average surplus per household than is achieved with full insurance, and reduces average medical spending per household by $138. The top panel of Table 8 under Simulations A.I–A.II summarizes these results.

In Figure 5, we plot the change in average employee surplus from different financial coverage levels relative to a single plan with a zero-dollar deductible and no coinsurance (i.e., full insurance). The top line of the figure plots the change in average surplus generated by varying the coinsurance rate with the deductible held fixed at the optimal zero level; this line corresponds to Figure 1 presented in the introduction. There are several features of this curve to emphasize. First, we find steep gains from increasing coinsurance from zero percent, consistent with the high levels of moral hazard and low levels of risk aversion estimated in our population. Second, as the coinsurance rate increases beyond the optimum level of 29%, average employee surplus falls, although less rapidly. Even with 100% coinsurance, average surplus is substantially higher than with full insurance; this is in part due to out-of-pocket maximums limiting overall risk exposure. Third, in a wide neighborhood around the optimal coinsurance rate, average surplus gains are still substantial—for example, gains remain over $100 per household for coinsurance rates between 20-60%. That is, the curve is relatively "flat" around the optimum coinsurance rate. Combined, these features suggest that the benefits from imposing moderate levels of cost sharing are fairly robust and not sensitive to getting the level "exactly right."

The other lines in Figure 5 depict how average surplus varies with coinsurance for different deductible levels. As noted, the zero-deductible plan is optimal in our setting. Further, the average consumer surplus gain from the optimal coinsurance rate, relative to full insurance, decreases monotonically as the deductible increases. The reason for this is as follows. While increasing the deductible does reduce the fraction of households that have positive medical spending, the corresponding optimal coinsurance rate also falls, implying that enrollees who spend past the deductible face a lower out-of-pocket price for care and consequently spend more. Overall, average spending declines very little—and in some cases even increases—so that losses from consumers bearing more risk as a result of higher deductibles are not offset by premium reductions.[32]

5.3 Multiple Plans: Household Assignment

We next turn to the potential benefits of offering multiple plans that vary in their level of cost sharing. We isolate potential gains from differentiation by assuming initially that the employer can assign households to particular plans.

Household-Specific Plans. To obtain a plausible upper-bound on the gains achievable by offering plans differentiated solely along financial dimensions, we initially examine the scenario where an employer can offer each household a "tailored" plan with a household-specific deductible and coinsurance rate. This is not feasible in reality due to both the complexity of the problem and the non-discrimination requirements contained in US employment regulations. Still, it serves as a benchmark for the potential gains that may be achievable by providing multiple plans to a heterogeneous population.

Maximizing average employee surplus makes the exercise straightforward. For each household, we determine the deductible level and coinsurance rate that maximizes its certainty equivalent (given by (8)) under the restriction that the household's premium is equal to its own expected medical spending net of OOP payments. This formulation ensures that each household fully internalizes the costs of its medical spending, and in turn, implies that the set of household-specific

[32] Table B8 in the Appendix provides the optimal coinsurance rate and change in average surplus for the deductible levels shown in Figure 5. The $2000 deductible line is flat, as in this case the deductible coincides with the individual OOP maximum, and the coinsurance rate has no effect on household liability.

Table 8: Simulated Results (HUGHP HMO Plans Only)

(HUGHP HMO Plans Only)		Coins. (%)	Δ Spending ($)	Δ Surplus ($)	ΔSurplus/(a)
Single Plan (Section 5.2)					
A.I Fixed Coins.		0	0	0	-
		-	-	-	-
A.II Optimal Coins.		29	-138.17	118.20$^{(a)}$	1.00
		[28, 31]	[-149.07, -71.32]	[60.38, 127.67]	-
Multiple Plans with Assignment (Section 5.3)					
A.III Tailored Plans	Mean	48	-191.61	150.84	1.28
		[46, 49]	[-206.64, -97.63]	[76.49, 161.91]	[1.25, 1.28]
A.IV Two Plans	Plan A	15	-168.44	137.08	1.16
		[14, 16]	[-182.42, -86.34]	[69.79, 147.95]	[1.15, 1.17]
	Plan B	51			
		[49, 53]			
Multiple Plans with Selection (Section 5.4)					
A.V Two Plans	Plan A	20	-138.75	119.68	1.01
		[15, 25]	[-151.41, -72.15]	[61.29, 129.14]	[1.01, 1.02]
	Plan B	35			
		[30, 35]			
By Family-Type (Section 5.5)					
B.I Two Plans	Single	47	-143.48	120.51	1.02
		[44, 49]	[-155.34, -73.44]	[61.49, 130.60]	[1.01, 1.02]
	Non-Single	26			
		[25, 27]			

Notes: The table reports the coinsurance rate that maximize average employee surplus, and changes in average total medical spending (Δ Spending) and average employee surplus (Δ Surplus) across households when only HUGHP-HMO plans are offered; changes are relative to a single HUGHP-HMO plan with zero coinsurance rates and zero deductibles (A.I). The last column represents the fraction of the change in the average employee surplus for a single HUGHP HMO Plan with the optimal coinsurance and deductible (A.II relative to A.I) obtained by the Simulation. Simulation A.III allows the coinsurance rate to vary by household. Simulations A.III and A.IV assign households to a particular plan; simulation A.V allows households to select which plan to enroll in, and restricts coinsurance rates to 5% increments. 95% confidence intervals are reported below results in brackets, and are obtained by re-estimating the model over 200 bootstrap samples of households and re-computing simulations.

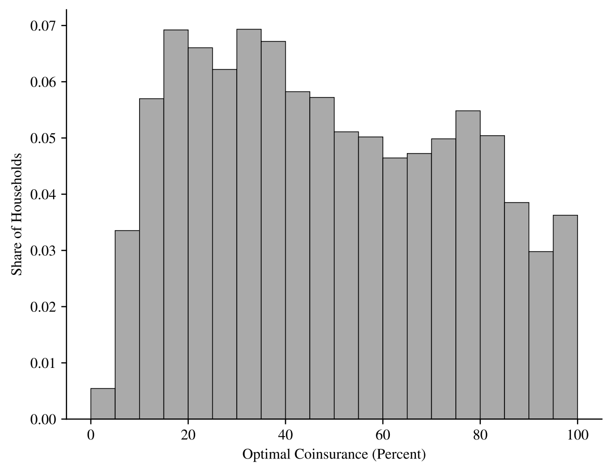
Table 9: Simulated Results (Household-Specific Tailoring of HUGHP-HMO Plans)

(HUGHP HMO Plans Only)		Deductible ($)	Coins. (%)	Δ Spending ($)	Δ Surplus ($)
Tailored Plans with Assignment					
A.III Zero Deductibles	Mean	-	48	-191.61	150.84
	Median	-	-	-204.03	86.53
	SD	-	26	81.20	725.66
A.III′ Optimal Deductibles	Mean	116.55	45	-192.68	151.04
	Median	0.00	-	-205.11	86.63
	SD	-	23	80.03	720.26

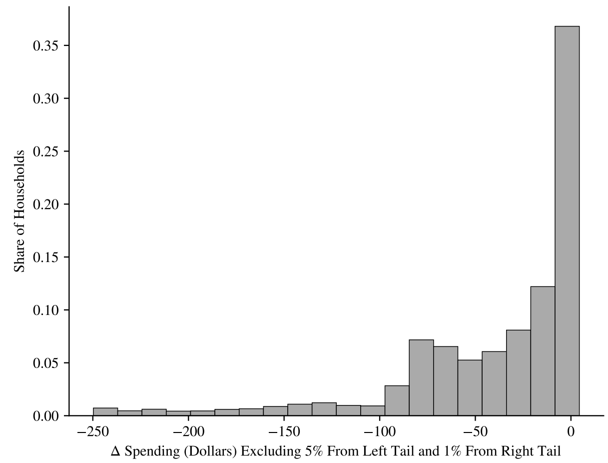
Notes: Predicted mean, median, and standard deviations of deductibles, coinsurance rates, and changes in spending and employee surplus across households from offering HUGHP-HMO plans with household-specific coinsurance rates and, for A.III' only, deductibles. Changes are relative to a single HUGHP-HMO plan with a zero deductible and zero coinsurance rate. ("Zero Deductibles" specification corresponds to Simulation A.III in Table 8.)

plans that are obtained from this household-by-household optimization maximizes average employee surplus.

Figure 6: Simulation A.III (Household-Specific Plans)



(a) Optimal Coinsurance Rates (%)



(b) Change in Spending ($)

Notes: Figure (a) shows the distribution of optimal coinsurance rates across households when each household is assigned to an HUGHP-HMO plan with zero deductible and a household-specific coinsurance rate. Figure (b) shows the corresponding distribution of changes in total household medical spending relative to a single plan with a zero deductible and zero coinsurance.

Results are presented in Table 9. When both deductibles and coinsurance rates are permitted to vary by household, we find a mean optimal coinsurance rate of 45% with a mean optimal deductible of $117 (Simulation A.III′). The average employee surplus gain compared to full insurance is $151, or $33 more than what is achievable by offering only a single plan (Simulation A.II in Table 8). The benefits from varying deductibles are limited: optimizing only over coinsurance rates with deductibles fixed at zero lowers achievable average surplus gains by only $0.20 (Simulation A.III in both Tables 8 and 9). With fixed zero deductibles, the mean optimal coinsurance rate is slightly higher, at 48%.

The final column in Table 8, which will be useful for comparing across simulations, displays the average gains from each simulation as a fraction of the gains from reaching the optimal coinsurance and deductible with a single plan. For example, the average gain from tailoring coinsurance at a zero deductible—which is $150.84—is 28% (or $32.64) more than the gains from optimizing the single plan.

Figure 6 plots the substantial cross-household dispersion in the optimal coinsurance rate, and resulting spending reductions per household (relative to full insurance), when deductibles are fixed at zero and coinsurance rates can be tailored to each household. As reported in Table 9, the standard deviation in optimal coinsurance rates is 26%, and the mean spending reduction relative to a single plan with full insurance is $192 with a large standard deviation of $81. The variation in optimal rates, and the gains from tailoring plans, speak to the significant heterogeneity in health needs and preferences of households in our population. Nevertheless, as in Figure 5 which depicts the population gains from adjusting a single plan's coinsurance rate, the gains from adjusting the coinsurance in a neighborhood around its optimum are fairly flat for any given household. As a result, moving from full insurance to a single plan with moderate cost sharing (Simulation A.I to A.II in Table 8) yields much larger gains than are achieved by further tailoring plans to the individual household level (Simulation A.II to A.III in Table 8).

Two Plans. We also consider the achievable gains when the employer can only offer two plans, but still assigns households to a plan. We determine the optimal pair of plans to offer in a manner similar to the previous simulation exercise: we search over all potential combinations of coinsurance rates and deductibles (in $250 increments) for two plans, and for each combination, we allocate each household to the plan that maximizes its certainty equivalent when the household's premium equals its expected spending net of OOP payments.

The results from this simulation are presented in Table 8 under Simulation A.IV. We find that offering two plans with assignment can achieve an additional $137 in average employee surplus above the full-insurance baseline, or 16% more than the single plan optimum. This achieves 58% of the additional benefits from offering (thousands of) tailored, household-specific plans. The optimal two plans with assignment have positive and significant coinsurance rates of 15% and 51%, and both have zero deductibles.

