NBER WORKING PAPER SERIES

HOW PRODUCT STANDARDIZATION AFFECTS CHOICE: EVIDENCE FROM THE MASSACHUSETTS HEALTH INSURANCE EXCHANGE

Keith M Marzilli Ericson
Amanda Starc

Working Paper 19527
http://www.nber.org/papers/w19527

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
October 2013

We thank Raj Chetty, David Cutler, David Dranove, Jonathan Gruber, Ben Handel, Larry Katz, Jon Kolstad, David Laibson, Ariel Pakes, Jim Rebitzer, Josh Schwartzstein, and Bob Town for thoughtful comments, along with seminar participants at the American Economic Association annual meeting, the Healthcare Markets Conference, the NBER Insurance Meeting, and the University of Wisconsin. We acknowledge funding from the Lab for Economic Applications and Policy (LEAP) at Harvard University, and Boston University and Wharton internal research funds. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

How Product Standardization Affects Choice: Evidence from the Massachusetts Health Insurance Exchange
Keith M Marzilli Ericson and Amanda Starc
NBER Working Paper No. 19527
October 2013
JEL No. D14,D80,H31,I11,L15

ABSTRACT

Standardization of complex products is touted as improving consumer decisions and intensifying price competition, but evidence on standardization is limited. We examine a natural experiment: the standardization of health insurance plans on the Massachusetts Health Insurance Exchange. Pre-standardization, firms had wide latitude to design plans. A regulatory change then required firms to standardize the cost-sharing parameters of plans and offer seven defined options; plans remained differentiated on network, brand, and price. Standardization led consumers on the HIX to choose more generous health insurance plans and led to substantial shifts in brands' market shares. We decompose the sources of this shift into three effects: price, product availability, and valuation. A discrete choice model shows that standardization changed the weights consumers attach to plan attributes (a valuation effect), increasing the salience of tier. The availability effect explains the bulk of the brand shifts. Standardization increased consumer welfare in our models, but firms captured some of the surplus by reoptimizing premiums. We use hypothetical choice experiments to replicate the effect of standardization and conduct alternative counterfactuals.

Keith M Marzilli Ericson
Boston University School of Management
595 Commonwealth Avenue
Boston, MA 02215
and NBER
kericson@bu.edu

Amanda Starc
The Wharton School
University of Pennsylvania
3641 Locust Walk
Room 303 CPC
Philadelphia, PA 19104
and NBER
astarc@wharton.upenn.edu

1 Introduction

Market efficiency relies on consumers making informed choices. Yet in many contexts, consumers face difficult trade-offs because they may not understand the product itself, may have difficulty comparing different products[1], may be overwhelmed by large choice sets[2], or may be unable to observe important dimensions of product quality. Market makers and regulators often seek to help consumers with choice architecture that simplifies complex problems. For example, policy-makers may require firms to disclose certain types of information about their products (e.g. gas mileage on cars; see Dranove and Jin 2010), may create certification schemes that indicate whether a product has met minimal levels of quality (e.g. Leslie and Jin 2003 on restaurant quality, Houde 2013 on energy efficiency). Relatedly, policy-makers may "nudge" consumers into making different (potentially better) choices, for instance by changing how information is presented, what the default option is, or other aspects of the decision interface (Choi, Laibson, and Madrian 2009). Providing clear, transparent information or simplifying a choice menu is often seen as a benign nudge that can improve market outcomes and consumer surplus.

We examine the effects of product standardization, a policy designed to improve consumer choice and increase competitive pressure on firms. Using insights from the industrial organization literature and behavioral public economics, we show how standardization affects consumer choice, welfare, and firm behavior in a complex market setting. Our context– consumer choice of health insurance plans on the Massachusetts health insurance exchange (HIX)– shares similarities to other markets for complex products, and it is an important market itself: consumer choice of health insurers is a key foundation of the U.S. health care system. Moreover, the insights and methodological tools utilized in this study are applicable to a wide range of policy interventions.

We identify the effects of standardization using a natural experiment: the standardization of health insurance plans on the Massachusetts HIX. Initially, the Massachusetts HIX gave firms wide latitude to design the terms of insurance plans; these plans were then grouped into tiers based on actuarial value (a measure of the plan's overall level of coverage). However, a regulatory change in 2010 standardized the financial cost-sharing characteristics of plans on the exchange, allowing only seven distinct plan types. Firms were still permitted to set prices and differentiate themselves by brand and physician/hospital network. This differentiation

[1]See Kling et al. (2011) on comparison frictions in the Medicare Part D market.

[2]On choice overload in general, see Iyengar and Lepper (2000). In the health insurance context, Hanoch et al. (2009) and Bundorf and Szrek (2010) show experimentally that decision making difficulty grows with choice set size. Finally, Frank and Lamiraud (2009) examine health insurance markets in Switzerland, and find that as the number of choices offered to individuals grow, their willingness to switch plans for a given gain declines.

matters for consumers.[3]

We first show reduced-form evidence that standardization had a substantial effect on brand and characteristics of plan chosen: consumers who enroll just before and just after the change look similar but make different choices. Overall, the generosity of plans chosen rose post-standardization (there was a drop in the share of enrollees choosing bronze tier and high deductible health plans). There were also major shifts in insurers' relative market share. We show that these changes were a result of the policy, rather than confounding factors such as changes in the composition of consumers or seasonal effects. We also show that these shifts did not result merely from changes in relative prices.

Standardization could have affected choice through two major channels: an "availability" and a "valuation" effect. For the availability effect, choices could change because the mix of products available changed and the utility-maximizing choice might have differed between the old and new choice sets. This is easily accommodated in the standard, utility-based model of consumer choice. However, preferences may be context-dependent (Tversky and Simonson 1993). Thus, we allow for a valuation effect: a change in decision weights after standardization (that is, the parameters of the decision utility function that rationalize observed choices change). Note that the decision utility function may be distinct from welfare-relevant utility,[4] and may be affected by seemingly-irrelevant factors (see e.g. Ericson and Starc 2012a on heuristics). Here, standardization may alter decision utility because it shifts consumers' attention (DellaVigna 2009) or changes the salience of product characteristics (e.g. as in Bordalo, Gennaioli, and Shleifer 2012).

To distinguish between the availability and valuation effects, we estimate a discrete choice model in which a consumer's choice is a function of underlying preferences and context-dependent decision weights. Our methodology allows us to identify both utility parameters and the differential effects of changes in choice architecture that resulted from the policy change. We allow decision weights on various insurance attributes (deductible, brand, etc.) to vary pre- and post-standardization. We find that decision weights differ significantly pre- and post-standardization. We then present a more structured model of consumer choice in which the salience of bundles of product characteristics is allowed to vary across regimes (e.g. financial characteristics, brand, etc.).[5] We decompose the total effect of standardization on

[3] Starc (2012) finds that consumers have preferences over the brands in the Medigap market, despite plans having identical financial and network characteristics. We note that in this context, carrier denotes both network breadth and less tangible characteristics like customer service and trustworthiness.

[4] For example, Kahneman, Wakker, and Sarin (1997) distinguish "decision utility"–the function that rationalizes observed choices– from "experienced utility", the hedonic flow from actual consumption. Changes in the choice interface may change the decision weights, but are unlikely to change the hedonic flow from insurance plans.

[5] We present two additional models– an optimization friction model, and a substitutability model– to

market shares into components and find that the valuation effect is larger in magnitude than the availability effect; changes in relative prices play a minor role. We find that standardization increased consumer surplus and overall welfare.[6] But firms reoptimized prices and captured some of the additional surplus and consumer welfare would have been higher if prices had not changed.

In practice, standardization of products often entails two changes: a change in the choice menu (i.e. the options offered), as well as changes to the decision interface (e.g. simplifications enabled by standardization). While both types of changes are relevant for estimating the policy-relevant effect of standardization, it can be valuable to disentangle them in order to make predictions for other contexts. To separate these, we conduct an experiment in which participants make hypothetical insurance choices from menus and choice interfaces similar to the HIX's pre- and post-standardization menu, as well as a new counterfactual condition that separates the changes in product availability from the consumer interface.

Our experimental design replicates the effect of standardization on choice: although baseline choices differ between experiment participants and HIX enrollees, the experiment replicates the major shifts (toward more generous plans and among brands) resulting from standardization in the actual HIX. This supports the validity of our research design on the HIX: standardization itself, rather than a shift in the composition of enrollees, explains the results.[7] We run a third condition to dissociate the effect of the choice menu change from the choice interface: in this treatment, participants see the post-standardization choice menu using the pre-standardization choice interface. Results show that choices would have been different if standardization had not been accompanied by a change to the decision interface. Moreover, we find that standardization induces shifts in the reported importance of plan attributes.

The complexity of insurance plans make HIXs an ideal context to examine standardization. The multiple dimensions of these products are difficult for consumers to evaluate (e.g. coinsurance, copayments, deductibles, maximum out-of-pocket limits). Health insurance markets have particularly high levels of consumer confusion (e.g. see Abaluck and Gruber 2010 on Medicare Part D), and evidence indicates individuals misunderstand impor-

illustrate natural, complementary interpretations of the consumer's decision process.

[6] We use two different welfare criteria, the pre- and post-standardization decision utility indices, as welfare evaluation in the presence of choice inconsistencies is controversial (see Bernheim and Rangel 2011, Beshears et al. 2008). While the two criteria give different magnitudes of the effect of standardization, both agree that standardization increased welfare. For simplicity, we discuss the pre-standardization criterion in the text and leave the post-standardization criterion for an appendix.

[7] These results also support the use of hypothetical choice experiments in health insurance; see also Kesternich, Heiss, McFadden, and Winter (2012); Ericson and Kessler (2013); and Krueger and Kuziemko (2013).

tant aspects of insurance contracts (Handel and Kolstad 2013). While most existing health insurance markets do not have standardized plan types, Medigap (Medicare Supplement Insurance) is an exception. Suggestive evidence from interviews with program administrators indicates that Medigap's standardization reduced consumer confusion (Fox, Snyder, and Rice 2003; also see Rice and Thomas 1992).[8]

Our results provide guidance for the design of HIX, while also adding to the behavioral public finance literature (e.g. on tax salience, see Finkelstein 2010, Chetty, Looney, and Kroft 2012) and the industrial organization literature. Designing HIXs well is important: they are part of a movement toward consumer-driven markets for health insurance, and approximately 20 million consumers will receive coverage via the exchanges under the Patient Protection and Affordable Care Act (ACA). Various states have established HIXs as a result of the ACA; the federal government will run the exchanges for other states. There is a debate among economists and policy-makers over the extent to which HIX regulators should actively shape the offerings in the market, including whether to standardize plans. Understanding consumer demand for insurance, and how it is affected by standardization, is critical in implementing the ACA effectively and is relevant for a number of other insurance markets as well (e.g. employer-sponsored insurance, Medicare Part D).

The paper is organized as follows. Section 2 describes the policy change in Massachusetts and our data sources. Section 3 provides reduced form evidence of the impact of standardization. Section 4 outlines the empirical model and presents the structural estimates. Section 5 conducts counterfactual analyses and welfare analyses, and Section 6 describes the hypothetical choice experiment. Section 7 concludes.

2 The Massachusetts HIX

2.1 History and Existing Literature

The Massachusetts HIX was established by the state's 2006 health reform. The HIX we examine is an unsubsidized health insurance exchange (termed "Commonwealth Choice") for individuals and families with incomes over 300% of the poverty line who were not offered employer-sponsored insurance; a separate, subsidized program serves individuals under 300% of the poverty line. In the time frame we analyze (2009-2010), the exchange had been

[8] A well-identified analysis is difficult, since many regulations changed simultaneously. Relatedly, Finkelstein (2004) finds that the introduction of minimum standards in the Medigap market reduced the fraction of the population holding such insurance. However, we are unaware of any work examining the effect of the Medigap standardization on price competition or consumer choice among brands.

operating for 2 years, and was highly regulated by the Connector Authority.[9] The Massachusetts reform was widely seen as a success: uninsurance rates fell to nearly zero and costs rose no faster than in neighboring states (Kolstad and Kowalski 2012). The HIX played an important role in this reform by providing a marketplace for choosing among a variety of regulated insurance options. Consumers purchasing an exchange plan can choose a plan through an internet portal or by phone; most enroll through the HIX's website. On the website, consumers input demographic information that affects pricing and then are able to compare various plans.

In previous work, we have modeled consumer demand on the Massachusetts HIX prior to standardization (Ericson and Starc 2012a), and examined pricing regulation in the presence of imperfect competition (Ericson and Starc 2012b). In Massachusetts, insurers can only price on age, family size, and location; critically, the oldest consumer in a plan can only be charged twice the rate of the youngest consumer. We show that tighter restrictions on how prices can vary with age lead to higher prices for younger consumers, but increase overall welfare in the presence of an effective mandate. See Ericson and Starc (2013) for more detail on the HIX.

2.2 Plan Standardization

Under the ACA, states have a great deal of latitude in designing exchanges, including plan design. However, throughout its existence, the Massachusetts HIX has taken an active approach. Initially, a number of tiers were defined (bronze, silver, gold) by actuarial value, in a model that was subsequently duplicated by the Affordable Care Act. The Connector Authority required insurers to offer a minimum number of products (six, distributed across tiers) and it awarded a seal of approval only to selected providers (Toolkit 2011). This system evolved over time so that in late 2009 25 distinct plans were offered by 6 insurers: Blue Cross Blue Shield, Neighborhood Health Plan, Tufts, Health New England, Harvard Pilgrim, and Fallon.

Interest in standardizing the plans grew out consumer feedback, as consumers were confused by the existing choice architecture. Board members further saw this as an opportunity to improve choice, stating that "consumers don't have to worry that there's some sort of 'gotcha' in the health insurance purchase. They can know that they are comparing equivalent products and so make better informed decisions based on premium and provider differences".[10] However, board members faced a difficult choice of *how* to standardize the products; the Connector had little research to guide their decision.

