How Product Standardization Affects Choice: Evidence from the Massachusetts Health Insurance Exchange*

PRELIMINARY, For NBER Conference. Please Do Not Cite

Keith M Marzilli Ericson and Amanda Starc

Boston University and University of Pennsylvania

March 18, 2013

Abstract

Standardization of complex products is widely touted as improving consumer decisions and intensifying price competition, but evidence on standardization's effect 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 contracts, which were then grouped into tiers of quality. A regulatory change then forced firms to standardize the financial features of their insurance plans and offer seven defined plans; plans remained differentiated on network

*We thank Raj Chetty, David Cutler, Jonathan Gruber, Larry Katz, Jon Kolstad, David Laibson, Ariel Pakes, Jim Rebitzer, Josh Schwartzstein, and Bob Town for thoughtful comments. We acknowledge funding from the Lab for Economic Applications and Policy (LEAP) at Harvard University, and Wharton internal research funds.

and brand. We find that standardization altered consumers' choices, their valuation of plan attributes, and the market equilibrium. Post-standardization, consumers shifted to relatively more generous health insurance plans and to alternative brands. Using a discrete choice model, we estimate how standardization affected the weights consumers placed on plan attributes. We evaluate the welfare effects of standardization and conduct a number of counterfactuals. We use hypothetical choice experiments with different insurance menus to replicate the effect of standardization and conduct alternative counterfactuals.

1 Introduction

Effective markets rely on consumers making informed choices. Yet in many contexts, consumers face difficult, complex choices: they may not understand the product itself, they may have difficulty comparing across products (Kling et al. 2011), they may be overwhelmed by large choice sets (Iyengar and Lepper 2000)[1], and they may be unable to observe important dimensions of product quality. Market makers and regulators often seek to help consumers with "choice architecture". For example, policy-makers may require firms to disclose certain types of information about their products (e.g. mileage on cars; see Dranove and Jin 2010) as well as creating certification schemes that 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 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.

We examine product standardization, a choice architecture intervention designed to both

[1]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. Hanoch et al. (2009) and Bundorf and Szrek (2010) show experimentally that decision making difficulty grows with choice set size.

improve consumer choice and increase competitive pressures on firms. When products vary across many dimensions, consumers may not be able to identify which dimensions are most important, may not pay attention to certain characteristics (Lacetera, Pope, and Sydnor 2012) and therefore may not be able to rank products based on quality. Standardization entails defining dimensions on which products cannot vary, and could entail limiting firms to a single "standard product" or a set of standardized products. Policy-makers makers hope standardization will facilitate comparisons across types of products, improving consumer matching by making quality easier to observe. Moreover, policy-makers hope standardization will facilitate comparison shopping within a type of product, reducing product differentiation and thereby reducing firms' markups over cost. Despite the promise of standardization, there has been little research on standardization's effects.[2]

We examine a natural experiment: the standardization of health insurance plans on the Massachusetts health insurance exchange (HIX). Initially, the Massachusetts HIX gave firms wide latitude to design the terms of insurance plans. However, a regulatory change in 2010 standardized the financial characteristics plans on the exchange, allowing seven distinct plan types. Brand, physician/hospital network, and price were still allowed to vary; network differentiation matters for consumers, as may brand preferences.[3] We examine the impact of standardization on consumer choice, firm pricing, and equilibrium outcomes. These results are of general interest for economists examining how product regulation can alter consumer choice and affect equilibrium outcomes. They are also directly policy-relevant, as HIXs in other states are considering whether to standardize plans or not.

HIXs are an ideal context to examine standardization. Insurance plans are multi-dimensional products (e.g. coinsurance, copayments, deductibles, maximum out-of-pocket limits) that are difficult for consumers to evaluate. Health insurance markets have particularly high levels of consumer confusion (e.g. see Abaluck and Gruber 2010 on Medicare Part D) and evidence

[2]Relatively more research has been done on standardizing information disclosure, which is related to, but distinct from, standardizing the products themselves.

[3]Starc (2012) finds that consumers have preferences over the brands in the Medigap market, despite plans having identical financial and network characteristics.

indicates individuals misunderstand important 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).

HIX are an example of an overall shift to consumer-driven markets for health insurance, and the HIXs play an important role in the Patient Protection and Affordable Care Act (ACA). Approximately 20 million consumers will receive coverage via the exchanges, with subsidies totalling about $100 billion per year, tied to the premiums set by insurers. Various states have– or will– set up set up HIXs as a result of the ACA; the federal government will run the exchanges for other states. Policy-makers are actively debating whether to standardize health insurance plans on the HIX. 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).