Since we have shown that allowing for positive deductibles produces at best very small benefits with household-level tailoring and no benefits in other cases, we restrict attention to zero-deductible plans for our remaining analyses.

5.4 Multiple Plans: Endogenous Plan Choice

Having quantified the benefits of imposing moderate levels of cost sharing with a single plan, and computing the potential gains achievable when households can be assigned to customized plans, we now consider the more realistic environment where households can freely select among multiple plans that are offered by the employer.

We examine the benefits that a firm can obtain by offering two plans that vary only in their premiums and coinsurance; both plans are restricted to have zero deductibles and OOP maximums given by Harvard's 2014 levels. As noted above, we also allow the employer to flexibly cross-subsidize the plans' premiums by choosing a subsidy level $\kappa \in [0,1]$ that determines the extent to which the average cost difference between the two plans is reflected in premiums. We conduct our simulations by computing, for every subsidy level and pair of coinsurance rates in 5% increments, plan premiums and the corresponding plan choices and spending decisions for all households such that: (i) households make optimal plan choices and spending decisions, given premiums and coinsurance rates; and (ii) premiums paid by all households cover total medical spending net of OOP payments, and premium differences between plans reflect κ times the average net medical spending difference between plans. We find the pair of coinsurance rates and subsidy level that maximizes average employee surplus.[33]

[33] Our approach is equivalent to optimizing both plans' coinsurance rates under the constraints that total premiums

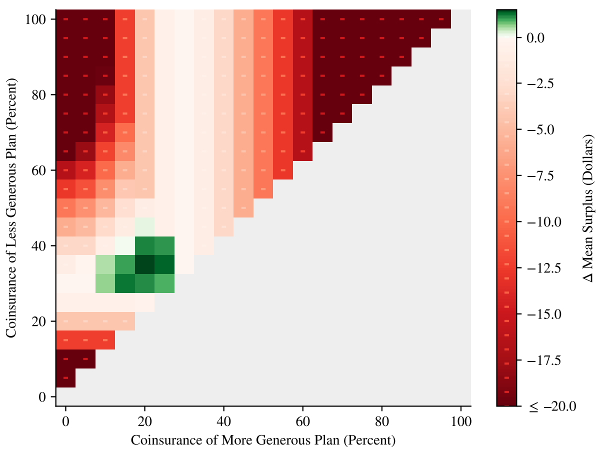
We find that the optimal two plans to offer, with selection, have coinsurance rates of 20% and 35%, with a subsidy level of 20%—i.e., only 20% of the more generous plan's cost difference is reflected in its premium difference from the less generous plan. We predict very modest benefits from offering this choice of plans: a gain of $119.68 relative to the single-plan baseline. This is just 1% above the single-plan optimum, an improvement that is both economically and statistically much smaller than the 16% gained from two plans with assignment. These results are summarized in Table 8 under Simulation A.V.

There are two main reasons why these gains, though positive and statistically significant, are small in magnitude. First, when households choose the plan that is optimal for them, they do not choose the option that equates their benefits from coverage to their impact on total spending (reflected in premium changes). Households that benefit more from insurance controlling for the cost they impose on others (for example, households that are most risk averse) may inefficiently choose less generous plans (Marone and Sabety, 2020); conversely, households that have a greater tendency to over-consume medical care may choose more generous plans. Second, in order to stabilize the market when two plans have different coinsurance rates, the employer finds it optimal to significantly cross-subsidize premiums for the more generous plan. This makes the generous plan more attractive for all households, including those for whom higher cost sharing would be optimal, thereby reducing the extent to which households with different efficient levels of cost sharing can be screened into different plans. Consequently, the gain from differentiating plans is reduced, and the difference in coinsurance rates between the optimal two plans with selection—here, 20% and 35%—is smaller than would be prescribed with assignment (15% and 51%, Simulation A.IV in Table 8).

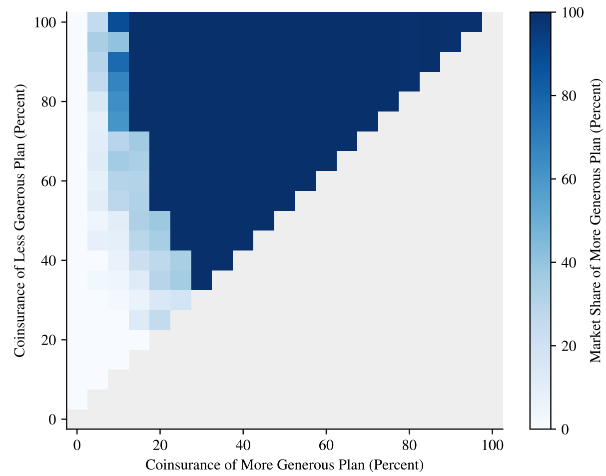
Figure 7 further explores these findings. In both panels the axes represent the coinsurance rates of the two plans, with the more generous plan on the horizontal axis. For each pair of coinsurance rates, we determine the average employee surplus-maximizing premium subsidy level κ, and show: (a) the average employee surplus gain relative to a single plan with zero deductible and the optimal level of coinsurance (Simulation A.II in Table 8); and (b) the market share of the more generous plan. Figure 7(a) shows that there are small gains from offering two plans with different coinsurance rates only locally around the 29% coinsurance rate that is optimal for a single plan. Figure 7(b) shows that, at the optimal premium subsidy level for each pair of coinsurance rates, market shares are often 0% for one of the plans (depicted by either the darkest or lightest regions of the Figure). In areas that are the lightest (i.e., where the more generous plan has 0% share), it is optimal for the employer to let the more generous plan "death spiral" to zero enrollment; in these areas, the gains from steering all employees to the higher cost-sharing option are greater than those from supporting different coverage levels across households. In areas that are the darkest (i.e., where the most generous plan has 100% share), it is optimal for the employer to completely cross-subsidize premium levels ($\kappa = 0$) so that the premiums for the two plans are equal and all employees optimally

collected cover total expected medical expenditures net of out-of-pocket spending, the plan with a lower coinsurance rate has weakly greater premiums than the higher coinsurance rate plan, and households enroll in the plan that maximizes their expected utility. Further details are provided in the Appendix.

Figure 7: Simulation A.V (Two HUGHP-HMO Plans, with Selection)



(a) Change in Average Surplus (Relative to A.II)



(b) Market Share of More Generous Plan

Notes: Figure (a) shows the change in average employer surplus from offering a pair of HUGHP-HMO plans with different coinsurance rates and zero deductibles, relative to offering a single HUGHP-HMO plan with a zero deductible and the optimal 29% coinsurance rate (Simulation A.II). The vertical (horizontal) axis represents the coinsurance rate for the less (more) generous plan. Figure (b) shows the market share of the more generous plan for each pair of coinsurance rates. In both figures, for each pair of coinsurance rates, the premium subsidy level is chosen to maximize average employee surplus. The subsidy level and coinsurance rates are restricted to 5% increments.

choose the more generous plan.

5.5 Multiple Plans: Managing Selection

So far, we have shown that although there are substantial average employee surplus gains from offering multiple plans with assignment, the gains are essentially eliminated once households can freely select among plans. When plans are differentiated solely along financial dimensions, effectively screening households is inherently difficult. As we have mentioned, one reason for this is that at the level of premium cross-subsidization required to stabilize enrollment, households may not choose the plan that provides them with the efficient level of coverage.

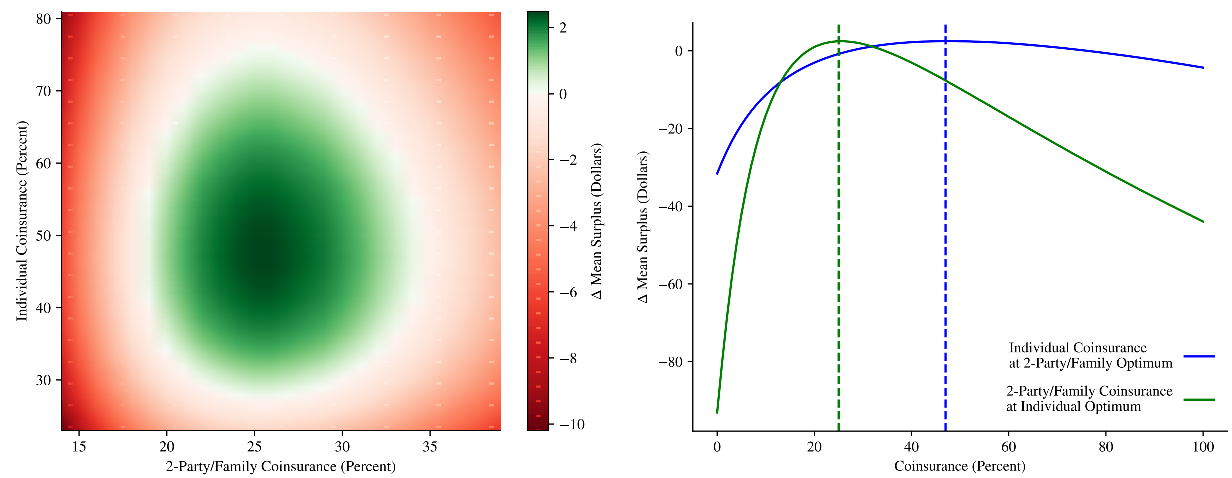
This raises the possibility that an employer can benefit from additional instruments that can be used to "manage selection." To guide the development of strategies to screen enrollees into different plans, we investigate the characteristics of households that our model predicts would be more efficiently served by higher coverage plans. In Simulations A.III and A.IV, described in Section 5.3, we constructed both household-specific tailored plans, and the optimal two plans under assignment. In both simulations, we find that (i) larger households, and (ii) households with members with greater sickness severity (as measured by DCG severity quartile) are optimally provided with higher coverage. For the tailored plan case (Simulation A.III), we perform an OLS regression of the optimal coinsurance rate for each household on household characteristics. Results, reported in Table B9 in the Appendix, show that households in the highest and second-highest DCG quartiles are optimally assigned coinsurance rates that are 44 and 31 percentage points lower than the lowest DCG quartile; and families (with three or more members) are optimally assigned 6–12 percentage points lower coinsurance rates than single-coverage households. In the two-plan simulation (Simulation A.IV), we find that the median DCG quartile of households assigned to the higher-coverage plan is 4, while it is 2 for the lower-coverage plan; and households in the higher-coverage plan have 2.4 members on average while those in the lower-coverage plan have 1.8 members.

Coverage Based on Household Size. Since larger households are, on average, optimally provided with greater insurance coverage when coinsurance rates are household-specific, we explore the effectiveness of creating plans with different coverage levels based on whether the household has single- or non-single (i.e., two-party or family) coverage. In contrast to the assignment policies evaluated earlier in Section 5.3, this policy based on the number of dependents is likely feasible to implement. In this exercise, we allow for a separate coinsurance rate for single-coverage households, and one for non-single-coverage households. We again require that collected premiums cover total expected medical spending net of OOP payments, hold deductibles fixed at zero and OOP maximums at the level provided by Harvard in 2014, and choose the pair of coinsurance rates that maximize average employee surplus.

The results from this exercise are summarized in Table 8 under Simulation B.I. We find that the optimal coinsurance rates for single- and non-single coverage plans are 47% and 26%. These two plans generate only 2% more in average employee surplus than is achievable with a single plan at the optimal coinsurance (Simulation A.II), and only 1% more than is achieved with two plans with selection (Simulation A.V).