[9] Eventually, it will transition to being an ACA exchange, with slightly different regulation.
[10] Nancy Turnbull, quoted in Toolkit (2011).

The initial standardization led to the creation of seven product categories: Gold, Silver-High, Medium and Low, and Bronze-High, Medium, and Low. The plans were initially offered by the same set of six insurers. As a result, while standardization lowered the number of contract designs (financial parameters) used in the market, it actually increased the total number of plans, in the sense of contract design-carrier combinations. Crucially, standardization also unbundled the decision making process into one decision about a contract design, followed by a decision about an insurer. (The standardization process is an ongoing one; the silver medium plan was later eliminated due to low demand, and additional insurers have been added.)

Standardization was a policy change that did two things: it altered both the plans available and the display of information in the marketplace. Pre-standardization, plans were simply listed in ascending premium order. Post-standardization, consumers first choose a standardized financial package and then choose among carriers. This choice process decoupled the choice of financial characteristics from the choice of carrier, potentially leading to different decision weights on carrier and characteristics like deductibles and copayments. Screenshots that show the choice interface both pre- and post-standardization are available in the appendix.

Figure 1 shows the plan designs available before and after standardization for each insurer, focusing on plan's deductible and coinsurance for hospital admissions. (There are additional plan design parameters not displayed, including out-of-pocket maximum, physician copay, etc.) Each marker represents a plan that is available; the size of the marker indicates that plan's relative market share. Note that virtually all the pre-standardization plans were available in similar forms post-standardization; the offset markers in the pre-standardized panel indicates when an insurer's plan design differed slightly from the standardized version. In effect, standardization aligned all the plan options on the grid and filled in the holes in the grid.

Plan standardization can affect choice in a number of ways. First, relative prices may change. Second, the addition of new plans to the choice set may alter choices; we refer to this as the "availability" effect. Third, the relative weights consumers place on plan attributes may be altered by the option menu itself and the information presentation; we refer to this as the "valuation" effect. The pricing and availability effects are easily accommodated in a standard plan choice model, while a model that incorporates alternative decision processes is necessary to capture the valuation effect.

Plan standardization can also be interpreted as lowering search costs– here, the cost of using acquired information to compare plans. Existing literature has examined search costs in markets with homogenous goods (see e.g. Cebul et al. (2011) and Hortascu and

Syverson (2004)). Lowering search costs through improved information disclosure can lead to increased price competition (Sorensen (2000)) or improved quality (Jin and Leslie (2003)).[11] In our context, standardization does not turn plans into homogenous products–they are still differentiated on network/brand– but does reduce differentiation on cost-sharing. Because the prices are listed clearly on the HIX, the search problem is not primarily about finding prices, but about network quality (e.g. "is my doctor covered?") and plan generosity ("Is 20% coinsurance on hospital spending better a $500 copayment per hospital admission?").

2.3 Data

Our dataset is transaction-level data (purchase, cancellation, and payments) from the unsubsidized market (Commonwealth Choice) from the beginning of the Connector's existence in July 2007 until July 2010. We observe approximately 50,000 transactions. There are large spikes in initial enrollment during the first month of the Connector's existence as well as just before the individual mandate took effect in December 2007, with a steady-state enrollment of approximately 1,000 households per month.

Our main analyses focuses on the subset of the data for which we have detailed price information: Nov. 2009-Feb. 2010. Because our data from the Connector is transaction-level data, we do not observe all the plan prices that individuals face. However, for this subsample, we collected an extensive set of price quotes from the Connector website using a Perl script. In this narrow timeframe there are no other major changes in the consumer population or HIX regulation beyond plan standardization. We supplement this data sample with data dating back to Nov. 2008 to examine trends in plan choice, control for seasonality, and to conduct difference-in-difference analyses.

3 Reduced-Form Evidence

3.1 Plans Chosen Pre- and Post-Standardization

The reduced-form evidence shows that standardization had a significant effect on the level of insurance generosity chosen, as all as on the choice of insurer (brand/network). Figure 1 graphically shows each insurance plan's design and market share before and after standardization, and Figure 2 gives precise, aggregate numbers. It is immediately apparent that standardization leads to a large increase in Neighborhood Health Plan's market share (38.7%

[11] In the health insurance setting, Dafny and Dranove (2005) show that health plan report cards do "tell consumers something they don't know" and increase enrollment beyond the role of market-based information, and Jin and Sorensen (2006) find that plan ratings have a meaningful effect on quality of health plan chosen.

to 49.3%), and a large decrease in Fallon's market share (21.1% to 7.6%); both differences are highly statistically significant (p<0.001). Other insurers also saw shifts in market share to a lesser degree.

Turning to the generosity of health plan, standardization increased the overall generosity of health plan chosen. Bronze plans are popular during both time periods, but their market share declines by about 5 percentage points post-standardization (p=0.01); the decline in bronze enrollment is absorbed roughly equally by silver and gold plans. The largest difference is a shift away from high-deductible health plans (HDHPs, a particular sub-type of bronze plans[12]): while 54% of enrollees overall chose HDHPs pre-standardization, only 29% chose HDHPs post-standardization (p<0.001).

Table 1 shows that enrollees chose more generous plans post-standardization. We calculate each plan's actuarial value– the fraction of health care costs that are insured for a representative sample of the population–using the federal government's formula for the ACA exchanges.[13] The mean actuarial value of chosen plans rises from 78% pre-standardization to 83% post-standardization.[14] The corresponding mean out-of-pocket cost declines by $259/year, with a reduction in the standard deviation of out-of-pocket spending as well.

3.2 Difference-in-Difference Estimates and Trends

Our preferred reduced form estimates come from a comparison of the months immediately prior to and following standardization described above. Our enrollee age-mix is similar pre and post-standardization (Table 1) and we find no evidence of trends in choice in the six months prior to standardization (see Table A.3 and Figure A.1.)

We also estimate a difference-in-difference model that directly accounts for any potential seasonality. In this model, we use data from the year prior to standardization (Nov-Dec. 2008 and Jan.-Feb. 2009) as a placebo test. This time period that did not see any major changes to the HIX. Table A.2 (column 3) shows that there were no significant changes in tier chosen for Nov-Dec. 2008 versus Jan.-Feb. 2009, alleviating concerns that seasonality is driving our results. Nov-Dec. 2008 versus Jan.-Feb. 2009 saw only very small changes in brands chosen (less than 5 percentage point shift in market share), and these shifts were in the opposite direction from the shift seen after standardization.

Thus, the formal difference-in-difference estimates in column 2 of Table A.2 are very similar to our main specifications, and in fact show a slightly larger effect of standardization.

[12] We define high-deductible health plans (HDHP) following the tax code as plans with at least a $1200 individual deductible ($1150 in 2009).

[13] Available at http://www.cms.gov/cciio/resources/regulations-and-guidance/index.html

[14] For bronze plans, actuarial value is generally within the 60-70% range. Silver plans range between 70-90%, and gold plans have actuarial values above 90%.

However, our difference-in-difference analysis is limited in the data it can examine: because we do not have detailed plan design data for the period prior to Nov. 2009, we cannot include actuarial value, HDHP status, or out-of-pocket costs in our model for this period. However, the similarity for the attributes we can observe indicates that our primary estimates do in fact identify the effect of standardization.

3.3 Hedonic Pricing

To examine the effect of the standardization on prices, we present a series of hedonic regressions in Table 2. Unadjusted for generosity, plans are slightly cheaper in 2010 (column 1). However, column 2 shows that this is largely because in 2010, there were more plans available in the lower tiers (bronze and silver) than in 2009. If anything, plans became slightly more expensive, adjusted for financial generosity. Columns 3 and 4 show that this effect is not uniform across plan types. Less generous tiers become *relatively* more expensive, while the cost savings associated with choosing a HDHP, controlling for actuarial value, are larger in 2010. The observed changes in relative prices predict shifts in choice: for example, we should expect more consumers to choose gold plans and HDHPs in 2010 since they became relatively cheaper. Since we in fact see a shift away from HDHPs our consumer choice model must be flexible enough to allow for other factors to affect choice: the choice menu available and the decision process itself.

4 Standardization and Consumer Choice

4.1 Shifting Decision Weights: Results

We model consumer choice as a discrete choice problem. First and most fundamentally, products are differentiated based on network and brand. Second, we are interested in modeling consumer choice of plans, and so infer valuation of insurance attributes from those choices.[15] In our model, consumers face a discrete choice from a set of plans, and attach decision weights to various plan characteristics. In our baseline model, we assume that consumer i's decision utility of plan j is given by $u_{ij} = \delta_j + \mu_{ij} + \varepsilon_{ij}$, with the mean value of a plan denoted by δ_j and the individual-specific component denoted by μ_{ij}. Given the

[15] The alternative method of forming consumers' subjective expected distribution of out-of-pocket costs under each plan, and evaluating plans using a risk-averse utility function for money, is not feasible in our application. First, we do not know consumers' subjective expected distribution of costs, since beliefs may not match actual claims. Second, evidence shows that consumers do not value insurance plans according to the standard expected utility model (Abaluck and Gruber 2009, Barseghyan et al. forthcoming). Third, claims data are unavailable for this population. Fourth, plans would still be differentiated based on network.

assumption that the error term ε_{ij} is i.i.d. extreme value, the probability that consumer i purchases product j can be written as:

$$s_{ij} = \frac{\exp\left(\delta_j + \mu_{ij}\right)}{1 + \Sigma_j \exp\left(\delta_j + \mu_{ij}\right)}$$

While a standard model of consumer choice can account for both the availability effect and the effect of price changes, we also seek to explore how consumer valuations differ pre- and post-standardization. To capture the valuation effect, we allow the decision utility index to depend on both primitive preferences and context: changes in context can alter how an attribute is valued.[16]

In Table 3, we separately estimate decision-weights pre- and post-standardization. Utility is given by $u_{ij} = \mathbf{X}_{ij}\left(\boldsymbol{\beta} + \mathbf{1}_{Post}\boldsymbol{\Delta}_{\beta}\right) + \left(\alpha + 1_{Post}\Delta_{\alpha}\right) p_{ij} + \varepsilon_{ij}$, where $\boldsymbol{\beta}$ and α are pre-standardization decision weights, $\mathbf{1}_{Post}$ is a post-standardization indicator, and Δ is the additive shift in each decision weight post-standardization. To cleanly identify price sensitivity α, we follow a similar identification strategy to that used in our previous work (Ericson and Starc 2012b). Becuase prices set by firms jump discontinuously at round numbered ages (e.g. 30, 35, 40, etc.), while preferences evolve continuously with age, α can be identified by allowing preferences to change continuously with age. To implement this by allowing for an age-trend in price-sensitivity. For a detailed discussion on the validity of this strategy and its robustness to a variety of specifications, see Ericson and Starc (2012b).[17]

Column 1 of Table 3 presents a conditional logit specification that allows for heterogeneity in α based on age, but no further heterogeneity in decision weights. Then, in column 2, we present a mixed logit specification that allows α to take on a log-normal distribution, so that demographically-identical individuals in a given year vary in how much weight they put on premiums. In the mixed logit specification, the estimated standard deviation on the premium coefficient is substantial and statistically significant.

Both specifications show little evidence that standardization increased price sensitivity. Column 1 shows that the premium coefficient α indeed becomes more negative (more price sensitive) post-standardization, a difference of about 4 percent; this difference is not statistically significant. Column 2 estimates a distribution of α separately pre- and post- standard-

[16] While we use the language of decision utility indices and decision weights, the idea is related to that of Bernheim and Rangel's (2010) "ancillary conditions" which, although they do not change the characteristics of a product, can still affect choice.

[17] Only relative changes in decision weights are identified by comparing decisions in two contexts. We thus need to normalize the utility of one of the plans (or the outside option, in the absence of an effective mandate) to be zero in both years. We normalize the utility of the gold BCBS plan to be zero and unchanged across years. This plan was chosen both because it is contractually identical and has similar market share before and after the policy intervention.

ization. The mean values of α pre- and post-standardization are -0.043 and -0.040, respectively; the mean individual is slightly less price sensitive post-standardization. (Similarly, the medians are -0.046 and -0.042, indicating slightly less price sensitivity post-standardization). One can also see relative changes in the valuation of tiers (bronze is the comparison category): HDHPs receive more negative weight post-standardization, and the perceived differences between silver/gold plans and bronze plans increase post-standardization, even holding constant actuarial value. Table A.4 shows that a similar pattern results even if actuarial value is not included.

4.2 Context and Salience Model

In order to interpret and generalize our findings, we add additional structure. First, in our salience model, we allow standardization to change the relative salience of bundles of attributes (e.g. estimating whether tier becomes more or less important relative to brand). This change in salience can result from changes in attention (DellaVigna 2009) due to how standardization changed the decision process and information presentation. Changes in salience can also arise from simple shifts in the option menu (see Bordalo, Gennaioli, and Shleifer 2012 on how the option set itself can alter the salience of characteristics), or can result because consumers perceive implicit recommendations contained in the design of choice set (contextual inference; see Kamenica 2008).

In this salience model, we capture the change in salience by letting a parameter σ multiply bundles of attributes. Thus, if β represents the primitive preference for characteristic x, $\beta\sigma_{a,x}$ represents the decision weight under a particular context a. (This keeps the decision weights more closely linked to the underlying utility of the product.) We estimate three different values of σ post-standardization for different bundles: $\sigma_{Post,f}$ for financial characteristics (tier and HDHP), $\sigma_{Post,b}$ for brands, and $\sigma_{Post,p}$ for premium. This contrasts to the approach in Table 3, which simply estimates different decision weights before and after standardization.