To analyze the impact of standardization on choice we first show reduced form evidence that it 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. There are two potential sources of this shift: an "availability" and a "valuation" effect. For availability, choices could change because the mix of products available changed and the utility-maximizing choice may differ between the old and new choice sets. Alternatively, standardization could have changed how consumers value plan attributes, either by changing understanding of product attributes or because preferences are choice-set dependent (Tversky and Simonson 1993). With the valuation effect, there is a change in the "decision utility function" (in the language of Kahneman, Wakker, and Sarin 1997[4]), which simply means that the weights on attributes that rationalize observed choices differ. The decision utility

[4]I.e. "decision utility" is distinguished from "experienced utility".

function can be influenced by both elements of the product that are directly relevant for consumer welfare, as well as by aspects of the decision making process (e.g. the use of heuristics, see Ericson and Starc 2012).

To distinguish between the availability and valuation effects, we estimate a discrete choice model in which decision weights on various insurance attributes (deductible, brand, etc.) are allowed to vary pre- and post-standardization. We find that decision weights differ significantly pre- and post-standardization. The differential decision weights can be interpreted as changes in information, attention (DellaVigna 2009) or salience (Bordalo and Shleifer 2012); relatedly, consumers may perceive implicit recommendations contained in the design of choice set (contextual inference; see Kamenica 2008). The finding of different decision weights pre- and post-standardization contributes to a literature that finds important effects of context on preferences.

We use our discrete choice estimates to break down the total effect of standardization into components. Because standardization changes how consumers choose among products, it will also change how firms optimally set prices. We examine how firm markups over cost change as a result of standardization. To do so, we run counterfactuals that examine the equilibrium policy implications of standardization. First, we show that choices are affected by both the salience of product characteristics and firm decisions, examining how many more consumers would have chosen a high deductible health plan (HDHP) in the absence of a firm price response.

In practice, standardization of products entails two change: an actual change in the choice menu, 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 these two sources of standardization's effect in order to make predictions for other contexts. To separate these effects, we conduct an experiment in which participants make hypothetical insurance choices from with different menus and choice interfaces. Using our experimental design, we replicate the effect of

standardization on choice. This establishes both the validity of our research design using the HIX experiment, and supports the use of the hypothetical choice experiment. We further test an additional treatment, in which participants see the post-standardization menu using the pre-standardization choice interface. We find that the choice menu also has an effect on choice, and that choices would have been different if standardization had not been accompanied by a change to the decision interface.

Finally, did standardization improve consumer welfare? Welfare evaluation in the presence of choice inconsistencies is controversial (see Bernheim and Rangel 2011, Beshears et al. 2008). We evaluate standardization's effect on welfare using two different welfare criteria: the estimated utility function pre-standardization, and the estimated utility function post-standardization. While the two criteria give different magnitudes of the effect on standardization, both agree that standardization increased welfare.

Behavioral biases, as well as nudges or choice architectures meant to counter these biases, are often evaluated in a partial equilibrium setting in which only consumers choices change. However, as noted by Grubb (2009) and Handel (2013), firms are likely to respond to both behavioral biases and nudges as well. Our unique context allows us to explore both demand- and supply- side reactions to the policy change; perhaps unsurprisingly, we find a role for both in our consumer welfare calculations. A study that examined consumer choices in isolation would overstate that value of the policy intervention. Therefore, we argue that our approach provides additional context for policy-makers considering nudges.

The paper is organized as follows. Section 2 describes the Massachusetts Connector and the policy change. Section 3 describes the data and 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 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 2006 Massachusetts health reform. The HIX we examine is an unsubsidized health insurance exchange (termed "Commonwealth Choice") for individuals and families making 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 operating for 2 years, and was regulated by the state. Eventually, it will transition to being an ACA exchange, with slightly different regulation. See Ericson and Starc (2013) for more detail on the exchange. Previous work on the Massachusetts HIX has modeled consumer demand (Ericson and Starc 2012a), and pricing regulation in the presence of imperfect competition (Ericson and Starc 2012b).

2.2 Plan Standardization

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 required insurers to offer a minimum number of products (six, distributed across tiers) and awarded a seal of approval only to selected providers (cite: Toolkit Series). This system evolved over time to a situation in late 2009 in which 25 distinct plans were offered by one of 6 carriers: 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".[5] However, this left board members with a difficult choice of *how* to standardize the products; rather than relying on competitive pressure forcing insurers to offer the products demand by consumers, the decision would suddenly be top-down. Furthermore, the Connector had little research to guide their decision, and relied heavily on common sense.[6]

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; in addition, Fallon began offering plans with tiered/limited networks. As a result, while standardization lowered the number of contract designs (financial parameters) used in the market, it actually increased the number of plans, in the sense of contract design-carrier combinations. Crucially, standardization also unbundled the decision making process into a decision about a contract design, followed by a decision about an insurer, as seen below in Figure 1. The standardization process is an ongoing one; the silver medium plan has been eliminated due to low demand and additional insurers have been added.