To explore why the gains are so limited, we depict in Figure 8(a) the average employee surplus gain relative to a single plan with zero deductible and the optimal level of coinsurance (Simulation A.II in Table 8) as the coinsurance rates for single and non-single coverage varies. Note that here, in contrast to Figure 7(a) depicting gains achievable with two plans with selection, there is a large area of coinsurance rates for which there is an improvement relative to a single plan. However, the gains are always small, and mostly less than $2 per household on average. Figure 8(b) depicts cross sections of Figure 8(a), with each line varying the coinsurance for one plan while holding fixed the other at its optimal level. Note that the two curves are quite different. The single-coverage plan has a much higher optimal coinsurance rate (47%) than both the optimal rate for the non-single-coverage plan (26%) as well as the optimal coinsurance rate for a single plan (29%). Furthermore, due to their lower expected health needs and relatively low out-of-pocket maximums, individuals have relatively flatter curves around their optimum coinsurance rate than do two-party households and families. It therefore follows that gains are limited here relative to offering a single plan: two-party and family households are assigned coinsurance rates close to the single plan optimum, and single-coverage households realize only small benefits from moving their coverage away from that level.

Figure 8: Simulation B.I (Coverage Based on Household Size)



(a) Change in Average Surplus (Relative to A.II) (b) Change in Average Surplus, By Household Size

Notes: Figure (a) shows the change in average employer surplus from offering a single-coverage and non-single-coverage plan with different coinsurance rates, relative to a single plan with zero deductible and the optimal level of coinsurance (Simulation A.II in Table 8). The vertical (horizontal) axis represents the coinsurance rate for the single-coverage (non-single-coverage) plan. Figure (b) depicts the average employee surplus change as either the coinsurance rate for the single-coverage plan changes, holding fixed the coinsurance rate for the non-single-coverage plan at the optimal (blue); or vice-versa (green).

Non-Financial Plan Differentiation. Given the limited gains from screening based on family size, we turn next to adjusting coverage based on households' underlying sickness severity. Though employers are prohibited from varying the level of coverage in response to employee health conditions, they can offer all households the same menu of plans that differ both along financial *and* non-financial dimensions. If these non-financial attributes differentially appeal to households based on their health status, then the employer can pair these characteristics with different coverage levels to better align privately and socially optimal choices.

In our empirical setting, there are two salient non-financial characteristics of plans available to the employer that satisfy this condition: the carrier (HUGHP or Harvard Pilgrim) and plan type (HMO or POS). Harvard Pilgrim (HP) has a denser network of physicians in the suburbs and outlying geographic areas than HUGHP, and is anecdotally viewed as more attractive for larger families living in those areas. POS plans provide access to out-of-network providers, and to specialists without a referral from a primary care physician, and hence are considered to be more attractive than HMO plans for those facing greater sickness severity. The estimates from our model described in Section 4.5 are consistent with such perceptions: both HP and POS plans are more preferred by households with higher expected health care needs.[34]

Motivated by these observations, our last set of simulations examines the gains that are achievable when a firm is no longer restricted to only offering HUGHP-HMO plans, but can also offer

[34]In our model, both having more dependents and having a higher average DCG severity score is predicted to increase a household's expected health needs (λ).

Table 10: Simulated Results (Managing Selection)

		Coins. (%)	Δ Spending ($)	Δ Surplus ($)	ΔSurplus/(*a*)
C. HUGHP HMO and HP POS Plans					
C.I Two Plans, Zero Coins.	HUGHP	0	0.00	0.00	-
(Selection)		-	-	-	-
	HP	0			
		-			
C.II Two Plans, Same Coins.	HUGHP	30	-139.00	121.64[(a)]	1.00
(Selection)		[25, 30]	[-149.92, -69.12]	[63.87, 131.61]	-
	HP	(same)			
C.III Tailored Plans	HUGHP	65 (mean)	-191.50	156.64	1.29
(Assignment)		[63, 66]	[-206.50, -97.51]	[82.54, 168.32]	[1.26, 1.30]
	HP	26 (mean)			
		[25, 28]			
C.IV Two Plans, Assignment	HUGHP	63	-162.72	134.74	1.11
(Assignment)		[61, 65]	[-175.58, -82.54]	[70.69, 144.56]	[1.09, 1.11]
	HP	20			
		[19, 21]			
C.V Two Plans, Optimal Coins	HUGHP	60	-159.38	131.29	1.08
(Selection)		[55, 65]	[-170.19, -82.25]	[68.27, 140.93]	[1.05, 1.09]
	HP	20			
		[20, 25]			

Notes: The table reports the coinsurance rate that maximize average employee surplus, and changes in average total medical spending (Δ Spending) and average employee surplus (Δ Surplus) across all households when both HUGHP-HMO and HP-POS plans with zero deductibles are offered; changes are relative to a single HUGHP-HMO plan and a single HP-POS plan, both with zero coinsurance (C.I). The last column represents the fraction of the change in average employee surplus with only a single HUGHP-HMO and single HP-POS plan (C.II relative to C.I) obtained by the Simulation. All simulations except C.III restrict there to be a single HUGHP-HMO and a single HP-POS plan offered. Simulations C.I and C.II restrict coinsurance rates to either be zero (C.I) or the same (C.II) across plans. Simulation C.III allow both the plan-type and coinsurance rate to vary by household. Simulations C.I, C.II, and C.V allow households to select which plan to enroll in (and restrict the subsidy level and coinsurance rates to 5% increments); simulations C.III and C.IV assign households to a particular plan. 95% confidence intervals are reported below results in brackets, and are obtained by re-estimating the model over 200 bootstrap samples of households and re-computing simulations.

HP-POS plans (i.e., the maximally differentiated plan offering). We conduct five simulations with these two categories of plans, denoted C.I-C.V and summarized in Table 10. These simulations parallel the earlier Simulations A.I-A.V in examining the benefits of: introducing some positive amount of cost sharing (C.II); assigning households to either household-specific plans (C.III) or one of two plans (C.IV); and allowing households to make an unconstrained choice between two plan options (C.V). As before, when consumers are able to choose among multiple plans, we allow the employer to flexibly cross subsidize plans' premiums and restrict the subsidy level and coinsurance rates to 5% increments.

Our measure of average employee surplus for Simulations C.II-C.V is reported relative to a new baseline scenario, referred to as Simulation C.I. In this baseline, the employer offers a single HUGHP-HMO plan and a single HP-POS plan, each with a zero deductible and 0% coinsurance, and employees can freely choose between the two plans.[35] This is the simplest option that could be provided by an employer seeking to differentiate its offerings based on carrier and plan type. It is also a reasonable approximation to Harvard's offerings in the first year of our data.[36]

In Simulation C.II in Table 10, we modify Simulation C.I by allowing the employer to set a positive coinsurance rate that is common to both plans. A coinsurance rate of 30% maximizes average employee surplus, and generates a $122 gain. Note that this is similar to the gain achieved by moving from full insurance to the optimal coinsurance with a single plan (Simulation A.I to A.II). The intuition for this similarity, even with two plans and selection, is straightforward. Each household's preferred plan (HUGHP-HMO or HP-POS) is unaffected by changes in coinsurance, and hence is the same in both C.I and C.II; the determination of the optimal coinsurance rate is then the same as in the single plan case.[37]

Our next two simulations with non-financial differentiation explore the gains from assignment, both with household-specific plans (C.III) and with only two plans (C.IV). In Simulation C.III, each household is assigned their optimal plan and coinsurance rate; in C.IV, they are assigned to either a HUGHP-HMO or a HP-POS plan, each with a different coinsurance rate. The average gains from Simulations C.III and C.IV are 29% and 11% (relative to the two-plan single-coinsurance optimum), again similar to the analogous results with only HUGHP-HMO plans (A.III and A.IV). Figure 9 shows the distribution of optimal coinsurance rates, separately for the HP-POS and HUGHP-HMO plans, from Simulation C.III (household-specific plans). The overall distribution is essentially the same as before, but the low coinsurance rates are heavily focused on the relatively high-severity enrollees who choose the HP-POS plan (dark bars). Spending reductions are focused on enrollees who choose HUGHP-HMO (pale bars). Consistent with this, in Simulation C.IV (two plans with assignment), the HP-POS plan has a much lower coinsurance rate (20%) than the the HUGHP-HMO plan (63%). Taken together, these results suggest that non-financial plan differentiation can be very helpful in managing selection: the coverage levels prescribed under household-specific tailoring are correlated with the underlying carrier and plan-type preferences in the population.

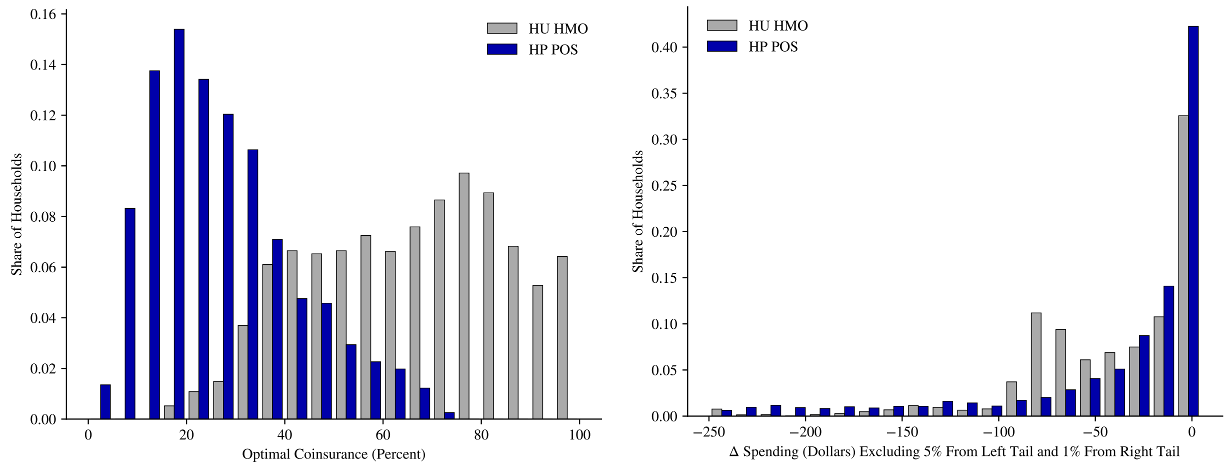
To explore the extent to which this is the case, we repeat Simulation C.II (where there is a single HUGHP-HMO and a single HP-POS plan), but now allow the plans to charge different

[35] All simulations with selection are conducted using the same procedure as described in Section 5.4, where the employer optimally chooses the extent to which it cross subsidizes premiums.

[36] Using our estimated model, we calculate that a firm generates over a $1000 gain in average employee surplus per year by offering both a HUGHP-HMO and a HP-POS plan relative to only a HUGHP-HMO plan. While we believe that there is likely to be a substantial benefit from offering such horizontal differentiation (for example, allowing households to choose their carrier and plan type based on location, sickness severity and preferences), there are several reasons why this estimate may not correspond to the true benefits from offering a second type of plan. Most importantly, our simulations do not account for any cost differences across plans—potentially arising from network, provider, and reimbursement rate differences across carriers, or from differences in seeking out-of-network and specialist care across plan types—or any fixed costs of offering additional plans. By using a new baseline in Simulation C.I to to normalize these exercises, we focus on the effectiveness of introducing differentiation on non-financial dimensions as a tool to manage selection, and avoid quantifying the absolute benefits of offering such differentiation (which is outside the scope of our analysis).