We can identify how standardization *changes* decision weights from one context to another, but of course cannot identify context-free fundamental utility parameters.[18] Formally, let δ_{jPre} be the pre-standardization decision index of a plan with characteristics $\mathbf{X}_j$ and δ_{jPost} be the post-standardization decision index of a plan with the same characteristics Our discrete choice model measures $\delta_{jPre} = \mathbf{X}_j\boldsymbol{\beta}\boldsymbol{\sigma}_{Pre}$ and $\delta_{jPost} = \mathbf{X}_j\boldsymbol{\beta}\boldsymbol{\sigma}_{Post}$, but we cannot

[18] In a model in which context affects the decision weight placed on product characteristics, underlying utility is not easily identifed (or even defined). Is a product popular because it contains a bundle of popular characteristics or because its good characteristics are particularly salient to consumers in the context? The literature has taken varied approaches to this problem. While some papers rely of documenting dominated decisions or modelling things that directly affect utility, like switching costs (Handel 2013), other papers rely on restrictions from theory (Abaluck and Gruber 2012).

separately identify the vectors $\boldsymbol{\beta}, \boldsymbol{\sigma}_{Post}$, and $\boldsymbol{\sigma}_{Pre}$. Hence, we normalize $\boldsymbol{\sigma}_{Pre} = \mathbf{1}$, and identify $\boldsymbol{\beta}$ (the valuation of characteristics pre-standardization) and $\boldsymbol{\sigma}_{Post}$ (the changes in valuation post-standardization). When $\sigma_{Post,x} > 1$, the decision weight placed on that characteristic increased post-standardization, and when $\sigma_{Post,x} < 1$, the weight decreased.

In this model, the decision index for the conditional logit in 2009 is written as:

$$v_{ijt} = \mathbf{B}_{jt}\boldsymbol{\beta}_b + \mathbf{F}_{jt}\boldsymbol{\beta}_f + \alpha_i p_{ijt} + \varepsilon_{ijt}$$

where $\mathbf{B}_{jt}$ is a matrix of brand dummies and $\mathbf{F}_{jt}$ is a matrix with silver, gold, and HDHP dummies. The 2010 decision index is then given by:

$$v_{ijt} = \mathbf{B}_{jt}\boldsymbol{\beta}_b\sigma_{Post,b} + \mathbf{F}_{jt}\boldsymbol{\beta}_f\sigma_{Post,f} + \alpha_i p_{ijt}\sigma_{Post,p} + \varepsilon_{ijt}.$$

Identifying $\boldsymbol{\sigma}_{Post}$ requires some additional assumptions, which are likely to be valid in our context. There cannot be differential measurement error across the two years of data. The characteristics we measure (such as brand and metal dummies) cannot have increased or decreased in value: bronze plans must be equally generous across the two regimes. This is likely to be satisfied: the post-standardization plans were modelled after pre-standardization plans, and we include the plans' actuarial values to control for any additional differences. Also, there can be no heteroskedasticity that requires rescaling of the coefficients (see Train 2003). In our example, this assumption is likely to be satisfied: in our optimization frictions model below, we specifically model the heteroskedasticity by allowing the variance of the idiosyncratic error term to change post-standardization. We find only a slight change in the variance of the error term.

4.3 Optimization and Substitutability Models

In addition to our main salience model, we estimate an optimization friction model and a substitutability (nested logit) model. All of these models build on a discrete choice framework and have similar practical interpretations despite the slightly different conceptual interpretations. The optimization frictions model is consistent with reduced search or optimization costs, while the substitutability model is consistent with reduced comparison frictions (Kling et al 2009).

Our optimization frictions model allows context to affect the degree of "optimization frictions" facing consumers. Here, we do not interpret the error term ε_{ijt} as a primitive preference unobserved to the econometrician, but rather as a friction (error) impacting the decision index. Then we write the 2009 decision index as above and the 2010 decision index as:

$$v_{ijt} = \mathbf{B}_{jt}\boldsymbol{\beta}_b\sigma_{Post,b} + \mathbf{F}_{jt}\boldsymbol{\beta}_f\sigma_{Post,f} + \alpha_i p_{ijt} + \sigma_{Post,\varepsilon}\varepsilon_{ijt},$$

where $\sigma_{Post,\varepsilon}$ allows the variance of the error term to vary post-standardization. It is straightforward to see that the optimization frictions model is just a different normalization than the salience model: rather than normalizing the variance of the error terms across years, we normalize the price coefficient across years. Furthermore, the economic content of these normalizations are complementary: in both, the relative importance of financial and network characteristics is allowed to vary across the two regimes.

In the final type of structure, we estimate a nested logit model, allowing for the substitutability of products to vary pre- and post-standardization. After standardization, as shown by screenshots in the appendix, the choice process is structured such that consumers first choose a tier and then choose an insurer brand, making a nested logit model a natural choice. This nested logit model will allow us to specify a correlated error structure within pre-specified nests of bronze, silver, and gold plans. In the nested logit model, the deterministic part of utility is still given by $u_{ij} = \delta_j + \mu_{ij}$, but u_j is decomposed into X_j,the part that varies within a tier, and Z_s, the part that is constant within a tier. Tiers–bronze, silver, or gold–are indexed by s. We estimate the nested logit plan model in two steps: first, we estimate the probability of choosing a plan, conditional on the tier chosen. In the second step, we estimate the probability of choosing a given tier.[19]

We denote the dissimilarity parameter (an inverse measure of the correlation of the error terms within a nest) by ρ. This parameter should change if product standardization leads consumers to view products as closer substitutes and strengthens the degree of price competition as a result. We thus allow the nesting parameters to vary pre and post-standardization, and so estimate both ρ_{Pre} and ρ_{Post}. In the model, product standardization may lead consumers to view products as closer substitutes and strengthen the degree of price competition as a result.

Econometrically, the dissimilarity parameter ρ can play a similar role to the salience parameters described above if it is allowed to vary across years. To see that, note that the dissimilarity parameter performs two distinct functions in estimation. First, it determines the correlation coefficient between error terms within a tier: the closer ρ is to zero, the higher the correlation. In addition, the dissimilarity parameter affects the relative weighting of characteristics within the nest X_{ij}, relative to characteristics outside of the nest Z_s: the

[19] In the nested logit model, the probability of choosing a plan j conditional on choosing a plan within tier s is given by $s_{ij} = \frac{\exp\left(\rho^{-1}(X_{ij}\beta)\right)}{\Sigma_{l\in s}\exp(\rho^{-1}(X_{il}\beta))}$,and the probability of choosing a given tier is given by $s_{is} = \frac{\exp(Z_s\alpha)\exp\left(X_{ij}\widehat{\beta/\rho_s}\right)^{\rho}}{\Sigma_s \exp(Z_s\alpha)\exp\left(X_{ij}\widehat{\beta/\rho_s}\right)^{\rho}}$.

closer ρ is to zero, the higher the relative weight on the characteristics Z_s. For a ρ of zero, only the characteristics in the outer nest are salient to consumers. The larger the ratio of ρ_{Pre}/ρ_{Post}, the greater is $\sigma_{Post,f}$ (multiplying tier) in the previous models.

4.4 Discussion

A few modelling choices deserve additional discussion. First is the impact of consumer heterogeneity. We estimate mixed logit specifications in both the decision weight and structured models. However, because it is often difficult to identify consumer heterogeneity in preferences separately from variation in decision processes (Ericson and Starc 2012a), we also estimate a nested logit specification to estimate flexible patterns of substitution between plans in our substitutability model.

Our focus is on consumer decision-making, rather than firm behavior. This is largely driven by the institutional environment and the nature of the policy change. For example, we take policies as given because the plan parameters are heavily regulated both pre- and post-standardization, and set entirely by the HIX post-standardization. By contrast, firms were allowed to reprice their policies. We do not need to assume that the post-standardization prices are equilibrium prices, as firms are likely learning about demand. However, we do incorporate the observed change in prices into our counterfactual simulations and welfare calculations. In addition, this paper focuses only on first-time choosers with no history in the HIX. We are interested in how standardization affects consumer choice; individuals who are inertial and do not make a new choice are, in effect, not exposed to standardization. Also, the vast majority of enrollees in this time period (and all the enrollees in our analysis sample) are making a first-time choice, so this analysis is a good guide to firm incentives as well. (See Ericson and Starc 2012b for an extended discussion of inertial consumers.)

One limitation is that our dataset does not contain information on consumer costs, and we cannot examine the impact of standardization on (adverse) selection across plans.[20] Since our primary analyses focus on consumer choice, rather than firm pricing, cost information and/or adverse selection would not alter our results. Change in selection would affect firm pricing strategies, and we cannot model that change. However, note that the policy change was not accompanied by a large price increase in more generous plans, as would be predicted by an adverse selection death spiral; if anything, the relative premium for more generous plans fell.

[20] As highlighted by Handel (2013), changes in selection could have potentially serious consequences in markets where consumers are susceptible to behavioral biases.

4.5 Results

Table 4 presents our estimation of the optimization friction, salience, and nested logit models in which the multiplicative change in decision weight σ is constrained to be the same within groups of characteristics: for brand σ_b, for financial cost-sharing characteristics σ_f, for premiums σ_p, and/or for the error term σ_ε. Each model strongly rejects the hypothesis that the decision utility is constant across years. The first column in Panel A allows for a different variance of the idiosyncratic error term post-standardization, as well as a change in the weight placed on financial characteristics σ_f and brand σ_b.[21] The results show that the variance of ε_{ij} is slightly lower post-standardization ($\sigma_\varepsilon = 0.96$), consistent with choices being less "noisy" post-standardization. In this model, we estimate that the tier/HDHP financial characteristics become much more important post standardization ($\sigma_f = 3.39$), while brand becomes only slightly less important ($\sigma_b = 0.98$).

The second and third columns of coefficients in Panel A estimate the salience model using a conditional logit model (Salience Model 1) and a mixed logit model that allows for unobserved consumer heterogeneity in premium sensitivity (Salience Model 2). Both models show a very similar pattern: σ_p is about 1, indicating that standardization did not increase (or decrease) price sensitivity, and the σ_b coefficient on brand is again slightly lower than one, indicating less weight placed on brand post-standardization. Finally, both specifications show that financial characteristics become much more important post-standardization ($\sigma_f > 2$).

Finally, Panel B presents the nested logit specification. The dissimilarity parameter in 2009 in much higher than the dissimilarity parameter in 2010 (0.28 vs. 0.14). This indicates a higher correlation of errors within a tier post-standardization and a much higher weight on tier post-standardization, consistent with the results in Panel A. In both years, plans within a nest are very close substitutes. The ratio of the dissimilarity parameters is closely related to the tier salience parameter; the model implies a σ_{tier} of 1.96, which is similar to that estimated in the optimization friction and salience models. In addition, the nested model implies a higher degree of substitutability between plans within a tier post-standardization. We argue that standardization led to a shift in the valuation of plan attributions, largely because certain plan characteristics, such as tier, were more heavily emphasized and became more salient to consumers.

[21] We hold salience of premiums fixed across the two years, as the model needs one normalization.

5 Counterfactuals and Welfare

5.1 Simulations

Given estimates of primitive preferences and the change in the salience (or the valuation effect), we now turn to counterfactual exercises that allow us to decompose the changes in market shares into availability and valuation effects. We also estimate the welfare impact of standardization. We first run a counterfactual experiment that attempts to disentangle the supply- and demand- side forces that explain changes in market shares across plans. Using the Salience-1 model estimated in Table 4, we simulate the percentage of consumers choosing bronze plans pre- and post-standardization, using both old and new hedonic price vectors. To be precise, in the first two columns, we simulate choices using the pre-standardization decision weights ($\boldsymbol{\beta}$) and choice set, under either the 2010 or 2009 hedonic price vectors. In the second two columns, we use the post-standardization decision weights ($\boldsymbol{\beta}\boldsymbol{\sigma}_{Post}$) and choice set, again under each hedonic price vector.

The decision to model counterfactual prices using the hedonic model relies on the fact that the hedonic model predicts prices extremely well, with a R^2 of over .9 in nearly every specification. Hedonic prices provide the best representation of the idea of "holding prices fixed" at the 2009 levels. The use of hedonic prices allows us to capture the key features of the market with minimal additional assumptions.[22]

Figure 3 reports results of the counterfactuals, plotting the percent of enrollees in each brand or tier by year. The upper left panel shows actual 2009 and 2010 choices. The lower left panel shows availability effect: the effect of shifting the choice set, holding constant decision weights and prices at their 2009 levels. This choice set shift is what leads to a large increase in Neighborhood's market share, and decreases in Fallon's share and the market share of HDHP plans. The upper right panel shows the valuation effect: shifting decision weights (holding constant prices at 2009 levels and using the 2010 choice set) leads to the increase in generosity of plan chosen: enrollees shift away from HDHP and bronze plans. Finally, the lower right panel shows the effect of shifting prices. The result of the price shift actually counteracts the availability and valuation effects: changes in prices alone would have led to a decline in Neighborhood's market share, and in increase in HDHP and bronze plans.[23]

Decomposing the valuation and availability effects is important: our simulations in which we shift thechoice set alone (availability effect) also show that more enrollees would have

[22] Also, as detailed in Ericson and Starc (2013b), an equilibrium model of pricing requires incorporating modified community rating regulation and is outside the scope of this paper.

[23] Appendix Table A.6 provides more detailed counterfactuals in numerical form.

chosen bronze plans. In reality, the fraction choosing bronze plans fell. The simulations in which we shift decision weights alone show that the valuation effect is the source of this decline in bronze plans chosen. (Recall, standardization increased the weight enrollees attached to tier). Similarly, product availability disproportionately affects Neighborhood's market share. Our simulations predict that the market share of Neighborhood plans would be an extremely high 72% in 2010 if individuals had still used pre-standardization decision weights and faced pre-standardization prices. A large part of this availability effect is due to the existence of the relatively inexpensive, low deductible "Bronze High" plan. However, the valuation effect's increased weight on tier counteracts this availability effect, reducing Neighborhood's market share to 48% with post-standardization decision weights.