[37] The small differences in the optimal coinsurance rate and surplus gains across the two sets of simulations arise from the curvature of the CARA utility function, combined with differences in the starting level of utility for those households on the HP-POS plan compared to the HUGHP-HMO.

Figure 9: Simulation C.III (Distribution of Optimal Coinsurance Rates and Change in Spending)

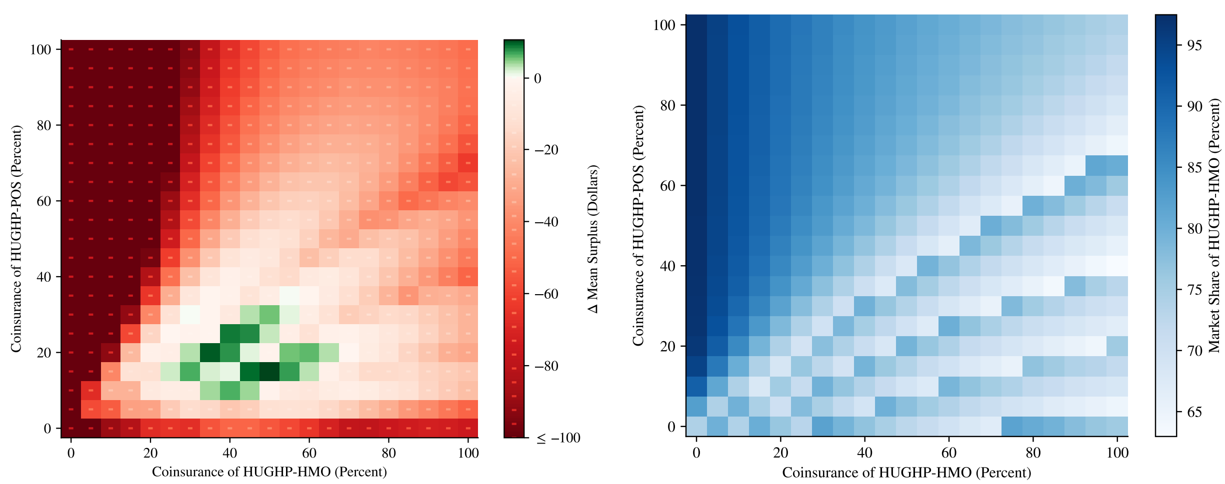


(a) Optimal Coinsurance Rates

(b) Change in Spending

Notes: Figure (a) shows the distribution of optimal coinsurance rates across households when each household is assigned to either an HUGHP-HMO or HP-POS plan, each with a zero deductible and a household-specific coinsurance rate. Figure (b) shows the corresponding distribution of changes in total household medical spending relative to the household being assigned to either a HUGHP-HMO or HP-POS plan, each with a zero deductible and zero coinsurance.

Figure 10: Simulation C.V (HUGHP-HMO and HP-POS plans)



(a) Change in Average Surplus (Relative to C.II)

(b) Market Share of HUGHP-HMO Plan

Notes: Figure (a) shows the change in average employer surplus from offering a single-coverage and non-single-coverage plan with different coinsurance rates, relative to offering a single HUGHP-HMO and single HP-POS plan with zero deductibles and the optimal uniform coinsurance rate of 30% (Simulation C.II). The vertical (horizontal) axis represents the coinsurance rate for the HP-POS (HUGHP-HMO) plan. Figure (b) depicts the market share for the HUGHP-HMO plan. In both figures, for each pair of coinsurance rates, the premium subsidy level is chosen to maximize average employee surplus. The subsidy level and coinsurance rates are restricted to 5% increments.

coinsurance rates. Again, the employer can choose the optimal premium subsidy level, and enrollees are allowed to freely select among plans. Results are reported under Simulation C.V in Table 10. The optimal difference in coverage levels between plans—60% and 20%—is very close to the values chosen with two plans under assignment (63% and 20% in Simulation C.IV). The average employee surplus gain relative to the two-plan single-coinsurance optimum is 8%, achieving 73% of—and not significantly different from—the gains from two plans with assignment. That is, the loss from endogenous selection is effectively removed when we allow for non-financial differentiation. Furthermore, Figure 10 (analogous to Figure 7 with only financial differentiation) confirms that a much larger difference in coinsurance rates can now optimally be supported without one plan's market share being zero, and that modest gains are present for a greater range of coinsurance rates. We find that positive gains are only achieved when coinsurance rates for the HP-POS plan are lower than for the HUGHP-HMO plan.

Thus, by differentiating plans on a dimension that naturally separates the consumers who optimally enroll in high-coverage plans—the relatively sick—from those who do not, and choosing coinsurance rates to complement that separation, the employer may be able to capture sizable benefits from plan choice that might otherwise be lost due to selection.

5.6 Robustness

Risk Aversion. As discussed in Section 4.5, our estimated CARA coefficient is lower than that found in previous work on insurance choice. At our estimate, the average employee is indifferent between receiving nothing with certainty and a 50–50 lottery winning \$100 or losing \$99.97 (or winning \$10,000 and losing \$9,700). Handel (2013) suggests that a CARA coefficient that implies indifference between receiving nothing and a 50–50 lottery winning \$100 or losing an amount between \$98 and \$99 is more typical with what is estimated in the literature. Though we provide potential explanations for this difference in the previous section, we investigate robustness of our simulation results by repeating our main simulations with a larger CARA coefficient.[38] All other parameters are held fixed at their estimated values.

Our findings are summarized in Appendix Table B10. Not surprisingly, a higher degree of risk aversion implies a lower optimal coinsurance rate of 15% for a single HUGHP HMO plan, compared to 29% in our main simulations. A zero deductible remains optimal. The average gain from reaching the single-plan optimum is not statistically different. As a proportion of the gains from moving to the single-plan optimum, the incremental gains from introducing tailored plans are somewhat higher than our main results with a lower CARA coefficient: fully tailored plans generate 33% gains relative to the single plan optimum, compared to 28%; two plan tailoring gains 20% compared to 16%. These differences are fairly small in economic terms, and the extent to which the gains are mitigated when we allow for selection is essentially unchanged from our main results. Thus, aside from the exact level of optimal cost sharing that is predicted by the model, we find no evidence that this larger level of risk aversion meaningfully affects our predictions.[39]

[38] We set $\beta_\psi = -2$, which implies a CARA coefficient of approximately 1.4×10^{-4} and indifference between receiving nothing and a 50–50 lottery winning \$100 or losing \$98.7.

[39] We do not repeat our simulations with plans differentiated by carrier or plan type. Our model, which interacts health needs λ with both carrier and plan type, implies that a household realizes different incremental utility from these characteristics under different realized levels of λ. One implication is that the HP-POS plan is "riskier" than

Moral Hazard. Our results are obtained from a particular model of household behavior that imposes a "multiplicative" relationship between health needs (λ) and price responsiveness (i.e., moral hazard, ω): the increase in a household's spending when moving from no insurance to full insurance is given by $\omega \times \lambda$. This is different from the "additive" specification used in the main analyses of EFRSC and in Marone and Sabety (2020). Our modeling choice is motivated and supported by patterns observed in our data (see Section 2).

We note that while, in theory, this modeling choice could affect our findings, in practice it is not an important determinant for many of our conclusions. For example, introducing some positive amount of cost sharing in order to curb moral hazard is likely to be optimal even with an additive moral hazard specification, as is using non-financial plan differentiation to help manage selection. In addition, we find that relatively sick households are optimally enrolled in higher-coverage options when multiple plans are offered, even though they are predicted to increase their spending by the most when provided with insurance. This finding would likely be strengthened under additive moral hazard.[40]

6 Discussion and Concluding Remarks

An employer designing an insurance plan menu faces many choices. How much cost sharing is optimal? What form should this cost-sharing take? How many plans should be offered?

We provide some guidance on these questions. First, we add to the growing empirical evidence that moral hazard—the sensitivity of medical spending to cost sharing—is significant in the population covered by large employers. Given this finding, we show that requiring a moderate level of enrollee cost sharing can substantially reduce spending and—when this additional spending is less beneficial than its cost—generate substantial average surplus gains for employees. A positive coinsurance rate of approximately 30% with a zero-dollar deductible can be optimal for an employer offering a single plan, yielding average household gains of over $100 per year relative to full coverage and demonstrating the effectiveness of simple cost-sharing arrangements. We also find that gains increase steeply as cost sharing increases away from zero, but then become less sensitive to further changes. This implies that setting some reasonable level of cost sharing, even if not exactly at the optimal level, is still clearly preferable to no cost-sharing at all. Achieving further gains through offering multiple plans requires additional nuance. Though there are substantial benefits from tailoring different levels of coverage to households based on the risk that they face, achieving these gains when households are freely able to choose among plan options is difficult. We

the HUGHP-HMO plan, making it difficult to compare the value of these characteristics across models with different levels of risk aversion. See footnote 29 for related discussion.

[40]Our finding that the amount of increased medical spending due to insurance is increasing in health spending—as opposed to being constant, as would be the case under an "additive" moral hazard specification—may nevertheless be important for our result that a zero deductible is optimal for a single plan. Under our multiplicative specification, the households with the highest increase in dollar spending due to insurance coverage are also those who are sickest and hence most likely to spend past the deductible, even when the deductible is high. Increasing their deductible does not affect their spending in our model. This limits the savings generated by a high-deductible plan, and hence its benefit to consumers, relative to a plan with low deductible and higher coinsurance rate.

show that offering plans that are differentiated along non-financial dimensions that appeal to sicker households can be an effective means of screening enrollees into more generous plans.

Our findings can broadly be interpreted as supporting the actions being taken by some large employers (including the one that is the focus of our study) to impose limited cost sharing on plans where it was absent, and to introduce plans that vary along non-financial dimensions, such as provider networks and policies for obtaining out-of-network care. That said, we intentionally limited our simulations to comparing a variety of relatively generous coverage options as all plans had quite low out-of-pocket maximums. We chose not to expose our households to more risk since this would have taken us further afield from our data and hence from our estimated model of consumer behavior. We thus caution that our findings do not speak directly to the effectiveness of plans that impose greater levels of cost-sharing than those evaluated here (see, for example, Brot-Goldberg et al., 2017, in the context of high-deductible health plans).

We conclude by identifying two promising avenues for future research. First, there may be benefits from developing more finely-tuned cost-sharing schemes, where out-of-pocket costs can vary depending on the individual's diagnosis (Zeckhauser, 1970).[41] Second, there is an open question of how many insurance carriers (as opposed to plans) an employer should offer as options to its employees. Working with multiple insurers may induce them to compete more fiercely with one another on premiums or administrative fees, as documented in Cutler and Reber (1998).[42] It may also protect an employer from hold-up in future negotiations with insurers, particularly if there are large switching costs borne by employers when changing insurers. However, by offering plans from too many insurers, a large employer may limit insurers' bargaining leverage with medical providers, making it harder for them to negotiate favorable medical rates (Ho and Lee, 2017). Balancing these trade-offs requires a more developed understanding of the dynamic issues that underlie negotiations between employers and insurers.

[41] Zeckhauser (1970) develops a model where all individuals have the same preferences and underlying risks, and a single insurance plan is offered whose premiums equal average costs. Under these assumptions, the plan that maximizes individuals' expected utilities provides higher coverage for medical conditions characterized by high non-discretionary spending and low moral hazard than for less severe conditions where moral hazard is substantial.