To summarize: the reduction in the market shares of bronze plans is largely due to valuation, rather than availability or supply-side factors. The reduction in HDHPs is due to both availability and valuation effects. The large increase in Neighborhood's market share is largely due to the availability effect. Demand-side factors and the change in plan offerings due to regulation, rather than firm pricing, are largely responsible for the shifts in choices we see.

5.2 The Effect of Standardization on Welfare

In order to calculate standardization's effect on welfare, we make two additional choices. First, we choose to calculate welfare under a revealed preference welfare criterion that utilized the post-standardization decision utility index. This welfare criterion reflects the choices consumers make when regulators feel they are better informed about their options.[24] Second, we have to accommodate the expansion in the number of plans in 2010. Recall that there are actually more plans post-standardization than pre-standardization. Our welfare analyses hold the number of plans fixed, choosing the 18 most popular plans in 2010 to compare to the 18 plans in 2009. Thus, we potentially underestimate the positive effects of standardization on welfare by reducing some of the availability effect in addition to correcting for the additional error draws from the logit model.

We present our estimates of the change in consumer welfare in Table 5 using our salience model and nested logit model.[25] For each welfare criterion, we measure equivalent variation

[24] Following the welfare framework of Bernheim and Rangel (2012), we note that one cannot rank the inconsistent choices. Fortunately, these different welfare criteria give similar results, though of slightly different magnitudes.

[25] We don't present welfare estimates based on the optimization friction model, as the ε terms can be interpreted either as errors or preference shocks. Under an error interpretation, making fewer errors clearly improves welfare.

using the standard formula of Nevo (2001) and McFadden (1999), which is given by:

$$EV_w = (1/\alpha_w)\left[\log\left(\Sigma_{j\in\{post\}}\exp\left(\delta_{wj}\right)\right) - \log\left(\Sigma_{j\in\{post\}}\exp\left(\delta_{wj}\right)\right)\right]$$

where $\delta_j = \alpha_w p_j + X_j\beta_w$ is the estimated mean utility of plan j, which can be decomposed into the disutility of price $\alpha_w p_j$ and the positive utility of plan characteristics (or fixed effects) $X_j\beta_w$. Note that the coefficients have subscripts w, which reflect the welfare-relavant decision utility index based on the post-standardization choices.

When consumers make choices according to a decision utility function that differs from a welfare-relevant utility function, the formula for utility needs to be adjusted. That is, we may want to evaluate the impact of 2009 choice set under the 2010 decision utility function. This requires a simple offset term to the formula for utility in the logit model. In order to do this, we assume that the idiosyncratic component of utility is held fixed and integrate over this component. This assumes that the choice is made based on the decision utility index and is represented by the first, traditional term in the calculation. Then, we subtract off the expected, deterministic component of utility, represented by the first share-weighted average, conditional on the decision utility index. We then add in the expected, deterministic component of utility conditional on the welfare criterion, noting that the i.i.d. error term as already been accounted for in the first term. Letting w be the welfare criterion and d be the decision criterion, we then have:

$$EU = (1/\alpha_w)\log\left(\Sigma_j \exp\left(\delta_{wj}\right)\right) - \left[\Sigma_j s_{jw} X_{ij}\frac{\beta_w}{\alpha_w} + \Sigma_j s_{jw} X_{ij}\frac{\beta_d}{\alpha_d}\right].$$

There are a number of components of each simulation: a decision utility index, a welfare criterion, a price vector, and a choice set. In each cell of Table 5, we compute the equivalent variation associated with facing the pre-, rather than post-standardization choice set. In the first row, decision utility indices and price vectors are allowed to vary across years. In the second row, the post-standardization decision utility function is held fixed. This implies that by construction, the compensating variation is higher (pre-standardization utility is lower), as the offset term is always negative because consumers are not fully optimizing. In the third row, the price vector is fixed at the 2009 level.

The welfare results across our different models are qualitatively and quantitatively similar. Row 1 presents our estimate of the total effect of standardization on welfare, which includes a shift in menu, premiums, and decision utility index. Using either welfare criterion, we find that standardization improved welfare by 6-7% percent of premiums (or $23-26/enrollee per month), depending on the exact specification. When the 2010 choice set is expanded to the full set in the second panel, we see that the effect grows to \$41-\$46/enrollee per month.

We then consider the welfare gain in the absence of a shift in the decision utility index (i.e. assuming that individuals would always optimize using post-standardization decision weights). Row 2 shows that the welfare change is slightly larger holding the decision utility index constant, indicating that standardization had a positive impact on consumer surplus due to both the availability and valuation effects. If the valuation effect was zero, row 2 would be equal to row 1.

We then examine the effect of prices for plans, which differed pre and post standardization. To conduct counterfactual simulations that hold prices constant across years, we use a hedonic pricing model. Thus, in row 3 we evaluate welfare under a standardization event that (counterfactually) held prices constant at pre-standardization levels– that is, a simulation in which, were a plan to be offered in the identical form pre- and post-standardization, it would have the same price. The simulation, therefore, does allow for changes in prices that result from changes in plan generosity. Thus, we use the EV_w formula above, but plug in the counterfactual prices at the pre-standardization level: $\hat{p}_j^{pre}$. Here, we find that standardization would have increased welfare even more if price levels had remained constant.

The standardization policy change improved welfare. This is due to both the valuation and availability effects, with the bulk of the welfare gains coming from the availability effect. Standardization both helped consumers trade-off various product characteristics in a constructive way and made high value, low deductible, low premium, restrictive network plans - such as the Neighborhood "Bronze High" plan - available to consumers. The results of the constant pricing simulations imply that changes in firms' premiums capture part of the surplus that results from standardization: the effect of standardization is smaller than it would have been absent a supply-side effect. Nudges that affect consumer behavior are also likely to affect equilibrium outcomes as firms respond to changing consumer demand.

5.3 Discussion

We estimate that standardization improves outcomes for consumers and that firms can capture some of the newly generated surplus. Why didn't firms offer such an assortment of plans initially? At least part of the explanation is that standardization introduced additional choice while providing additional decision support tools that allowed consumers to express their preferences. Another potential explanation is that firms were still learning and did not know this deviation would be profitable: the market is relatively new (approximately 4 years old at the end of our sample period) and that may not have been enough time for firms to learn about both costs and demand. Since the HIX may be a relatively small proportion of insurers' books of business, they may not have a huge incentive to perfect their offerings in this particular market. Finally, selection could have led to the product assortment in the pre-standardization period. A single firm introducing one of the new standardized plans might have attracted a relatively high cost subset of the population, making deviations from a pre-standardization equilibrium unprofitable.

The policy change was not without risk. If the insurers had been forced to offer only one plan at each level - bronze, silver, and gold - the consumers would have been made worse off under the 2010 prices regardless of their preferences and the exact array of plans. (Having a bronze plans with and without a high deductible is critical.) The regulator redesigned the menu in a way that provided valuable additional choice to consumers. However, this type of policy intervention requires caution: without expanding the choice set and providing a high-value option to consumers, the policy change could have reduced welfare.

6 Experiment

The standardization on the Massachusetts HIX involved two changes. First–and most importantly– the choice set changed. Second, the choice interface changed. Recall that post-standardization, plans within the same sub-tier had identical financial characteristics– this is the change in the choice menu. However, this change also enabled a change in the choice interface: instead of choosing a plan from the list of plans available[26], post-standardization enrollees first chose a tier of insurance generosity, and then chose an insurer. In addition, slightly different information was displayed pre versus post-standardization.

We conduct an experiment to examine the extent to which standardization had an effect through a) the change in choice menu versus b) the change in choice interface. The experiment disassociates these two changes. We assign participants to one of three conditions: The "Pre-Stdz." condition replicated the HIX's pre-standardization choice menu and interface, while the "Post-Stdz." condition replicated the HIX's post-standardization choice menu and interface. The third condition, "Alt-Post." has exactly the same plans as in the "Post-Standardization" condition, but uses the pre-standardization decision interface (plans are presented in a list, and characteristics of plans were presented as they were in the pre-standardization interface). Comparing Pre-Stdz. to Post-Stdz. choices allows us to establish the validity our experimental design (and the validity of our analysis of the HIX data). Comparing Post-Stdz. choices to choices in the counterfactual Alt-Post condition

[26] On the HIX pre-standardization, participants had the option to filter this list to just "tier" (e.g. just look at the bronze, silver, and/or gold policies), but the characteristics of each tier were not described at the filtering stage. There was no ability to filter more narrowly. See the Online Appendix for details.

allows us to examine the extent to whether the observed shifts in choice or due to the menu or the interface.

We recruited participants from an online panel (run by the firm Qualtrics) who roughly matched the demographics of individuals purchasing insurance on the HIX: they lived in one of these northeastern states (ME, VT, NH, MA, CT, RI, and NY), and had relatively high household incomes ($35k+ for an individual or $65k+ for a family of four). Participants answered some demographic questions. They were then assigned to a condition, and asked to pick the insurance plan they preferred. This is our primary variable of interest. After making their choice, participants were asked to rate the salience of various plan characteristics. They were then shown another choice menu, and asked to make a second choice, and then asked to rate the salience of various plan characteristics in this second menu.

We first examine the reduced form effect of the various conditions. Our hypothesis of interest is not about the levels chosen in our experiment, but in differences between conditions. Examining the actual choices on the HIX (Table 1), we make predictions for the comparison of choices in the Pre-Stdz and Post-Stdz conditions. Although there are many differences between observed choices in 2009 and 2010, we focus our hypotheses on the three largest effect sizes (>10 percentage point differences) seen in the actual HIX data. Our hypotheses are that standardization should:

H1: Reduce the fraction of participants choosing high deductible health plans (HDHP)

H2: Increase the market share of Neighborhood Health Plan

H3: Decrease the market share of Fallon

We have three additional weaker hypotheses (shifts in choice between 5 and 10 percentage points): standardization should decrease the fraction of bronze plans chosen, increase the fraction silver plans chosen, and increase the market share of Tufts Health Plan.

Table 6 shows the summary statistics for the experiment, by condition. First, note that experimental participants choose more generous plans than observed in the actual HIX. There are many potential explanations for this, including selection into the exchange; Ericson and Starc (2012a) show that plans chosen on the HIX are less generous than observed in employer-sponsored insurance. The distribution of brand choices is similar between the actual data and the observed data, with the biggest exception is that Tufts is relatively more popular among the experimental participants. (Note that we intentionally chose a geographic region in which the smallest insurer, Health New England, was not offered.)

The treatment effects in the experiment verify all three predictions, even though the baseline levels of choice differ between the experiment and the actual data. In the Post-Stdz.

condition, the fraction choosing HDHP drops by 16 percentage points, the market share of Neighborhood Health Plan increases by 17 percentage points, and the market share of Fallon drops by 4 percentage points. (All these differences are significant with $p < 0.01$.) Similarly, we find small directional support (though statistically insignificant) for a decrease in bronze and an increase silver plans. The only shift we do not replicate was the market share of Tufts Health Plan: experimental participants were slightly less likely to choose Tufts in the Post-Stdz. condition, while HIX enrollees were slightly more likely to choose Tufts post-standardization; this may be an artefact of the high rate of preference for Tufts among experimental participants. Appendix Table A.7 verifies these results using a regression framework; controlling for demographics alters point estimates of differences only slightly, but improves precision. Appendix Table A.8 runs conditional and mixed logit choice models on the experimental data–analogous to 3. It finds many similar shifts in decision weights: an increase post-standardization in valuation of the gold tier (relative to bronze) and the disutility from HDHP plans. However, we do not find a significant age trend in premiums, and valuation of silver tiered plans increases post-standardization only in the mixed logit specification. Finally, we find an increase in price sensitivity post-standardization in the conditional logits, larger than that found in the actual HIX data.

These results show that hypothetical choice experiments can approximately replicate actual behavior, and add to a growing literature validating such experiments in the health insurance context (Kuzeimko and Krueger 2013, Kesternich et al. 2013, Ericson and Kessler 2013). The experiment's results confirms the validity of our design analyzing the actual HIX data, providing evidence that observed shift in choices was due to standardization rather than some other factor (e.g. a shift in enrollee composition).

The counterfactual condition "Alt.-Post" uses the post-standardization menu with the pre-standardization choice interface. There are only small differences in the brands chosen, comparing this condition to the Post-Stdz. condition. However, the alternative interface leads experimental participants to make more extreme choices than in the Post-Stdz. condition: Alt.-Post participants are both more likely to choose a gold plan and more likely to choose a HDHP plan than Post-Stdz. participants. This is consistent with the post-standardization interface enabling consumers to differentiate among plans in a more accurate way; it can be difficult to differentiate among plans in a long list, and individuals may gravitate toward one end or another. Note that the change in interface is complementary to the change in choice menu, as the post-standardization interface simplifications would not have been possible without the concurrent change in the choice menu.

After participants made their choice in from their assigned menu, we asked them to rate "how important" various factors were in making their choice on a scale of 1-7 (not at all important to extremely important). Table 7 gives means by condition (Appendix Table A.8 shows that results are unchanged controlling for covariates). The most important category is, unsurprisingly, premium with a rating of about 6.0, with the following categories close behind (5.4 to 5.8): cost of hospital stay, cost of a doctor's visit, deductible and "maximum out of pocket expense". Tier was rated the least important dimension for all three conditions– while it may have been useful in organizing information, individuals seemed instead to rely on the financial characteristics of plans.

These importance ratings were affected by condition: standardization increased the importance of tiers relative to the other characteristics of plans. The regression point estimates indicate that the measured importance of every other listed attribute declined, except brand. However, these results show that the increase in the importance of tier came primarily from the interface redesign, rather than the choice menu. The Post-Alt condition did not show any significant change in the importance of tier, as compared to the Pre-Stdz. condition. This suggests that theories of salience that only rely on the attributes of choice (rather than how they are presented) miss important elements of salience.