[42] Note that an employer does not necessarily need to contract with multiple insurers to benefit from competition. For example, the employer could run an auction or play insurers off against one another in negotiations (Ho and Lee, 2019).

References

Abaluck, Jason, and Jonathan Gruber. 2011. "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program." *American Economic Review*, 101: 1180–1210.

Arrow, Kenneth. 1963. "Uncertainty and the Welfare Economics of Medical Care." *The American Economic Review*, 53(5): 141–149.

Bradbury, James, Roy Frostig, Peter Hawkins, Matthew James Johnson, Chris Leary, Dougal Maclaurin, and Skye Wanderman-Milne. 2018. "JAX: composable transformations of Python+NumPy programs."

Brot-Goldberg, Z., A. Chandra, B. Handel, and J. Kolstad. 2017. "What Does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics." *Quarterly Journal of Economics*, 132(3): 1261–1318.

Brown, Zach Y. 2019. "An Empirical Model of Price Transparency and Markups in Health Care." Unpublished.

Buchanan, Joan, Emmett Keeler, John Rolf, and Martin Homer. 1991. "Simulating Health Expenditures under Alternative Insurance Plans." *Management Science*, 37(9): 1067–1090.

Budish, Eric, Robin S. Lee, and John Shim. 2020. "A Theory of Stock Exchange Competition and Innovation: Will the Market Fix the Market?" NBER Working Paper 25855.

Bundorf, M. Kate, Jonathan Levin, and Neale Mahoney. 2012. "Pricing and Welfare in Health Plan Choice." *American Economic Review*, 102(7): 3214–48.

Cardon, James H., and Igal Hendel. 2001. "Asymmetric Information in Health Insurance: Evidence from the National Medical Expenditure Survey." *RAND Journal of Economics*, 32(3): 408–427.

Claxton, G., M. Rae, N. Panchal, A. Damico, N. Bostick, K. Kenward, and H. Whitmore. 2014. "Employer Health Benefits Annual Survery 2014." Kaiser Family Foundation and Health Research and Educational Trust.

Cohen, Alma, and Liran Einav. 2007. "Estimating Risk Preferences from Deductible Choice." *American Economic Review*, 97(3): 745–788.

Cutler, David M., and Sarah J. Reber. 1998. "Paying for Health Insurance: The Tradeoff between Competition and Adverse Selection." *Quarterly Journal of Economics*, 113(2): 433–466.

Dafny, Leemore S., Kate Ho, and Mauricio Varela. 2013. "Let Them Have Choice: Gains from Shifting Away from Employer-Sponsored Health Insurance Towards an Individual Exchange." *American Economic Journal: Economic Policy*, 5(1): 32–58.

Dalton, Christina, Gautam Gowrisankaran, and Robert Town. 2020. "Salience, Myopia, and Complex Dynamic Incentives: Evidence from Medicare Part D." *Review of Economic Studies*, 87: 822–869.

Dubin, Jeffrey A., and Daniel L. McFadden. 1984. "An Econometric Analysis of Residential Electric Appliance Holdings and Consumption." *Econometrica*, 52(2): 345–362.

Einav, Liran, Amy Finkelstein, and Paul Schrimpf. 2015. "The Response of Drug Expenditure to Non-linear Contract Design: Evidence from Medicare Part D." *Quarterly Journal of Economics*, 130(2): 841–899.

Einav, Liran, Amy Finkelstein, Stephen P. Ryan, Paul Schrimpf, and Mark R. Cullen. 2013. "Selection on Moral Hazard in Health Insurance." *American Economic Review*, 103(1): 178–219.

Ericson, Keith Marzilli, and Justin Sydnor. 2017. "The Questionable Value of Having a Choice of Levels of Health Insurance Coverage." *Journal of Economic Perspectives*, 31(4): 51–72.

Geruso, M., T. Layton, G. McCormack, and M. Shepard. 2020. "The Two Margin Problem in Insurance Markets." *Working Paper.*

Ghili, Soheil, Ben Handel, Igal Hendel, and Michael D. Whinston. 2020. "Optimal Long-Term Health Insurance Contracts: Characterization, Computation, and Welfare Effects." Unpublished.

Gowrisankaran, Gautam, Aviv Nevo, and Robert Town. 2015. "Mergers When Prices Are Negotiated: Evidence from the Hospital Industry." *American Economic Review*, 105(1): 172–203.

Handel, Benjamin. 2013. "Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts." *American Economic Review*, 103(7): 2643–2682.

Handel, Benjamin, and Jonathan Kolstad. 2015. "Health Insurance for "Humans": Information Frictions, Plan Choice and Consumer Welfare." *American Economic Review*, 105(8): 2449–2500.

Handel, Benjamin, Igal Hendel, and Michael D. Whinston. 2015. "Equilibria in Health Exchanges: Adverse Selection Versus Reclassification Risk." *Econometrica*, 83(4): 1261–1313.

Heckman, James J. 1981. "The incidental parameters problem and the problem of initial conditions in estimating a discrete time–discrete data stochastic process." In *Structural Analysis of Discrete Data with Econometric Applications.* , ed. Charles F. Manski and Daniel McFadden, 179–195. Cambridge, MA:MIT Press.

Ho, Kate, and Robin S. Lee. 2017. "Insurer Competition in Health Care Markets." *Econometrica*, 85(2): 379–417.

Ho, Kate, and Robin S. Lee. 2019. "Equilibrium Insurer-Provider Networks: Bargaining and Exclusion in Health Care Markets." *American Economic Review*, 109(2): 473–522.

Ho, Kate, and Robin S. Lee. 2020. "Optimal Health Insurance in the Presence of Market Power." Unpublished.

Kowalski, Amanda E. 2015. "Estimating the tradeoff between risk protection and moral hazard with a nonlinear budget set model of health insurance." *International Journal of Industrial Organization*, 43: 122–135.

Manning, Willard, and M. Susan Marquis. 1996. "Health Insurance: The Tradeoff between Risk Pooling and Moral Hazard." *Journal of Health Economics*, 15(5): 609–639.

Marone, Victoria R., and Adrienne Sabety. 2020. "Should There be Vertical Choice in Health Insurance Markets?" Unpublished.

Pakes, Ariel. 1986. "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica*, 54(4): 755–784.

Pauly, Mark V. 1968. "The economics of moral hazard: comment." *The American Economic Review*, 531–537.

Prager, Elena. forthcoming. "Health Care Demand Under Simple Prices: Evidence From Tiered Hospital Networks." *AEJ: Applied Economics.*

Riley, John G. 1979. "Informational Equilibrium." *Econometrica*, 47(2): 331–359.

Rothschild, M., and J. Stiglitz. 1976. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics*, 90(4): 630–649.

Saltzman, Evan. forthcoming. "Managing Adverse Selection: Underinsurance vs. Underenrollment." *RAND Journal of Economics.*

Train, Kenneth E. 2003. *Discrete Choice Methods with Simulation.* Cambridge, U.K.:Cambridge University Press.

Veiga, Andre, and EG Weyl. 2016. "Product Design in Selection Markets." *Quarterly Journal of Economics*, 131(2): 1007–1056.

Wilson, Charles. 1977. "A Model of Insurance Markets with Incomplete Information." *Journal of Economic Theory*, 16(2): 167–207.

Zeckhauser, Richard. 1970. "Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives." *Journal of Economic theory*, 2(1): 10–26.

A Additional Modeling, Estimation and Simulation Details

A.1 Optimal Medical Spending

We derive the analytic form of optimal medical spending for a household on plan j at time t. For exposition, we omit all τ subscripts and assume the following discussion is household-type (single, 2-party, or family with three or more members; and union-status) specific.

First, we derive a household's optimal level of positive medical spending, denoted $m^*_{>0;j,t}(\cdot)$, given the household's ex-post utility in (1) and (4). This is given by:

$$m^*_{>0;j,t}(\lambda;\omega) = \begin{cases} \lambda & \text{if } \lambda < \min(\bar{\lambda}^{ded}_{j,t}, \bar{\lambda}^{oopmax}_{j,t}), \\ \lambda(1+\omega(1-coins_{jt})) & \text{if } \lambda \in [\bar{\lambda}^{ded}_{j,t}, \bar{\lambda}^{oopmax}_{j,t}) \text{ and } \bar{\lambda}^{ded}_{j,t} < \bar{\lambda}^{oopmax}_{j,t}, \\ \lambda(1+\omega) & \text{if } \lambda \geq \max(\bar{\lambda}^{ded}_{j,t}, \bar{\lambda}^{oopmax}_{j,t}), \end{cases} \tag{10}$$

where

$$\bar{\lambda}^{ded}_{j,t} = \frac{2ded_{j,t}}{2+\omega(1-coins_{j,t})} \ ,$$
$$\bar{\lambda}^{oopmax}_{j,t} = \frac{2[oopmax_{j,t} - ded_{j,t}(1-coins_{j,t})]}{2coins_{j,t}(1+\omega) - coins^2_{j,t}\omega} \ ,$$

and $ded_{j,t}$ is the deductible, $oopmax_{j,t}$ the out-of-pocket maximum, and $coins_{j,t}$ the coinsurance rate.

Optimal positive medical spending has the following properties (where we assume $\bar{\lambda}^{ded}_{j,t} < \bar{\lambda}^{oopmax}_{j,t}$ for clarity of exposition).[43] When health needs $\lambda < \bar{\lambda}^{ded}_{j,t}$ (where $\bar{\lambda}^{ded}_{j,t}$ is strictly less than the deductible), a household consumes exactly λ. This arises since, below the deductible, out-of-pocket costs increase one-for-one with spending. However, once the household's health needs reach the threshold $\bar{\lambda}^{ded}_{j,t}$, the household will discontinuously increase its spending to be greater than the deductible. This arises because the household anticipates that spending above the deductible realizes a coinsurance rate strictly less than 100%. In the range $\lambda \in [\bar{\lambda}^{ded}_{jt}, \bar{\lambda}^{oopmax}_{j,t})$, a household spends an extra $\lambda \times (\omega(1-coins_{j,t}))$ above λ. Once $\lambda \geq \bar{\lambda}^{oopmax}_{j,t}$, similar logic explains why a household chooses to again discontinuously increase spending beyond the out-of-pocket maximum; beyond this point, a household spends an extra $\lambda \times \omega$ above λ.

Second, we examine whether a household consumes a strictly positive amount of medical care. It will do so if its increase in utility from consuming the optimal level of positive care $m^*_{>0;j,t}$ exceeds the incurred financial and hassle costs, as well as the opportunity cost of spending nothing; i.e., if:

$$h(\lambda, m^*_{>0;j,t}(\cdot);\omega) - h(\lambda, 0;\omega) - c_{j,t}(m^*_{>0;j,t}(\cdot)) \geq 0 \ . \tag{11}$$

Given the expressions for $m^*_{>0;j,t}(\cdot)$ and $c_{j,t}(\cdot)$, there exists a *threshold health need* $\underline{\lambda}_{j,t}(\omega) > 0$ such that equation (11) is satisfied for all $\lambda \geq \underline{\lambda}_{j,t}(\omega)$, and violated for all $\lambda < \underline{\lambda}_{jt}(\omega)$.[44] This threshold is given by:

$$\underline{\lambda}_{j,t}(\omega) = \begin{cases} \underline{\lambda}_{j,t,1}(\omega) \equiv 2\omega\zeta & \text{if } \underline{\lambda}_{j,t,1}(\omega) < \bar{\lambda}^{ded}_{j,t}, \text{ else} \\ \underline{\lambda}_{j,t,2}(\omega) \equiv \frac{2\omega(ded_{j,t}(1-coins_{j,t})+\zeta)}{(1+\omega(1-coins_{j,t}))^2} & \text{if } \underline{\lambda}_{j,t,2}(\omega) < \bar{\lambda}^{oopmax}_{j,t}, \text{ else} \\ \underline{\lambda}_{j,t,3}(\omega) \equiv \frac{2\omega(oopmax_{j,t}+\zeta)}{(1+\omega)^2} & \text{otherwise,} \end{cases} \tag{12}$$

and this threshold is strictly increasing in hassle cost ζ.