Two additional factors were related as less important in the Post-Stdz. condition, as compared to Pre-Stdz.: cost of hospital stay and maximum out of pocket expense. Both were surprising: ex ante, hospital stay seems equally prominent in both conditions. Moreover, only in the Post-Stdz. condition was information about maximum annual out-of-pocket cost directly listed. One hypothesis is that participants interpreted "maximum annual out of pocket expense" as referring to their subjective assessment of the total risk they would face in the plan, and that in the Post-Stdz. condition they relied more on tier instead. Finally, neither brand nor tier varied in importance across the three conditions. This result is consistent with our discrete choice models estimated on the actual HIX data, which did not find an increase in price sensitivity post standardization.

7 Conclusion

The HIX standardization provides an opportunity to see how change in choice architecture affected choice and welfare in a policy-relevant setting. We have provided a flexible template for other researchers examining how regulation and choice architecture affect markets in a model that can capture availability, valuation, and pricing effects. Such changes can alter the decision weights consumers attached to multiple product characteristics. In this market, consumers trade-off financial generosity, network breadth, and premiums. Standardization emphasized the financial, rather than network, characteristics of plans. As a result, post-standardization consumers chose more financially generous plans. Our experimental results confirm and extend our analysis of the HIX data. We show that product standardization allows consumers to more accurately differentiate between plans and that changes in the choice set were complementary to changes in the information interface.

Our study highlights the potential for regulators to help consumers when making choices from complicated set of products. Simple shifts can have large impacts, shifting choices and improving (or reducing) consumer welfare.

References

Abaluck, J. (2011). What Would We Eat if We Knew More: The Implications of a Large-scale Change in Nutrition Labeling. *Working Paper, Mass. Inst. Technol.*

Abaluck, J. T. and J. Gruber (2011). Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part D Program. *American Economic Review 101*(4), 1180–1210.

Barseghyan, Levon, Francesca Molinari, Ted O'Donoghue, and Joshua C. Teitelbaum. (Forthcoming). The Nature of Risk Preferences: Evidence from Insurance Choices. *American Economic Review*, forthcoming.

Bernheim, B. D. and A. Rangel (2009). Beyond Revealed Preference: Choice-theoretic Foundations for Behavioral Welfare Economics. *The Quarterly Journal of Economics 124*(1), 51–104.

Beshears, J., J. J. Choi, D. Laibson, and B. C. Madrian (2008). How are Preferences Revealed? *Journal of Public Economics 92*(8), 1787–1794.

Bordalo, P., N. Gennaioli, and A. Shleifer (2012). Salience Theory of Consumer Choice. *NBER Working Paper 17947*(3).

Bundorf, M. K. and H. Szrek (2010). Choice Set Size and Decision Making: The Case of Medicare Part D Prescription Drug Plans. *Medical Decision Making 30*(5), 582–593.

Cebul, R., J. Rebitzer, L. Taylor, and M. Votruba (2011). Unhealthy Insurance Markets: Search Frictions and the Cost and Quality of Health Insurance. *American Economic Review 100*(4), 1842–1871.

Chetty, R., A. Looney, and K. Kroft (2009). Salience and Taxation: Theory and Evidence. *American Economic Review 99*(4), 1145–1177.

Choi, J. J., D. Laibson, and B. C. Madrian (2011). $100 Bills on the Sidewalk: Suboptimal Investment in 401(k) plans. *Review of Economics and Statistics 93*(3), 748–763.

Dafny, L. and D. Dranove (2008). Do Report Cards Tell Consumers Anything They Don't Already Know? The Case of Medicare HMOs. *The Rand journal of economics 39*(3), 790–821.

Dafny, L., K. Ho, and M. Varela (2013). Let Them Have Choice: Gains from Shifting Away from Employer-sponsored Health Insurance and Toward an Individual Exchange. *American Economic Journal: Economic Policy 5*(1), 32–58.

DellaVigna, S. (2009). Psychology and Economics: Evidence from the Field. *The Journal of Economic Literature 47*(2), 315-372.

Dranove, D. and G. Z. Jin (2010). Quality Disclosure and Certification: Theory and Practice. *The Journal of Economic Literature 48*(4), 935-963.

Ericson, K. M. M. and J. Kessler (2013). The Articulation Effect of Government Policy: Health Insurance Mandates Versus Taxes. *NBER Working Paper 18913*.

Ericson, K. M. and A. Starc (2012a). Heuristics and Heterogeneity in Health Insurance Exchanges: Evidence from the Massachusetts Connector. *The American Economic Review 102*(3), 493–497.

Ericson, K. M. M. and A. Starc (2012b). Pricing Regulation and Imperfect Competition on the Massachusetts Health Insurance Exchange. *NBER Working Paper 18089*.

Ericson, K. M. and A. Starc (2013). Designing and Regulating Health Insurance Exchanges: Lessons from Massachusetts. *Inquiry*, 2012-2013 Winter;49(4):327-38.

Finkelstein, Amy. (2004). Minimum Standards, Insurance Regulation and Adverse Selection: Evidence from the Medigap Market *Journal of Public Economics*, 88(12):2515-2547.

Finkelstein, Amy. (2009). E-ztax: Tax Salience and Tax Rates. *Quarterly Journal of Economics*, 124(3):969-1010.

Fox, P. D., R. E. Snyder, and T. Rice (2003). Medigap Reform Legislation of 1990: A 10-year Review. *Health Care Financing Review 24*(3), 121–138.

Frank, R. G. and K. Lamiraud (2009). Choice, Price Competition and Complexity in Markets for Health Insurance. *Journal of Economic Behavior & Organization 71*(2), 550–562.

Grubb, M. D. (2009). Selling to Overconfident Consumers. *The American Economic Review 99*(5), 1770–1807.

Handel, B. (2013). Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. *The American Economic Review*, forthcoming.

Handel, B. R. and J. T. Kolstad (2013). Health Insurance for Humans: Information Frictions, Plan Choice, and Consumer Welfare. *NBER Working Paper 19373*.

Hanoch, Y., T. Rice, J. Cummings, and S. Wood (2009). How much Choice is Too Much? The Case of the Medicare Prescription Drug Benefit. *Health Services Research 44*(4), 1157–1168.

Hortaccsu, A. and C. Syverson (2004). Search Costs, Product Differentiation, and Welfare Effects of Entry: a Case Study of SP 500 Index Funds. *Quarterly Journal of Economics 119*(4), 403–456.

Iyengar, S. S. and M. R. Lepper (2000). When Choice is Demotivating: Can One Desire Too Much of a Good Thing? *Journal of Personality and Social Psychology 79*(6), 995.

Jin, G. Z. and P. Leslie (2003). The Effect of Information on Product Quality: Evidence from Restaurant Hygiene Grade Cards. *The Quarterly Journal of Economics 118*(2), 409–451.

Jin, G. Z. and A. T. Sorensen (2006). Information and Consumer Choice: the Value of Publicized Health Plan Ratings. *Journal of Health Economics 25*(2), 248–275.

Kahneman, D., P. P. Wakker, and R. Sarin (1997). Back to Bentham? Explorations of Experienced Utility. *The Quarterly Journal of Economics 112*(2), 375–406.

Kamenica, E. (2008). Contextual inference in Markets: On the Informational Content of Product Lines. *The American Economic Review 98*(5), 2127–2149.

Kesternich, Iris, Florian Heiss, Daniel McFadden, and Joachim Winter. (2013). Suit the Action to the Word, the Word to the Action: Hypothetical Choices and Real Decisions in Medicare Part D *Journal of Health Economics*.

Kling, J. R., S. Mullainathan, E. Shafir, L. C. Vermeulen, and M. V. Wrobel (2012). Comparison Friction: Experimental Evidence from Medicare Drug Plans. *The Quarterly Journal of Economics 127*(1), 199–235.

Krueger, Alan B., and Ilyana Kuziemko. (2013). The Demand for Health Insurance Among Uninsured Americans: Results of a Survey Experiment and Implications for Policy. *Journal of Health Economics*.

Lacetera, N., D. G. Pope, and J. R. Sydnor (2012). Heuristic Thinking and Limited Attention in the Car Market. *The American Economic Review 102*(5), 2206–36.

Rice, T. and K. Thomas (1992). Evaluating the New Medigap Standardization Regulations. *Health Affairs 11*(1), 194–207.

Sorensen, A. T. (2000). Equilibrium Price Dispersion in Retail Markets for Prescription drugs. *Journal of Political Economy 108*(4), 833–850.

Starc, A. (2012). "Insurer Pricing and Consumer Welfare: Evidence from Medigap. *Working Paper*.

Toolkit Series (2011). *Determining Health Benefit Designs to be Offered on a State Health Insurance Exchange*. Blue Cross Blue Shield of Massachusetts Foundation Health Reform Toolkit Series.

Train, K. E. (2003). *Discrete Choice Methods with Simulation*. Cambridge University Press.

Tversky, A. and I. Simonson (1993). Context-dependent Preferences. *Management Science 39*(10), 1179–1189.

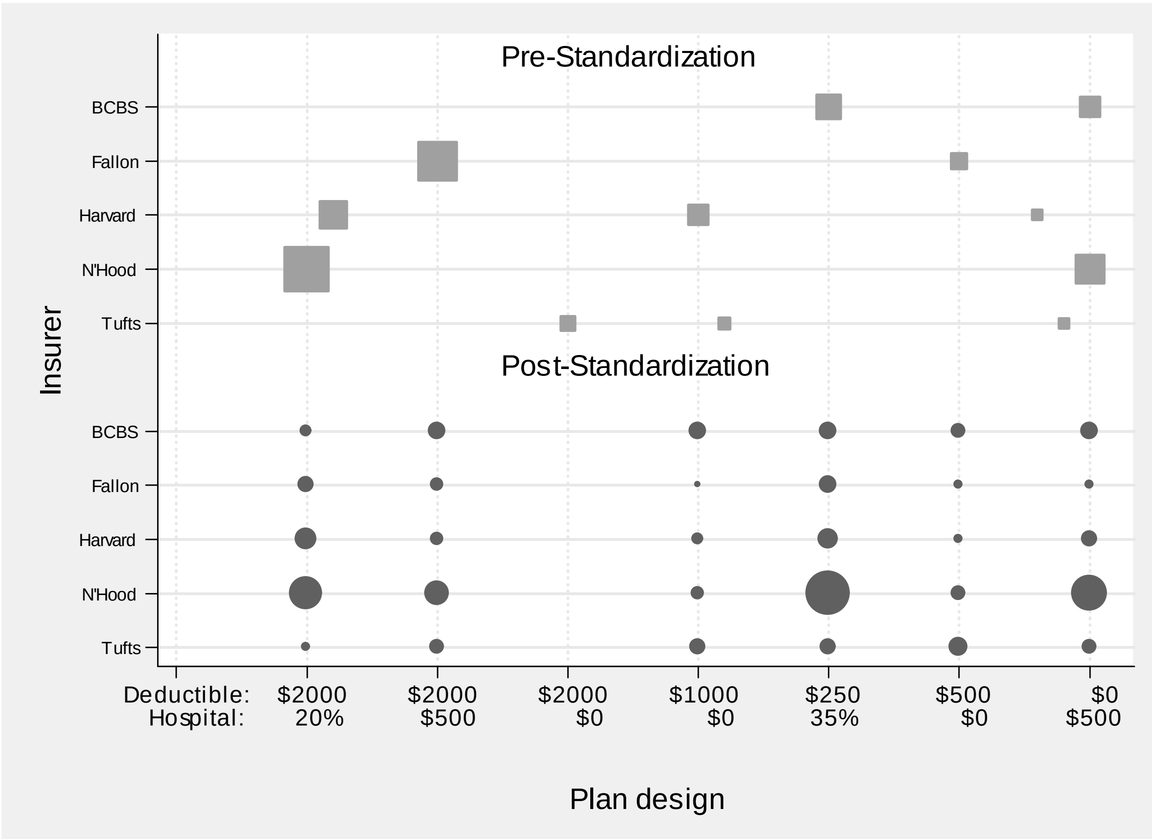


Figure 1: Plan Design and Market Share Before and After Standardization. The x-axis indicates plan design parameters. All post-standardization plans line up with one of the listed designs, while pre-standardization plans are offset from the vertical line when plans differ slightly in parameters. Size of markers indicates relative market share. Only bronze and silver plans are shown above.

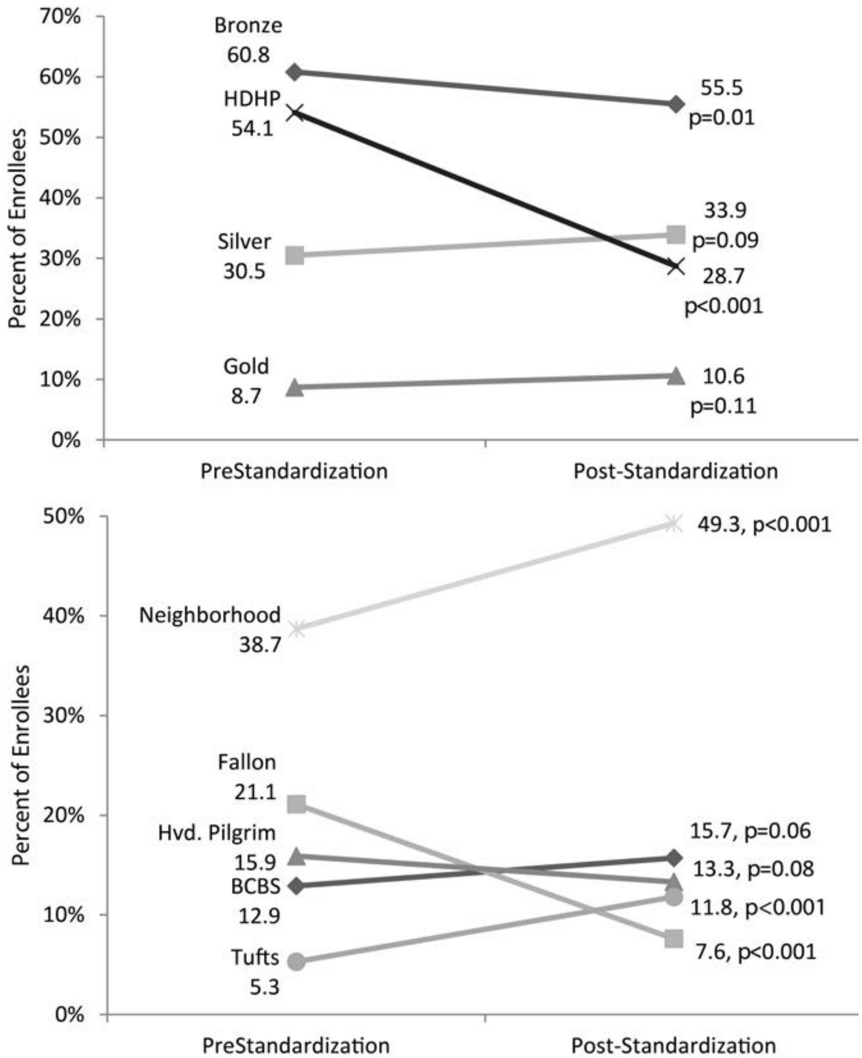


Figure 2: Enrollee Plan Choice, Pre- and Post-Standardization

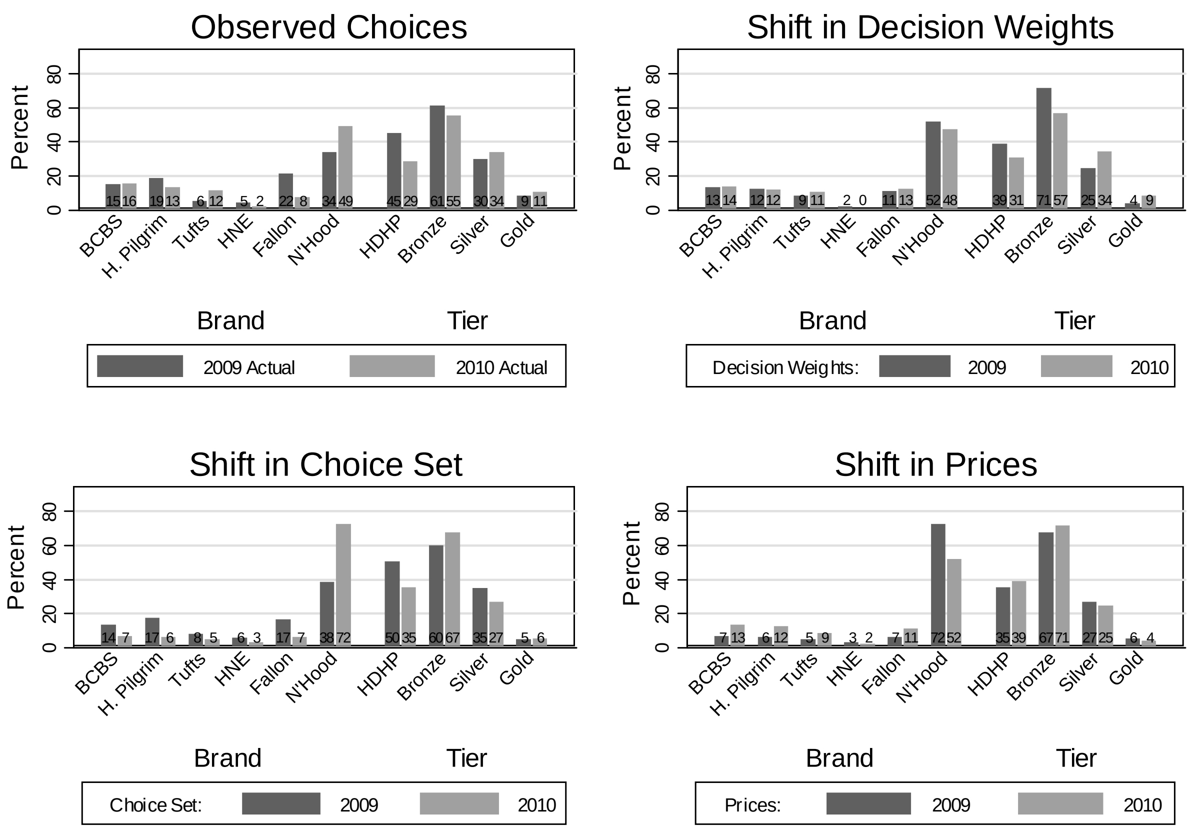


Figure 3: Counterfactual Simulations: Enrollment Under Alternative Decision Weights, Prices, and Choice Sets. Notes: Choice set shift holds constant decision weights and prices at 2009 levels. Price shift holds constant decision weights at 2009 levels and uses the 2010 choice set. Decision weight shift uses the 2010 prices and choice set.

Table 1: Plan and Enrollee Characteristics, by Year

	2009	2010	
Actuarial Value	77.8	82.5	$p<0.001$
Mean OOP Cost	1129	870	$p<0.001$
Std. Dev. OOP Cost	803	752	$p<0.001$
Monthly Premium Paid	$374	$389	$p=0.02$
Enrollee Age	42.5	43.3	$p=0.12$
N	982	1336	

Notes: *** $p<0.001$,** $p<0.01$,* $p<0.05$. Two sample tests of proportions (binary variables) or t-tests (continuous variables). Sample is restricted to consumers enrolling in a HIX plan for the first time between November 1, 2009 and February 28, 2010.

Table 2: Hedonic Regressions, List Prices

	(1)	(2)	(3)	(4)
2010 Dummy	2.019	11.71***	16.72***	116.2***
	(3.552)	(1.915)	(4.074)	(26.11)
Silver		111.4***	124.5***	112.3***
		(2.279)	(3.970)	(3.862)
Gold		227.6***	284.0***	244.5***
		(4.059)	(5.723)	(6.252)
Silver*2010			-10.07**	3.805
			(4.751)	(4.649)
Gold*2010			-38.36***	-17.88**
			(7.205)	(7.757)
HDHP			-19.48***	47.76***
			(5.253)	(6.854)
HDHP*2010			3.657	-15.07*
			(6.090)	(8.509)
Actuarial Value	6.469***	-11.89***		3.416***
	(0.128)	(1.493)		(0.229)
Actuarial Value^2		0.0832***		
		(0.00922)		
Actuarial Value*2010				-1.164***
				(0.298)
Constant	-206.4***	684.0***	277.3***	-21.65
	(11.51)	(59.34)	(3.379)	(20.50)
Fixed Effect	age cat.	age cat., insurer	age cat., insurer	age cat., insurer
Observations	70,577	70,577	70,577	70,577
R^2	0.719	0.920	0.917	0.922

Robust standard errors clustered at age category-plan-geography level. Sample is restricted to consumers enrolling in a HIX plan for the first time between November 1, 2009 and February 28, 2010.

Table 3: Discrete Choice Model: Decision Weights Vary Pre- and Post-Standardization

	(1)	(2)
	Condit. Logit	Mixed Logit
Silver	0.542***	1.282***
	(0.175)	(0.187)
Gold	1.299***	2.427***
	(0.259)	(0.293)
Silver*2010	0.0901	0.336*
	(0.171)	(0.177)
Gold*2010	0.342	0.455*
	(0.229)	(0.256)
HDHP	1.248***	0.575***
	(0.189)	(0.202)
HDHP*2010	-1.091***	-0.248
	(0.172)	(0.196)
Premium	-0.0254***	
	(0.00183)	
Premium*2010	-0.000981	
	(0.000733)	
Premium*Age	0.000269***	0.000364***
	(2.47e-05)	(3.78e-05)
Actuarial Value	0.0447***	0.0593***
	(0.00555)	(0.00568)
Mixed Logit: Lognormal Distribution of α_i		
Premium*2009, Mean[$\ln -\alpha_i$]		-3.140***
		(0.0662)
Premium*2010, Mean[$\ln -\alpha_i$]		-3.223***
		(0.0632)
Premium*2009, SD[$\ln -\alpha_i$]		0.340***
		(0.0301)
Premium*2010, SD[$\ln -\alpha_i$]		0.288***
		(0.0250)
Insurer Fixed Effect	Yes	Yes
N Person	2318	2318
N Person-Plan	70,577	70,577

*** p<0.001,** p<0.01,* p<0.05. This table presents estimates from conditional and mixed logit models in which the weights on product characteristics are allowed to vary by year. The mixed logit models the price coefficient as distributed log-normally in the population. Sample is restricted to consumers enrolling in a HIX plan for the first time between November 1, 2009 and February 28, 2010.

Table 4: Structured Change in Decision Weights

Panel A Models:	Optim. Frict.		Salience 1		Salience 2	
σ, Error Term	0.96	[0.0560]				
σ, Premium			0.96	[0.0438]	1.01	[0.0102]
σ, Financial Char.	3.39	[0.419]	2.30	[0.270]	2.67	[0.579]
σ, Brand	0.98	[0.145]	0.95	[0.0864]	0.94	[0.176]
Pre-Standardization Decision Utility Index:				Distribution of α_i:		
				Mean[ln $-\alpha_i$]	-3.63	[0.0730]
				SD[ln $-\alpha_i$]	0.0448	[0.0081]
Premium	-0.0249	[0.0020]	-0.0258	[0.0013]		
Premium*Age	0.0003	[$1.80*10^{-10}$]	0.0003	[$2.38*10^{-5}$]	0.0003	[$2.67*10^{-5}$]
Silver	0.170	[0.0566]	0.187	[0.0320]	0.241	[0.0934]
Gold	0.526	[0.1251]	0.599	[0.0715]	0.740	[0.2642]
HDHP	-0.0796	[0.0300]	-0.0823	[0.0228]	-0.0581	[0.0536]
Actuarial Value	0.0197	[0.0033]	0.0135	[0.00160]	0.0234	[0.0465]
Fixed Effects	insurer		insurer		insurer	

Panel B: Nested Logit Model:							Upper Nest:		
	Bronze Nest		Silver Nest		Gold Nest			Choice of Tier	
Premium	-0.0528	[0.0034]	-0.0422	[0.0039]	-0.0179	[0.0084]	ρ, 2010	0.141	[0.0422]
Premium*Age	0.0006	[0.0001]	0.0004	[0.0001]	0.0002	[0.0001]	ρ, 2009	0.276	[0.0469]
Actuarial Value	0.0387	[0.0037]	0.0755	[0.0068]			Silver	0.519	[0.0769]
Fixed Effects	insurer		insurer		insurer		Gold	1.16	[0.0762]

Notes. Bootstrapped standard errors in brackets. Salience model 1 estimates the salience model using a conditional logit, while salience model 2 uses a mixed logit. Panel B presents the substitutability model. The σ parameters in the first panel should be interpreted relative to 1, the normalized value for 2009. Actuarial value is measured on a 0 to 1 scale using the federal calculator. AV is not included in the gold nest of panel B due to lack of variation in AV among gold plans. Sample is restricted to consumers enrolling in a HIX plan for the first time between November 1, 2009 and February 28, 2010. N Enrollees = 2318 for all specifications.

Table 5: Welfare

	Conditional Logit	Mixed Logit	Nested Logit
Constant Size Choice Set:			
Total Welfare Change, in $/month	22.81	23.87	26.24
holding decision weights constant	27.04	27.95	30.26
holding prices constant (hedonics)	40.83	42.64	47.51
Full Choice Set:			
Total Welfare Change, in $/month	41.73	42.49	46.09
holding decision weights constant	45.96	46.66	50.10
holding prices constant (hedonics)	59.75	60.23	67.35
Avg. Premium Paid in 2010 , in $/month	386.70	386.70	386.70

Notes. This table presents welfare changes calculations in dollars per enrollee per month. The welfare criterion used corresponds to the 2010 estimated preferences (decision weights) in Table 4; evaluations using 2009 preferences are similar and contained in Appendix Table A.5. The first panel compares only the 18 most popular 2010 plans to the 2009 plans to eliminate welfare gains due to additional error draws in the logit model. The second panel compares the entire 2010 choice set to the 2009 choice set. When prices are held constant, the hedonic pricing model is used.

Table 6: Experiment: The Effect of Choice Menu and Interface

	Experiment			Observed in HIX	
	Pre	Post	Post-Alt	2009	2010
Bronze	33%	30%	40%	61%	55%
... Bronze HDHP	29%	13%	27%	54%	29%
Silver	41%	43%	28%	31%	34%
Gold	26%	26%	32%	9%	11%
Blue Cross	16%	18%	18%	13%	16%
Fallon	5%	1%	6%	21%	8%
Harvard Pilgrim	10%	6%	6%	16%	13%
Neighborhood	43%	59%	63%	39%	49%
Tufts	26%	16%	8%	5%	12%
N	299	307	304	982	1336

Compares choices in the experiment, by condition, alongside observed HIX choices.

Table 7: Experiment: Importance of Plan Characteristics by Condition

	Tier	Hospital	MaxOOP	Deduct.	Brand	Premium	Dr. Visit
Post-Alt	0.233	-0.0971	-0.0832	0.0767	0.174	0.0693	0.0764
	(0.153)	(0.107)	(0.105)	(0.105)	(0.142)	(0.0977)	(0.104)
Post-Stdz.	0.608***	-0.315***	-0.211*	-0.0205	0.212	-0.0453	-0.142
	(0.153)	(0.112)	(0.109)	(0.105)	(0.143)	(0.0995)	(0.110)
Constant	3.060***	5.706***	5.856***	5.535***	3.997***	5.987***	5.555***
(Pre-Stdz.)	(0.110)	(0.0748)	(0.0730)	(0.0720)	(0.101)	(0.0675)	(0.0759)

Notes. Dependent variable is level of importance (scale:1 to 7, higher is more important).