Thus, optimal medical spending for a household is:

$$m^*_{j,t}(\lambda;\omega) = \begin{cases} m^*_{>0;j,t}(\lambda;\omega) & \text{if } \lambda \geq \underline{\lambda}_{j,t} \\ 0 & \text{otherwise.} \end{cases} \tag{13}$$

[43] If instead $\bar{\lambda}^{ded}_{j,t} \geq \bar{\lambda}^{oopmax}_{j,t}$, then a household will spend λ until its health needs reach $\bar{\lambda}^{oopmax}_{j,t}$; upon exceeding that threshold, the household will spend $\lambda(1+\omega)$ as it will then have met its out-of-pocket maximum.

[44] Given our restriction to concave cost-sharing rules characterized by a deductible, coinsurance rate, and out-of-pocket maximum, the expression $h(\lambda, m^*_{j,t}(\cdot);\omega) - h(\lambda, 0;\omega) - c_{j,t}(m^*_{j,t}(\cdot))$ is strictly increasing in λ. Since this expression is strictly negative when $\lambda = 0$ and strictly positive for some positive $\lambda > 0$, the result follows.

A.2 Estimation Details

Consider household i and its observed decisions $\boldsymbol{d}_i = \{j_t^o, m_t^o\}_{t=1,\dots,3}$, where j_t^o is the plan choice for the household at time t and m_t^o is its medical spending.[45] Denote by $\boldsymbol{\gamma} \equiv (\{F_{\lambda,i,t}\}_t, \omega_i)$ the objects associated with a given household i, drawn from distribution $F_{\gamma,i}(\boldsymbol{\gamma};\boldsymbol{\theta})$ which is parameterized by $\boldsymbol{\theta}$. The likelihood of observing the household's decisions $\boldsymbol{d}_i$ is:

$$\mathcal{L}_i(\boldsymbol{d}_i;\boldsymbol{\theta}) = \int_\gamma \Bigg(\Big(\sum_{k\in\mathcal{J}_0} Pr_0(k;\boldsymbol{\gamma},\boldsymbol{\theta}) \times Pr_1(j_1^o|j_0 = k;\boldsymbol{\gamma},\boldsymbol{\theta}) \Big) \times \Big(Pr_2(j_2^o|j_1^o;\boldsymbol{\gamma},\boldsymbol{\theta}) \Big) \times \Big(Pr_3(j_3^o|j_2^o;\boldsymbol{\gamma},\boldsymbol{\theta}) \Big) \times \Big(\prod_{t=1,\dots,3} f_{m,t}(m_t^o|j_t^o,\boldsymbol{\gamma},\boldsymbol{\theta}) \Big) \Bigg) dF_{\gamma,i}(\boldsymbol{\gamma};\boldsymbol{\theta}) \ ,$$

where $\mathcal{J}_t$ denotes the set of plans available in period t; $Pr_0(k;\cdot)$ denotes the probability the household's period-0 insurance plan choice (which is unobserved) was $j_0 = k$; $Pr_t(j_t|j_{t-1};\cdot)$ denotes the probability of choosing plan j_t in period $t = 1\dots,3$ given prior enrollment in j_{t-1} (which is relevant due to the presence of switching costs); and $f_{m,t}(\cdot)$ denotes the probability density of medical spending.

For any candidate parameter vector $\boldsymbol{\theta}$, we evaluate each household's likelihood contribution via simulation by taking N_S draws of $\boldsymbol{\gamma}$ (each simulation indexed by s) and computing:

$$\hat{\mathcal{L}}_i(\boldsymbol{d}_i;\boldsymbol{\theta}) = \frac{1}{N_S}\sum_{s=1}^{N_S} \Bigg(\Big(\sum_{k\in\mathcal{J}_0} Pr_0(k;\boldsymbol{\gamma}_s,\boldsymbol{\theta}) \times Pr_1(j_1^o|j_0 = k;\boldsymbol{\gamma}_s,\boldsymbol{\theta}) \Big) \times \Big(Pr_2(j_2^o|j_1^o;\boldsymbol{\gamma}_s,\boldsymbol{\theta}) \Big) \times \Big(Pr_3(j_3^o|j_2^o;\boldsymbol{\gamma}_s,\boldsymbol{\theta}) \Big) \times \Big(\prod_{t=1,\dots,3} f_{m,t}(m_t^o|j_t^o;\boldsymbol{\gamma}_s,\boldsymbol{\theta}) \Big) \Bigg) \ , \tag{14}$$

where each object in (14) is computed as follows:

1. Each household's plan choice probabilities $Pr_t(j_t|j_{t-1}^o;\boldsymbol{\gamma},\boldsymbol{\theta})$ for each plan and period $t = 1,\dots,3$ is computed using the modified smoothed Accept-Reject function from Handel, Hendel and Whinston (2015) (see also Train, 2003):
$$Pr_t(j_t|j_{t-1};\boldsymbol{\gamma},\boldsymbol{\theta}) = \Big(\frac{(-v_{j,t}(\cdot))^{-1}}{\sum_{k\in\mathcal{J}_t}(-v_{k,t}(\cdot))^{-1}} \Big)^\eta \Big/ \sum_{l\in\mathcal{J}_t} \Big(\frac{(-v_{l,t}(\cdot))^{-1}}{\sum_{k\in\mathcal{J}_t}(-v_{k,t}(\cdot))^{-1}} \Big)^\eta \ ,$$
and the integral used to compute $\{v_{j,t}(\boldsymbol{\gamma}, j_{t-1}^o)\}_{\forall j}$ (corresponding to the household's expected utility from enrolling in plan j, given by (2)) is approximated using N_H draws of health shocks λ_t, and $\eta > 0$ is a smoothing parameter.

2. Each household's density of medical spending $f_{m,t}(m_t^o;j_t^o,\boldsymbol{\gamma},\boldsymbol{\theta})$ for each year is computed as follows. We assume that the observed medical spending, if positive, is given by $m^*_{>0;j,t}(\lambda,\omega)\times\nu$, where $m^*_{>0;j,t}(\lambda,\omega)$ is the optimal positive level of medical spending for the household on plan j (see (4)), and ν is multiplicative measurement error, where $\log(\nu) \sim \mathcal{N}(-\sigma_\nu^2/2, \sigma_\nu)$ and ν has mean 1. Then $f_{m,t}(\cdot)$ can be written as:
$$f_{m,t}(m;j,\boldsymbol{\theta}) = \begin{cases} F_{\lambda,t}(\underline{\lambda}_{j,t}(\omega)) & \text{if } m_t^o = 0; \\ (1 - F_{\lambda,t}(\underline{\lambda}_{j,t}(\omega)))\times \\ \quad \int \Bigg(\Big(\phi\big(\frac{\log(m/m^*_{>0;j,t}(\lambda,\omega)) + (\sigma_\nu^2/2)}{\sigma_\nu}\big)/\sigma_\nu \Big)\times & \text{if } m^o > 0, \\ \qquad \Big(\underbrace{(m^*_{>0;j,t}(\lambda,\omega))^{-1}}_{|d\nu/dm|} \Big) f_{\lambda,t}(\lambda|\lambda > \underline{\lambda}_{j,t}(\omega))d\lambda \Bigg) & \end{cases} \tag{15}$$
where $\phi(\cdot)$ is the probability density of the standard normal distribution. The presence of measurement error allows the model to rationalize any level of medical spending observed in the data.[46]

[45] For exposition, we focus on a household i that is present in all three periods $t = 1,\dots,3$ of our data, and did not enter or exit in any of these periods. For households that enter or leave during our sample time frame, only the years where the household is present are used in estimation, and medical spending is annualized for years in which the household is only partially present.

[46] There may be discontinuities in the function $m^*_{>0;j,t}(\cdot)$ due to non-linearities in the each plan's cost-sharing schedule $c_{j,t}(\cdot)$. Hence, although there is a unique $m^*_{>0;j,t}(\cdot)$ for any value of $\lambda \geq 0$, absent measurement error, certain levels of observed medical spending cannot be rationalized by any value of λ.

3. Each household's unobserved period-0 plan choice $Pr_0(j_0; \boldsymbol{\gamma}, \boldsymbol{\theta})$ is approximated as follows. Denote by $\boldsymbol{P}_0(\boldsymbol{\gamma}, \boldsymbol{\theta})$ the $|\mathcal{J}_0| \times 1$ vector with each element k corresponding to $Pr_0(k; \boldsymbol{\gamma}, \boldsymbol{\theta})$. Then:

$$\boldsymbol{P}_0(\boldsymbol{\gamma}, \boldsymbol{\theta}) \approx \boldsymbol{P}_1(\boldsymbol{\gamma}, \boldsymbol{\theta}) \times [\boldsymbol{T}_{j_t|j_{t-1}}(\boldsymbol{\gamma}, \boldsymbol{\theta})]^{tenure_i - 1}$$

   where $\boldsymbol{P}_1(\boldsymbol{\gamma}, \boldsymbol{\theta})$ is the $|\mathcal{J}_1| \times 1$ vector with each element k corresponding to $Pr_1(j_1 = k | j_0 = \emptyset; \boldsymbol{\gamma}, \boldsymbol{\theta})$ (i.e., the plan choice probabilities for a household with no prior choice of insurance plan); $\boldsymbol{T}_{j_t|j_{t-1}}(\boldsymbol{\gamma}, \boldsymbol{\theta})$ is a $|\mathcal{J}_1| \times |\mathcal{J}_1|$ matrix where element m, n is $Pr_1(j_1 = m | j_0 = n; \boldsymbol{\gamma}, \boldsymbol{\theta})$ (i.e., the plan choice transition matrix derived from $Pr(j_1 | j_0, \boldsymbol{\gamma}, \boldsymbol{\theta})$); and $tenure_i$ is the number of years the household was employed at $t = 1$ (observed in our data). Our approximation is exact if all plan and household characteristics are the same in period 1 as they were in prior years, and the household was not enrolled in any plan in the employer's choice set prior to employment by the firm.

Note that to control of unobserved heterogeneity, this procedure draws household objects $\boldsymbol{\gamma}$ from the distribution $F_{\gamma,i}(\cdot)$ and simulates forward its choices in a manner similar to Pakes (1986). We set $N_S = 100$, $N_H = 100$, $\eta = 300$, and $\sigma_\nu = 0.1$ (which implies that ν has a standard deviation of 0.1).

Our estimate of $\boldsymbol{\theta}$ is $\hat{\boldsymbol{\theta}} = \arg\max_{\boldsymbol{\theta}} \sum_i \left(\ln(\hat{\mathcal{L}}_i(\boldsymbol{d}_i; \boldsymbol{\theta})) + C_i \right)$, where C_i is a first-order asymptotic bias correction term for simulated maximum likelihood.[47] Implementation relies on the JAX software package (Bradbury et al., 2018) for automatic differentiation, JIT compilation, and GPU support.