Web Appendix

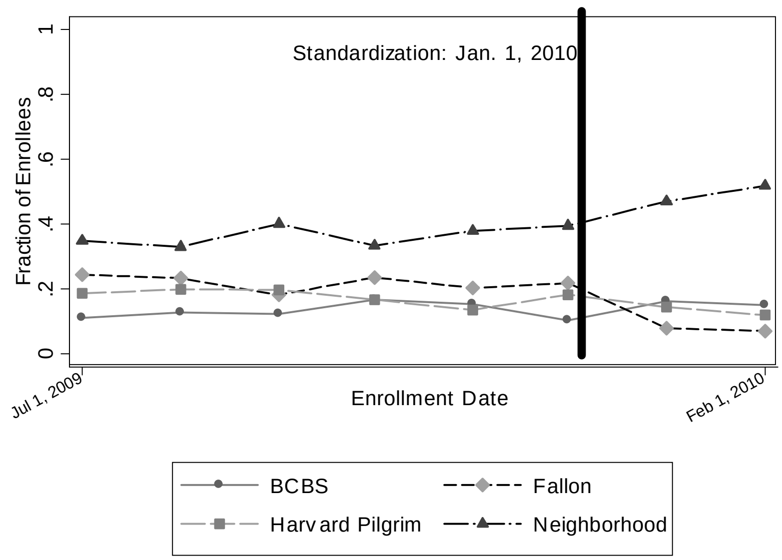


Figure A.1: No Evidence of Time Trends

Table A.1: Basic Summary Statistics

	2009	2010	
Bronze	60.8	55.5	p=0.01
Silver	30.5	33.9	p=0.09
Gold	8.7	10.6	p=0.11
HDHP	54.1	28.7	p<0.001
Chosen Plan:			
Actuarial Value	77.8	82.5	p<0.001
Mean OOP Cost	1129	870	p<0.001
Variance OOP Cost	803	752	p<0.001
Premium Paid	$374	$389	p=0.02
BCBS	12.9	15.7	p=0.06
Fallon	21.1	7.6	p<0.001
Harvard Pilgrim	15.9	13.3	p=0.08
HNE	6.1	2.3	p<0.001
Neighborhood	38.7	49.3	p<0.001
Tufts	5.3	11.8	p<0.001
Age	42.5	43.3	p=0.12
N	982	1336	

Notes: *** p<0.001,** p<0.01,* p<0.05. Two sample tests of proportions (binary variables) or t-tests (continuous variables). Sample is restricted to consumers enrolling in a HIX plan for the first time between November 1, 2009 and February 28, 2010.

Table A.2: Difference-in-Difference Estimates

	Pre. V. Post Std.		Difference in Difference		2008 v. 2009 Placebo	
Bronze	-0.0533**	(0.0207)	-0.0711**	(0.0302)	0.0178	(0.0220)
Silver	0.0336*	(0.0196)	0.0339	(0.0286)	-0.000372	(0.0208)
Gold	0.0197	(0.0123)	0.0372**	(0.0179)	-0.0174	(0.0130)
BCBS	0.0279*	(0.0146)	0.0193	(0.0228)	0.00857	(0.0175)
Fallon	-0.135***	(0.0149)	-0.170***	(0.0228)	0.0344**	(0.0173)
HNE	-0.0379***	(0.00869)	-0.0457***	(0.0114)	0.00777	(0.00732)
Neighborhood	0.106***	(0.0207)	0.149***	(0.0301)	-0.0434**	(0.0219)
Harvard Pilgrim	-0.0256*	(0.0149)	-0.0339	(0.0215)	0.00824	(0.0155)
Tufts	0.0653***	(0.0114)	0.0809***	(0.0165)	-0.0156	(0.0120)
N Enrollees	2318		4334		2016	

Displays coefficients on an indicator for post-standardization (or post-placebo). Heteroskedasticity robust standard errors in parentheses. Each row contains 3 regressions for that row's dependent variable. Sample for Pre. v. Post Std. is the paper's analysis sample (Nov 2009-Feb 2010). Sample for Difference Placebo is Nov. 2008-Feb. 2009 and otherwise has the same sample selection criteria. The difference in difference regressions use both samples.

Table A.3: No Pre-Existing Trends in Choice

	(1)	(2)	(3)	(4)	(5)	(6)
	BCBS	Fallon	HNE	N'Hood	Harvard	Tufts
Enroll Date	6.49e-05	-0.000153	0.000124	0.000297*	-0.000229	-0.000104
	(0.000124)	(0.000152)	(7.99e-05)	(0.000177)	(0.000141)	(8.51e-05)
Observations	2,699	2,699	2,699	2,699	2,699	2,699

	(7) Bronze	(8) Silver	(9) Gold
Enroll Date	0.000173	-8.42e-05	-8.84e-05
	(0.000180)	(0.000171)	(0.000105)
Observations	2,699	2,699	2,699

Sample: First-time enrollees, Jul. 2009 through Dec. 2009. Linear probability model. Coefficient on enrollment date presented, with robust standard errors in parentheses. See also Figure A.1

Table A.4: Standardization and Decision Weights, Robustness Excluding AV

	Condit. Logit	Condit. Logit	Mixed Logit	Mixed Logit
Premium	-0.0215***	-0.0227***		
	(0.00171)	(0.00176)		
Premium*2010	-0.000428	-0.00115		
	(0.000697)	(0.000729)		
Premium*Age	0.000224***	0.000241***	0.000269***	0.000305***
	(2.36e-05)	(2.40e-05)	(3.14e-05)	(3.41e-05)
Silver	0.188	0.575***	1.017***	1.120***
	(0.130)	(0.173)	(0.159)	(0.180)
Gold	1.004***	1.458***	2.032***	2.348***
	(0.234)	(0.255)	(0.269)	(0.282)
Silver*2010	0.655***	-0.0880	0.503***	0.115
	(0.122)	(0.170)	(0.139)	(0.176)
Gold*2010	1.019***	0.322	0.818***	0.396
	(0.205)	(0.229)	(0.229)	(0.252)
HDHP		0.374**		-0.327*
		(0.155)		(0.180)
HDHP*2010		-1.164***		-0.570***
		(0.172)		(0.192)
Mixed Logit: Lognormal Distribution of alphai				
Premium*2009, Mean[ln $-\alpha_i$]			-3.501***	-3.348***
			(0.0751)	(0.0724)
Premium*2010, Mean[ln $-\alpha_i$]			-3.525***	-3.394***
			(0.0723)	(0.0679)
Premium*2009, SD[ln $-\alpha_i$]			0.311***	0.325***
			(0.0355)	(0.0336)
Premium*2010, SD[ln $-\alpha_i$]			0.315***	0.308***
			(0.0320)	(0.0294)
Insurer Fixed Effect	Yes	Yes	Yes	Yes
N Person	2318	2318	2318	2318

*** $p<0.001$,** $p<0.01$,* $p<0.05$. This table presents estimates from conditional and mixed logit models in which the weights on product characteristics are allowed to vary by year. The mixed logit specifications model the price coefficient as distributed lognormally in the population. Sample is restricted to consumers enrolling in a HIX plan for the first time between November 1, 2009 and February 28, 2010.

Table A.5: Welfare

Welfare Criterion	Conditional Logit 2009	Conditional Logit 2010	Mixed Logit 2009	Mixed Logit 2010	Nested Logit 2009	Nested Logit 2010
Constant Size Choice Set:						
Total Welfare Change, in $/month	30.54	22.81	30.71	23.87	32.13	26.24
holding decision weights constant	25.22	27.04	26.62	27.95	29.13	30.26
holding prices constant (hedonics)	43.69	40.83	41.59	42.64	43.78	47.51
Full Choice Set:						
Total Welfare Change, in $/month	48.63	41.73	48.04	42.49	52.95	46.09
holding decision weights constant	44.75	45.96	45.10	46.66	51.15	50.10
holding prices constant (hedonics)	65.12	59.75	63.59	60.23	67.30	67.35
Mean Premium Paid, in $/month	352.45	386.70	352.45	386.70	352.45	386.70

Notes. This table presents welfare changes calculations in dollars per enrollee per month. The welfare criterion in the first column corresponds to 2009 and salience parameters of one for premium, brand, and financial characteristics; the criterion in the second column corresponds to 2010 and the salience parameters estimated in Table 4. The first panel compares only the 18 most popular 2010 plans to the 2009 plans to eliminate welfare gains due to additional error draws in the logit model. The second panel compares the entire 2010 choice set to the 2009 choice set. When prices are held constant, the hedonic pricing model is used.

Table A.6: Counterfactual Plan Enrollment: Alternative Decision Weights and Prices

Decision Weight	2009	2009	2010	2010	Observed
Prices	2010	2009	2010	2009	Observed
		2009 Choice Set			
Bronze	0.65	0.60	0.48	0.43	0.61
Silver	0.31	0.35	0.43	0.45	0.30
Gold	0.04	0.05	0.09	0.11	0.09
HDHP	0.54	0.50	0.39	0.36	0.45
BCBS	0.16	0.14	0.17	0.14	0.15
Fallon	0.22	0.17	0.19	0.15	0.22
Harvard Pilgrim	0.22	0.17	0.19	0.15	0.19
HNE	0.06	0.06	0.01	0.01	0.05
Neighborhood	0.25	0.38	0.28	0.40	0.34
Tufts	0.09	0.08	0.12	0.10	0.05
		2010 Choice Set			
Bronze	0.71	0.67	0.57	0.52	0.55
Silver	0.25	0.27	0.34	0.36	0.34
Gold	0.04	0.06	0.09	0.11	0.11
HDHP	0.39	0.35	0.31	0.28	0.29
BCBS	0.13	0.07	0.14	0.07	0.16
Fallon	0.11	0.07	0.13	0.08	0.08
Harvard Pilgrim	0.12	0.06	0.12	0.07	0.13
HNE	0.02	0.03	0.00	0.00	0.02
Neighborhood	0.52	0.72	0.48	0.68	0.49
Tufts	0.09	0.05	0.11	0.07	0.12

Notes. This table describes plan market shares across counterfactual scenarios in which decision weights and prices vary by year. The hedonic pricing model is used. The averages are presented for the 2009 consumers and choice set in the first panel and the 2010 consumers and choice set in the second panel.

Table A.7: Experiment: The Effect of Choice Menu and Interface

Panel A: Generosity				
	HDHP	**Bronze**	**Silver**	**Gold**
Pre-Stdz.	Comparison Condition			
Post-Stdz.	-0.161***	-0.0245	0.0113	0.0132
	(0.0332)	(0.0384)	(0.0411)	(0.0362)
PostAlt	-0.0260	0.0731*	-0.133***	0.0603
	(0.0369)	(0.0398)	(0.0393)	(0.0371)
Controls	Yes	Yes	Yes	Yes

Panel B: Brand					
	BCBS	**Fallon**	**HarvardPilgrim**	**N'Hood**	**Tufts**
Pre-Stdz.	Comparison Condition				
Post-Stdz.	0.0142	-0.0398***	-0.0451**	0.188***	-0.117***
	(0.0310)	(0.0136)	(0.0220)	(0.0394)	(0.0325)
PostAlt	0.0162	0.00565	-0.0429**	0.203***	-0.182***
	(0.0312)	(0.0187)	(0.0217)	(0.0395)	(0.0296)
Controls	Yes	Yes	Yes	Yes	Yes

Note: Robust standard errors in parentheses. Controls include age, state of residence, level of self-reported health, education, and income category. All regressions have N=910.

Table A.8: Salience of Plan Characteristics, by Condition

	Tier	Hospital Stay	Max OOP	Deduct.	Brand	Premium	Dr Visit
Pre-Stdz.	Comparison Category						
PostAlt	0.219	-0.108	-0.107	0.0651	0.165	0.0496	0.0485
	(0.153)	(0.106)	(0.103)	(0.105)	(0.144)	(0.0983)	(0.105)
Post-Stdz.	0.576***	-0.325***	-0.234**	-0.0433	0.160	-0.0659	-0.150
	(0.153)	(0.111)	(0.108)	(0.105)	(0.143)	(0.100)	(0.110)
Controls	Yes	Yes	Yes	Yes	Yes	Yes	Yes

Note: Robust standard errors in parentheses. Controls include age, state of residence, level of self-reported health, education, and income category. All regressions have N=910.

Table A.9: Results from Experiment: Decision Weights Vary Pre- and Post-Standardization

	(1)	(2)
	Condit. Logit	Mixed Logit
Silver	0.549	1.238***
	(0.386)	(0.458)
Gold	1.149**	2.379***
	(0.503)	(0.666)
Silver*2010	-0.0337	0.392
	(0.384)	(0.428)
Gold*2010	0.690	1.072**
	(0.430)	(0.484)
HDHP	0.728**	0.541
	(0.346)	(0.399)
HDHP*2010	-1.755***	-1.664***
	(0.412)	(0.445)
Premium	-0.00422	
	(0.00269)	
Premium*2010	-0.00347***	
	(0.00108)	
Premium*Age	2.32e-05	6.75e-05
	(3.59e-05)	(5.00e-05)
Mixed Logit: Lognormal Distribution of $-\alpha_i$		
Premium*2009, Mean[ln $-\alpha_i$]		-4.495***
		(0.405)
Premium*2010, Mean[ln $-\alpha_i$]		-4.115***
		(0.292)
Premium*2009, SD[ln $-\alpha_i$]		0.431***
		(0.132)
Premium*2010, SD[ln $-\alpha_i$]		0.580***
		(0.152)
Insurer Fixed Effect	Yes	Yes
N Person	543	543
N Person-Plan	14685	14685

Notes. Limits sample to Pre-Stdz. and Post-Stdz. conditions only (excludes Post-Alt.), and limits to ages 65 and under.