A.3 Simulations: Determination of Premiums and Enrollment with Selection

As discussed in the main text, we allow for an employer to manage adverse selection by allowing the premium difference between two plans to be less than the difference in their underlying costs: i.e., the employer can choose a subsidy level $\kappa \in [0, 1]$ that equals the ratio of the difference in plans' premiums to the difference in the plans' average costs (i.e., medical spending net out-of-pocket payments).

Formally, suppose there are N_1 enrollees in plan 1 and N_2 enrollees in the more-generous plan 2; for simplicity, suppose all households are individuals (there are no families). Define E_j to be total spending net of out-of-pocket payments in plan j for $j = 1, 2$; define average net spending across households to be $AC_j = E_j / N_j$. Then, the individual premium in plans 1 and 2 (denoted p_1 and p_2) are determined by the following two equations:

$$\begin{aligned} p_2 &= p_1 + \kappa(AC_2 - AC_1) && \text{(premium difference reflects } \kappa \text{ of average cost difference)}, \\ p_1 N_1 + p_2 N_2 &= E_1 + E_2 && \text{(premiums cover total spending).} \end{aligned} \tag{16}$$

To obtain outcomes for each (κ, c_1, c_2) triple, we utilize the following procedure. First, we initialize the premiums for each plan to equal average costs as if every household was enrolled on that plan; denote these premiums (p_1^0, p_2^0). Then, for each iteration $n = 1, 2, 3, \ldots$:

(i) Compute enrollment for each household given premiums (p_1^{n-1}, p_2^{n-1});

(ii) Given enrollment decisions in (i), compute expected net spending and average costs on each plan, and determine candidate premiums (p_1', p_2') to solve (16);[48]

(iii) For any household that switches plans in step (i) from iteration $n - 1$, determine whether it still wishes to do so given updated premiums (p_1', p_2'); if not, re-assign that household to its plan choice in iteration $n - 1$;

(iv) Given enrollment decisions in (iii), compute expected net spending and average costs on each plan, and update premiums (p_1^n, p_2^n) to solve (16).

---

[47] Following Gourieroux and Monfort (1997), we use the following correction term for each household i:

$$C_i = \frac{1}{2} \frac{\sum_s (L_{is}(\boldsymbol{\gamma}_s) - \bar{L}_i)^2 / N_S}{(\bar{L}_i)^2}$$

where $\bar{L}_i = (\sum_{s=1}^{N_S} L_{is}(\boldsymbol{\gamma}_s)) / N_S$, and $L_{is}(\boldsymbol{\gamma}_s)$ represents all terms inside the outer summation on the right-hand-side of (14).

[48] In our setting, premiums for 2-party and family households equal 2.7 times the single household premium. To account for this, we modify the second-line of (16) as follows:

$$(p_1 \times N_{1,single} + 2.7 \times p_1 \times N_{1,fam}) + (p_2 \times N_{2,single} + 2.7 \times p_2 \times N_{2,fam}) = E_1 + E_2 \ ,$$

where $N_{j,single}$ and $N_{j,fam}$ are the number of single- and non-single coverage households on plan j.

We repeat steps (i)-(iv) until $\max_j |p_j^n - p_j^{n-1}| < \$1$.[49,50]

B Additional Tables and Figures

Table B1: Difference-in-Difference Results, DCG Quartile 4

Observations	Mean	Fraction with zero spending	Percentile of Spending Distribution 10th	25th	50th	75th	90th
Individuals							
Reduced-Form Diff-in-Diff, All Severity Scores							
Difference-in-Differences (Levels)	-576	0.017	0	-64	-154	-353	-1,643
Difference-in-Differences (Percentages)	-15.1%	12.6%	-	-18.2%	-10.7%	-8.5%	-19.0%
Reduced-Form Diff-in-Diff, Mean Severity Score Quartile 4							
Difference-in-Differences (Levels)	-1,575	-0.002	-313	-347	-1,140	-2,166	-5,643
Difference-in-Differences (Percentages)	-19.2%	-12.5%	-33.1%	-15.6%	-24.2%	-21.8%	-28.3%
Two-Party and Family Households							
Reduced-Form Diff-in-Diff, All Severity Scores							
Difference-in-Differences (Levels)	-1,001	0.002	-307	-96	-349	-886	-3,830
Difference-in-Differences (Percentages)	-7.7%	19.7%	-18.6%	-2.7%	-5.1%	-6.3%	-13.3%
Reduced-Form Diff-in-Diff, Mean Severity Score Quartile 4							
Difference-in-Differences (Levels)	-1,516	-0.005	-92	-396	336	-4,100	-9,770
Difference-in-Differences (Percentages)	-6.6%	-	-2.0%	-4.0%	2.7%	-15.1%	-16.8%

Notes: Basic difference-in-difference results summarizing annual household spending for individuals (top panel) and two-party households and families (bottom panel) that comprise the 'base (2014-5) sample' summarized in column 2 of Table 1. For each panel, first set of results are repeated from Table 4 for ease of comparison. Second set of results focus just on households (in treatment and control groups) whose mean DCG score is in the highest quartile of the distribution for their family type and year. See Section 3.1 for details.

[49] Although the convergence criterion is computed using the implied premiums (p_1^n, p_2^n) that solve (16), for each $j = 1, 2$ we use $\tilde{p}_j^n = \alpha p_j^n + (1-\alpha)p_j^{n-1}$ as the value of premiums for each subsequent iteration, where $\alpha \in (0,1)$ is a tuning parameter.

[50] Steps (iii)-(iv) resolve convergence issues that arise with a finite number of households and the potential impact that a single household with a large amount of spending can have on premiums. For example, it can be the case that some household i with high medical spending enrolled on plan 1 would prefer to be on plan 2 if the premiums for the two plans were set given household i was enrolled on plan 1; but, if household i switched to plan 2 and premiums adjusted, household i would prefer to switch back to plan 1. Steps (iii)-(iv) require that at any set of enrollment decisions and premiums, those households that are allowed to adjust their enrollment decisions are those that would wish to do so even if premiums accounted for these adjustments. Our requirement that such "deviations" remain profitable to certain "reactions" shares similarities with alternative equilibrium concepts developed to address equilibrium non-existence issues in markets characterized by adverse selection (e.g., Wilson, 1977; Riley, 1979; Budish, Lee and Shim, 2020).

Table B2: Difference-in-Difference Results, Outpatient Visits

	Observations	Mean	Fraction with zero spending	Percentile of Spending Distribution 10th	25th	50th	75th	90th
Individuals								
Control (Union)								
2014 Spending	1,763	1,469	0.494	0	0	30	1,141	4,461
2015 Spending	1,763	1,673	0.493	0	0	36	1,201	4,838
Treated (Non-Union)								
2014 Spending	2,108	1,253	0.543	0	0	0	778	3,780
2015 Spending	2,108	1,286	0.558	0	0	0	798	3,470
Treated-Control Differences (Levels, Non-Union - Union)								
2014 Difference		-216	0.049	0	0	-30	-363	-680
2015 Difference		-387	0.064	0	0	-36	-403	-1,368
2015-2014 Differences (Levels)								
Control (Union)		204	-0.001	0	0	6	60	377
Treated		33	0.015	0	0	0	21	-311
Difference-in-Differences (Levels)		-171	0.015	0	0	-6	-40	-688
Difference (Percentages)								
Control		13.9%	-0.1%	-	-	19.8%	5.3%	8.5%
Treated		2.6%	2.7%	-	-	-	2.6%	-8.2%
Difference-in-Differences (Percentages)		-11.3%	2.8%	-	-	-	-2.6%	-16.7%
Two-Party and Family Households								
Control (Union)								
2014 Spending	1,141	5,109	0.15	0	345	1,956	5,964	13,631
2015 Spending	1,141	6,342	0.156	0	437	2,293	7,242	16,820
Treated (Non-Union)								
2014 Spending	2,695	4,774	0.164	0	338	1,823	5,764	12,534
2015 Spending	2,695	5,077	0.172	0	274	1,821	5,633	12,245
Treated-Control Differences (Levels, Non-Union - Union)								
2014 Difference		-335	0.014	0	-7	-133	-200	-1,097
2015 Difference		-1,265	0.016	0	-163	-473	-1,609	-4,575
2015-2014 Differences (Levels)								
Control (Union)		1,233	0.006	0	92	338	1,278	3,189
Treated		303	0.008	0	-64	-2	-131	-289
Difference-in-Differences (Levels)		-930	0.002	0	-155	-340	-1,409	-3,478
Difference (Percentages)								
Control		24.1%	4.1%	-	26.5%	17.3%	21.4%	23.4%
Treated		6.3%	5.0%	-	-18.9%	-0.1%	-2.3%	-2.3%
Difference-in-Differences (Percentages)		-17.8%	0.9%	-	-45.4%	-17.4%	-23.7%	-25.7%

Notes: Basic difference-in-difference results summarizing annual household outpatient spending for individuals (top panel) and two-party households and families (bottom panel) that comprise the 'base (2014-5) sample' summarized in column 2 of Table 1.

Table B3: Difference-in-Difference Results, Physician Office Visits

	Observations	Mean	Fraction with zero spending	Percentile of Spending Distribution 10th	25th	50th	75th	90th
Individuals								
Control (Union)								
2014 Spending	1,763	1,370	0.201	0	130	667	1,902	3,755
2015 Spending	1,763	1,532	0.175	0	187	831	2,070	4,106
Treated (Non-Union)								
2014 Spending	2,108	1,213	0.202	0	112	595	1,620	3,164
2015 Spending	2,108	1,293	0.192	0	128	610	1,727	3,395
Treated-Control Differences (Levels, Non-Union - Union)								
2014 Difference		-157	0.0	0	-19	-72	-282	-591
2015 Difference		-238	0.017	0	-59	-221	-343	-712
2015-2014 Differences (Levels)								
Control (Union)		162	-0.027	0	57	164	169	352
Treated		80	-0.01	0	16	15	107	231
Difference-in-Differences (Levels)		-82	0.017	0	-41	-148	-61	-121
Difference (Percentages)								
Control		11.8%	-13.2%	-	43.7%	24.6%	8.9%	9.4%
Treated		6.6%	-4.9%	-	14.4%	2.6%	6.6%	7.3%
Difference-in-Differences (Percentages)		-5.2%	8.3%	-	-29.3%	-22.0%	-2.3%	-2.1%
Two-Party and Family Households								
Control (Union)								
2014 Spending	1,141	4,054	0.041	543	1,579	3,134	5,489	8,487
2015 Spending	1,141	4,173	0.036	620	1,657	3,295	5,688	8,717
Treated (Non-Union)								
2014 Spending	2,695	4,519	0.017	930	1,873	3,509	5,989	9,326
2015 Spending	2,695	4,593	0.017	908	1,877	3,473	6,012	9,590
Treated-Control Differences (Levels, Non-Union - Union)								
2014 Difference		465	-0.024	387	294	375	499	839
2015 Difference		420	-0.019	288	219	178	324	873
2015-2014 Differences (Levels)								
Control (Union)		119	-0.005	77	78	161	199	230
Treated		74	0.0	-22	4	-36	24	264
Difference-in-Differences (Levels)		-44	0.005	-99	-75	-196	-175	34
Difference (Percentages)								
Control		2.9%	-12.8%	14.2%	5.0%	5.1%	3.6%	2.7%
Treated		1.6%	0.0%	-2.4%	0.2%	-1.0%	0.4%	2.8%
Difference-in-Differences (Percentages)		-1.3%	12.8%	-16.5%	-4.8%	-6.1%	-3.2%	0.1%

Notes: Basic difference-in-difference results summarizing annual household spending on physician office visits for individuals (top panel) and two-party households and families (bottom panel) that comprise the 'base (2014-5) sample' summarized in column 2 of Table 1.

Table B4: Difference-in-Difference Results, High Value Care Categories

	Observations	Mean	Fraction with zero spending	Percentile of Spending Distribution 10th	25th	50th	75th	90th
Individuals								
Control (Union)								
2014 Spending	1,763	139	0.436	0	0	27	260	330
2015 Spending	1,763	140	0.438	0	0	21	251	360
Treated (Non-Union)								
2014 Spending	2,108	119	0.435	0	0	19	221	307
2015 Spending	2,108	122	0.466	0	0	19	230	330
Treated-Control Differences (Levels, Non-Union - Union)								
2014 Difference		-20	-0.001	0	0	-8	-39	-22
2015 Difference		-18	0.028	0	0	-2	-22	-30
2015-2014 Differences (Levels)								
Control (Union)		0	0.002	0	0	-6	-9	31
Treated		2	0.031	0	0	0	9	23
Difference-in-Differences (Levels)		2	0.029	0	0	7	18	-8
Difference (Percentages)								
Control		0.3%	0.5%	-	-	-23.5%	-3.3%	9.3%
Treated		1.8%	7.1%	-	-	1.0%	4.2%	7.3%
Difference-in-Differences (Percentages)		1.5%	6.6%	-	-	24.4%	7.5%	-2.0%
Two-Party and Family Households								
Control (Union)								
2014 Spending	1,141	601	0.085	32	277	543	867	1,178
2015 Spending	1,141	586	0.086	19	275	523	850	1,160
Treated (Non-Union)								
2014 Spending	2,695	694	0.056	151	328	627	961	1,353
2015 Spending	2,695	630	0.071	111	323	588	891	1,188
Treated-Control Differences (Levels, Non-Union - Union)								
2014 Difference		92	-0.029	119	51	84	94	176
2015 Difference		44	-0.015	92	48	64	42	28
2015-2014 Differences (Levels)								
Control (Union)		-15	0.001	-13	-2	-20	-17	-18
Treated		-64	0.016	-40	-6	-40	-70	-165
Difference-in-Differences (Levels)		-48	0.015	-27	-3	-19	-52	-148
Difference (Percentages)								
Control		-2.5%	1.0%	-40.0%	-0.8%	-3.7%	-2.0%	-1.5%
Treated		-9.2%	28.0%	-26.6%	-1.7%	-6.3%	-7.3%	-12.2%
Difference-in-Differences (Percentages)		-6.6%	27.0%	13.3%	-0.9%	-2.6%	-5.3%	-10.7%

Notes: Basic difference-in-difference results summarizing annual household spending on high value care for individuals (top panel) and two-party households and families (bottom panel) that comprise the 'base (2014-5) sample' summarized in column 2 of Table 1. Following preventative care guidelines from the Affordable Care Act (2010, we define high value care to include HPV/Hepatitis vaccines; flu vaccines; preventive exams and counseling; and STI screening.

Table B5: Enrollment in POS-Plus Plan, by DCG Severity Score

DCG Quartile		Continuing Employees		New Employees	
		N	% POS-P	N	% POS-P
Family	1	133	26%	39	51%
(Max DCG)	2	248	28%	20	60%
	3	281	40%	21	62%
	4	320	41%	17	53%
Family	1	139	25%	34	53%
(Mean DCG)	2	220	29%	25	56%
	3	270	39%	20	60%
	4	353	41%	18	56%
Individual	1	101	11%	58	47%
	2	113	26%	37	59%
	3	150	27%	34	41%
	4	230	33%	25	60%

Notes: Comparison of probability of enrolling in a POS-Plus plan by DCG severity score and family type. Sample consists of non-unionized employees who are in the data in both 2015 and 2016; are enrolled in a POS plan in 2015; and who choose a POS or POS-Plus plan in 2016. DCG quartiles are computed by taking the mean or the max DCG score for each family and then finding quartiles within-family type and year. Right-hand column considers only new employees in 2016.

Table B6: Enrollment in POS-Plus Plan, by Family Size

Family Size	Continuing Employees		New Employees	
	N	% POS-P	N	% POS-P
1	594	27%	154	51%
2	327	32%	41	51%
3	220	37%	26	54%
4+	435	37%	30	63%

Notes: Comparison of family size distributions for the subsample of non-unionized employees who are in the data in both 2015 and 2016; are enrolled in a POS plan in 2015; and who choose a POS or POS-Plus plan in 2016. Right-hand column considers only new employees in 2016.

Table B7: Parameter Estimates

Parameter			Estimate	SE
$\boldsymbol{\theta}_1$	Plan Choice ($\boldsymbol{\beta}_x$)	POS(P)	-1.391	0.072
		HP	-0.441	0.048
		HP x POS(P)	0.828	0.066
		HP x Cambridge	-0.893	0.083
	(interacted w/ λ)	POS(P)	0.179	0.012
		HP	0.249	0.015
		HP x POS(P)	-0.129	0.012
		HP x Cambridge	-0.060	0.014
	Switching Cost	δ	3.789	0.123
$\boldsymbol{\theta}_2$	Health: Mean ($\boldsymbol{\beta}_\lambda$)	Tier 2	0.284	0.026
		Tier 3	0.408	0.032
		Age 40+	0.131	0.026
		Age 50+	0.032	0.029
		DCG Q2	0.727	0.029
		DCG Q3	1.225	0.032
		DCG Q4	1.831	0.038
		Single x Union	-1.226	0.044
		2-Party x Union	0.267	0.054
		Family x Union	0.549	0.047
		Single x Non-Union	-1.498	0.039
		2-Party x Non-Union	-0.044	0.055
		Family x Non-Union	0.415	0.043
		Single x 2015	0.020	0.025
		2-Party x 2015	0.031	0.045
		Family x 2015	0.035	0.024
		Single x 2016	0.111	0.024
		2-Party x 2016	0.075	0.043
		Family x 2016	0.064	0.031
	Health: Variance ($\ln(\boldsymbol{\sigma}_\lambda)$)	Single	0.083	0.017
		2-Party	-0.061	0.033
		Family	-0.312	0.025
	Health: Unobs. Variance in Mean	$\ln(\sigma_\mu)$	-0.388	0.027
	Moral Hazard	ω_1	-3.348	0.182
		ω_2	-1.335	0.235
		$\beta_{\omega,1}$	-0.565	0.083
	Risk Aversion	β_ψ	-5.785	0.036
	Hassle Costs ($\boldsymbol{\beta}_\zeta$)	Single x Union	-0.492	0.174
		2-Party x Union	-0.161	0.213
		Family x Union	0.538	0.196
		Single x Non-Union	-0.592	0.179
		2-Party x Non-Union	-0.487	0.256
		Family x Non-Union	-0.175	0.231

Notes: Parameter estimates from health plan choice and utilization model in Section 4. Utility measured in \$000s.

Table B8: Single Plan, Optimal Coinsurance with Variable Deductibles

Deductible	Coins. (%)	Avg. Spending ($)	Δ Surplus ($)
$0	29	8260.72	118.20
$250	28	8262.35	113.19
$500	27	8263.37	106.48
$750	25	8264.17	100.68
$1,000	22	8263.67	95.09
$1,250	19	8261.87	89.50
$1,500	15	8259.60	83.70
$1,750	9	8257.26	77.44
$2,000	-	8254.74	70.57

Notes: Each row provides the optimal (average employee surplus maximizing) coinsurance rate, the level of average spending across households, and the change in average employee surplus relative to full insurance for a single HUGHP HMO plan at a given deductible level. The first row ($0 deductible) corresponds to Simulation A.II in Table 8.

Table B9: Regression of Optimal Tailored Coinsurance Rate on Household Characteristics

	Coefficient	SE
Single x Union	0.787	0.005
Two-Party x Union	0.652	0.008
Family x Union	0.724	0.008
Single x Non-Union	0.853	0.004
Two-Party x Non-Union	0.705	0.007
Family x Non-Union	0.732	0.006
Tier 2	-0.065	0.006
Tier 3	-0.096	0.006
Age 40+	-0.034	0.005
Age 50+	-0.009	0.005
DCG Q2	-0.184	0.005
DCG Q3	-0.312	0.005
DCG Q4	-0.440	0.005
N	8827	
R^2	0.598	

Notes: OLS Regression of the optimal (average employee surplus maximizing) household-specific coinsurance rate (HUHGP HMO) from Simulation A.III (Table 8) on household characteristics.

Table B10: Simulated Results (HUGHP HMO Plans Only), Robustness to Risk Aversion

		Main Estimates ($\beta_\psi = -5.8$)			Robustness ($\beta_\psi = -2$)		
(HUGHP HMO Plans Only)		Coins.	Δ Surplus	$\frac{\Delta\text{Surplus}}{(a)}$	Coins.	Δ Surplus	$\frac{\Delta\text{Surplus}}{(b)}$
Single Plan (Section 5.2)							
A.I Fixed Coins.		0	0	-	0	0.00	-
		-	-	-	-	-	-
A.II Optimal Coins.		29	118.20$^{(a)}$	1.00	15	129.62$^{(b)}$	1.00
		[28, 31]	[60.38, 127.67]	-	[9, 16]	[50.25, 143.45]	-
Multiple Plans with Assignment (Section 5.3)							
A.III Tailored Plans	Mean	48	150.84	1.28	28	171.80	1.33
		[46, 49]	[76.49, 161.91]	[1.25, 1.28]	[19, 29]	[69.39, 188.36]	[1.30, 1.37]
A.IV Two Plans	Plan A	15	137.08	1.16	8	154.99	1.20
		[14, 16]	[69.79, 147.95]	[1.15, 1.17]	[5, 9]	[61.75, 170.01]	[1.18, 1.22]
	Plan B	51			32		
		[49, 53]			[19, 35]		
Multiple Plans with Selection (Section 5.4)							
A.V Two Plans	Plan A	20	119.68	1.01	10	136.74	1.05
		[15, 25]	[61.29, 129.14]	[1.01, 1.02]	[5, 10]	[52.30, 149.70]	[1.04, 1.06]
	Plan B	35			20		
		[30, 35]			[15, 30]		
By Family-Type (Section 5.5)							
B.I Two Plans	Single	47	120.51	1.02	31	132.92	1.03
		[44, 49]	[61.49, 130.60]	[1.01, 1.02]	[19, 33]	[51.68, 146.58]	[1.02, 1.03]
	Non-Single	26			13		
		[25, 27]			[8, 14]		

Notes: Simulation results corresponding to Table 8 (see main text for additional details). "Main Estimates" corresponds to results presented in the main text; "Robustness" presents results from adjusting each household's CARA coefficient to approximately 1×10^{-4} ($\beta_\psi = -2$). Coinsurance rates are in percentages; Δ Surplus is in dollars. 95% confidence intervals are reported below results in brackets, and are obtained by re-estimating the model over 200 bootstrap samples of households and re-computing simulations.