A.1 Experiment Design

Participants were recruited using Qualtrics Panels. Participants were limited to residents of northeastern states (Maine, Massachusetts, New Hampshire, Vermont, Connecticut, Rhode

Island, and New York.) Individuals gave their income, and were screened out of the experiment if their pre-tax household income was below the following thresholds: \$35,000 if single, \$45,000 for a household of 2, \$55,000 for 3, or \$65,000 for 4 or more. After answering some demographic questions, participants were assigned to a condition. Participants were divided into two major age groups: over 45 and under 45. Assignment to condition was balanced within age group, and each age group saw different prices (as prices on the HIX are age-dependent).

Participants saw the choice menu (plans and prices) available in zipcode 02130, and saw age 35 prices (if under 45) and age 50 prices (if over age 45). Zipcode 02130 (Jamaica Plain, Mass.) is similar to other zipcodes, with the exception that Health New England was not offered in this area; Health New England has relatively low market share.

Participants in the Pre-Stdz. condition simply chose their plan from the list of plans in an interface similar to the HIX's pre-standardization interface. (Compare Figure A.4 with Figure A.7.) In the Post-Stdz. condition, participants first chose a tier (Figure A.5), and then chose a plan (Figure A.6); compare to Figures A.2 and A.3. Finally, participants in the Alt-Post condition saw the post standardization plans using the pre-standardization interface (Figure A.8).

After making their choices, participants rated how important each of a list of attributes was for their decision, on a scale of 1-7. Then, participants made an additional choice in a different condition: participants in the Pre-Stdz. condition made their second choice in the Post-Stdz. condition, while participants in both the Post-Stdz. and Alt-Post conditions made their second choice in the Pre-Stdz. condition. (Our primary analyses rely only on participant's first choice, but this data was collected since the marginal cost was low). Participants then rated the importance of the factors for their second condition.

Figure A.2: Post Standardization Choice Interface: Choose Tier

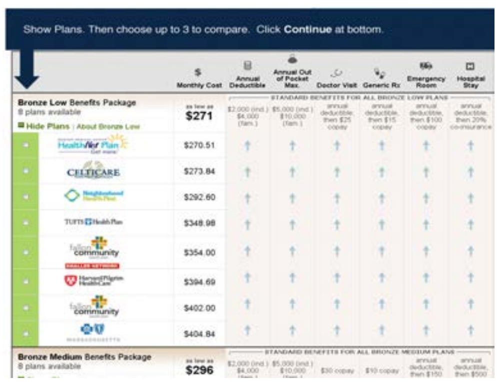
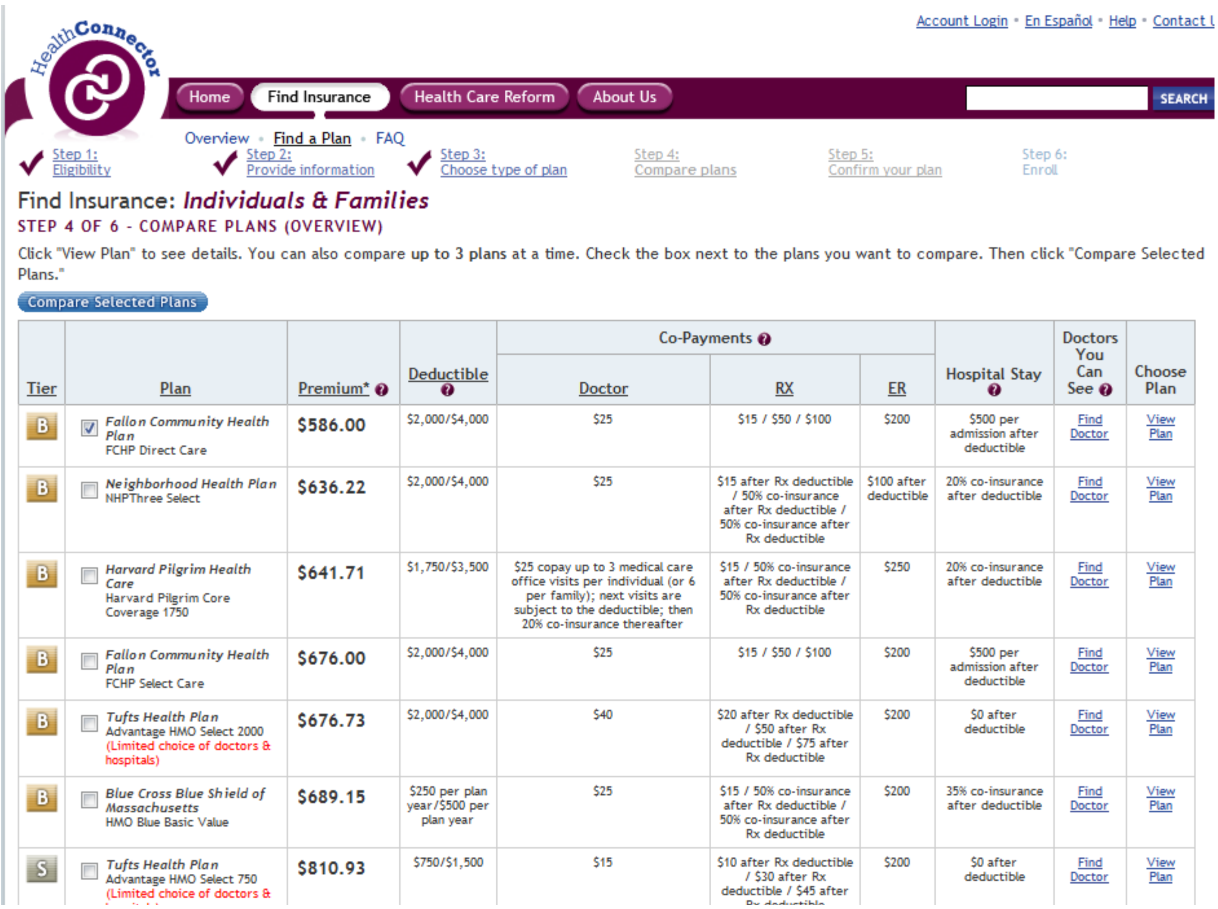


Figure A.3: Post Standardization Choice Interface: Choose Plan


Account Login • En Español • Help • Contact U

Health Connector

Home | Find Insurance | Health Care Reform | About Us | SEARCH

Overview • Find a Plan • FAQ

Step 1: Eligibility | Step 2: Provide information | Step 3: Choose type of plan | Step 4: Compare plans | Step 5: Confirm your plan | Step 6: Enroll

Find Insurance: *Individuals & Families*

STEP 4 OF 6 - COMPARE PLANS (OVERVIEW)

Click "View Plan" to see details. You can also compare up to 3 plans at a time. Check the box next to the plans you want to compare. Then click "Compare Selected Plans."

Compare Selected Plans

Tier	Plan	Premium*	Deductible	Co-Payments: Doctor	Co-Payments: RX	Co-Payments: ER	Hospital Stay	Doctors You Can See	Choose Plan
B	*Fallon Community Health Plan* FCHP Direct Care	$586.00	$2,000/$4,000	$25	$15 / $50 / $100	$200	$500 per admission after deductible	Find Doctor	View Plan
B	*Neighborhood Health Plan* NHPThree Select	$636.22	$2,000/$4,000	$25	$15 after Rx deductible / 50% co-insurance after Rx deductible / 50% co-insurance after Rx deductible	$100 after deductible	20% co-insurance after deductible	Find Doctor	View Plan
B	*Harvard Pilgrim Health Care* Harvard Pilgrim Core Coverage 1750	$641.71	$1,750/$3,500	$25 copay up to 3 medical care office visits per individual (or 6 per family); next visits are subject to the deductible; then 20% co-insurance thereafter	$15 / 50% co-insurance after Rx deductible / 50% co-insurance after Rx deductible	$250	20% co-insurance after deductible	Find Doctor	View Plan
B	*Fallon Community Health Plan* FCHP Select Care	$676.00	$2,000/$4,000	$25	$15 / $50 / $100	$200	$500 per admission after deductible	Find Doctor	View Plan
B	*Tufts Health Plan* Advantage HMO Select 2000 (Limited choice of doctors & hospitals)	$676.73	$2,000/$4,000	$40	$20 after Rx deductible / $50 after Rx deductible / $75 after Rx deductible	$200	$0 after deductible	Find Doctor	View Plan
B	*Blue Cross Blue Shield of Massachusetts* HMO Blue Basic Value	$689.15	$250 per plan year/$500 per plan year	$25	$15 / 50% co-insurance after Rx deductible / 50% co-insurance after Rx deductible	$200	35% co-insurance after deductible	Find Doctor	View Plan
S	*Tufts Health Plan* Advantage HMO Select 750 (Limited choice of doctors &	$810.93	$750/$1,500	$15	$10 after Rx deductible / $30 after Rx deductible / $45 after	$200	$0 after deductible	Find Doctor	View Plan

Figure A.4: Pre Standardization: Choice

In the table below, there are a number of different of health insurance plan designs. Each plan type is called a tier, with Gold plans being most generous and Bronze-Low plans being the least generous.

Each tier is offered by five different insurance companies at different prices. After choosing a tier, you will be given the opportunity to choose which insurance company you prefer.

If you had to buy one of these types of plans to get health insurance for yourself, which would you choose? (*Click next to the tier of plan you prefer.*)

Tier	Monthly Cost *as low as:*	Annual Deductible	Annual Out-of-Pocket Max	Doctor Visit	Hospital Stay
○ Bronze Low	$215	$2,000	$5,000	Deductible, then $25 copay	Deductible, then 20% coinsurance
○ Bronze Medium	$234	$2,000	$5,000	$30 copay	Deductible, then $500 copay
○ Bronze High	$225	$250	$5,000	$25 copay	Deductible, then 35% coinsurance
○ Silver Low	$303	$1,000	$2,000	$20 copay	Deductible, then no copay
○ Silver Medium	$321	$500	$2,000	$20 copay	Deductible, then no copay
○ Silver High	$305	None	$2,000	$25 copay	$500 copay
○ Gold	$373	None	None	$20 copay	$150 copay

Some useful terms:
Annual Deductible: What you'll pay during a plan year before your health plan will cover certain medical services.

Annual Out of Pocket Maximum: A cap on your combined deductible, co-insurance and copay costs for some medical services

Doctor Visit: What you'll pay out of pocket for a visit to your PCP. Plans will waive some or all of these costs for routine or 'wellness' visits."

Figure A.5: Experiment: Post-Stdz. Choice of Tier

Bronze Medium

Insurer	Monthly Cost	Annual Deductible	Annual Out-of-Pocket Max	Doctor Visit	Hospital Stay
Neighborhood Health Plan	$234	$2000	$5000	$30 copay	Deductible, then $500 copay
Tufts Health Plan	$269	$2000	$5000	$30 copay	Deductible, then $500 copay
Blue Cross Blue Shield of Massachusetts	$320	$2000	$5000	$30 copay	Deductible, then $500 copay
Fallon Community Health Plan	$336	$2000	$5000	$30 copay	Deductible, then $500 copay
Harvard Pilgrim Health Care	$353	$2000	$5000	$30 copay	Deductible, then $500 copay

Figure A.6: Experiment: Post-Stdz. Choice of Plan

In the table below, there are a number of different health insurance plans. They each have different levels of coverage and have different prices.

Please read through the list of plans.

If you had to buy one of these plans to get health insurance for yourself, which would you choose? (*Click next to the health insurance plan you prefer.*)

Tier	Insurer	Premium/month	Deductible/yr	Doctor visit	Hospital Stay
◯ Bronze	Neighborhood Health Plan	$221	$2,000	$25	20% co-insurance after deductible
◯ Bronze	Tufts Health Plan	$281	$2,000	$40	$0 after deductible
◯ Bronze	Harvard Pilgrim Health Care	$296	$1,750	$25	20% co-insurance after deductible
◯ Bronze	Fallon Community Health Plan	$306	$2,000	$25	$500 per admission after deductible
◯ Silver	Neighborhood Health Plan	$308	$0	$25	$500 per day (up to $2,000 per individual or $4,000 per family maximum per calendar year)
◯ Silver	Neighborhood Health Plan	$325	$0	$25	$500 per day (up to $2,000 per individual or $4,000 per family maximum per calendar year)
◯ Bronze	Blue Cross Blue Shield of Massachusetts	$333	$250	$25	35% co-insurance after deductible
◯ Silver	Tufts Health Plan	$336	$750	$15	$0 after deductible
◯ Gold	Neighborhood Health Plan	$385	$0	$15	$100 per admission
◯ Silver	Tufts Health Plan	$390	$0	$20	$600 per

Figure A.7: Experiment: Pre-Stdz. Choice

Tier	Insurer	Monthly Premium	Annual Deductible	Doctor Visit	Hospital Stay
Bronze	Neighborhood Health Plan	$215	$2000	Deductible, then $25 copay	Deductible, then 20% coinsurance
Bronze	Harvard Pilgrim Health Care	$296	$2000	Deductible, then $25 copay	Deductible, then 20% coinsurance
Bronze	Tufts Health Plan	$298	$2000	Deductible, then $25 copay	Deductible, then 20% coinsurance
Bronze	Blue Cross Blue Shield of Massachusetts	$309	$2000	Deductible, then $25 copay	Deductible, then 20% coinsurance
Bronze	Fallon Community Health Plan	$318	$2000	Deductible, then $25 copay	Deductible, then 20% coinsurance
Bronze	Neighborhood Health Plan	$234	$2000	$30 copay	Deductible, then $500 copay
Bronze	Tufts Health Plan	$269	$2000	$30 copay	Deductible, then $500 copay

Figure A.8: Experiment: Post-Stdz. Alternative Interface