Our paper is most closely related to two ideas within the empirical industrial organization literature. The first concerns costly consumer search, which can lead to equilibrium dispersion in prices for even homogeneous goods. Various studies, including Cebul et al (2011) and Hortascu and Syverson (2004) use this equilibrium dispersion in prices to trace out a distribution of search costs. This literature has primarily focused on homogenous goods markets, though the search problem is almost certainly exacerbated by product differentiation, which implies that consumers must search over a variety of product characteristics, in addition to price. By contrast, the marketing literature has considered the possibility that not all products enter the choice set, and instead allows for consideration sets (cite). Finally, within the health context, Sorenson (2000) considers the role of price posting regu-

[5] Nancy Turnbull, Toolkit Series.
[6] Toolkit.

lation in driving the equilibrium distribution of prices of prescription drugs. This last paper illustrates and import point: information disclosure can lead to increased price competition.

Our paper is also related to the literature studying the impact of information provision on market outcomes. Sorenson (2000) considers the role of information in affecting prices. Jin and Leslie (2003) consider the impact of posting restaurant report cards, finding that increased disclosure can improve quality, and Bollinger, Leslie and Sorenson (2010) find that posting calorie counts reduces calories consumed (see also Abaluck 2011). Again, within the health 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 Sorenson (2006) find that plan ratings have a meaningful effect on quality of health plan chosen.

3 Data and Reduced-Form Evidence

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 (with additional data to be added). 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 one thousand households per month.

Our choice analyses focuses on a subset of the data: Nov. 2009-Feb. 2010. Because we observe transaction-level data, we do not observe all the plan prices that individuals face. However, for the subsample, we collected an extensive set of price quotes from the Connector website using a Perl script. The choice of sample period does not have a strong effect on the results, and we show the robustness of our results to various sample selections.

First, we provide basic summary statistics in Table 1. Bronze plans are popular during both time periods, though they are slightly more popular in the earlier period. The shift

in enrollment from bronze plans is absorbed by silver plans. Interestingly, there is a big decrease in the number of consumers who choose high deductible plans (defined as in the tax code as plans with at least a $1200 individual deductible[7]): over half of all enrollees choose high-deductible plans in 2009, while just over one-fourth choose high-deductible plans in 2010. Furthermore, premiums do not decrease, which, while surprising, could be due to a change in composition of enrollees.

To examine the effect of the standardization on prices, we present a series of hedonic regressions in 2. Unadjusted for generosity, plans are slightly cheaper in 2010. However, column 2 shows that this is largely because more lower generosity (bronze and silver) plans were available in 2010. If anything, plans became slightly more expensive in 2010. However, 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 are lower in 2010. Therefore, we should expect consumers to choose more generous plans in 2010 given the price change.

4 Standardization and Decision Weights

We estimate a discrete choice model of demand for products in which 1) we model the utility of insurance plans as a function of their characteristics and 2) allow the weight on these characteristics to differ pre- and post- standardization. Estimating separate decision weights pre- and post-standardization allow us to model how demand differs post-standardization; these demand estimates can then be used to make inference about optimal firm response.

In this section, we remain agnostic about why the decision weights change post-standardization, but discuss and test among the alternatives in subsequent sections. Decision weights can result from context-dependent inference (Kamenica 2009), context-dependent salience (Bordalo, Gennaioli, and Shleifer 2012), or from attention in the way information is presented (DellaVigna 2009).

[7] $1150 in 2009.

4.1 Model of Consumer Choice

In the simplest discrete choice model, consumers face discrete choice from a set of plans, and value various plan characteristics; the mean value of a plan will be denoted by δ_j and the individual-specific component will be denoted by μ_{ij}. Initially, we assume that consumer i's decision utility of plan j is given by:

$$u_{ij} = \delta_j + \mu_{ij} + \varepsilon_{ij}$$

where ε_{ij} is an error term that i.i.d extreme value. Consumer's (welfare-relevant) utility may deviate from the decision index, due to the decision process (see also Ericson and Starc 2012). We model both utility and salience of product characteristics and use a conditional logit specification to estimate consumer demand for insurance plans. We assume the consumer chooses the plan with the largest decision index. Given the assumption that ε_{ij} is i.i.d. extreme value, this implies that a plan's enrollment share can be written as:

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

where s_{ij} represents the probability that consumer i purchases product j.[8]

The decision index depends on decision weights, which we assume are related to– but not necessarily identical to–the underlying utility of the product. Hence, if the mean (welfare-relevant) utility of the product is given by $\delta_j = X_j^*\beta$ in the standard model, we allow the decision index in a particular context f to be $\delta_{j|f} = X_j^*\beta\sigma_f$: the vector β represents the primitive preference and the vector σ_f represents the relative weight these characteristics are given in the decision under of the product characteristics. The context f that alters decision-weights can be a function of the way information is presented, the product's position in the choice set, or any "ancillary condition" in the language of Bernheim and Rangel (2011).

[8] In the absence of individual heterogeneity μ_{ij} and heuristics h_{ij}, the δ_j parameters simply represent an inversion of the observed market shares for each plan.

4.2 Identification

In a model in which certain product characteristics receive more weight than others, identification of underlying utility is likely to be difficult. It is easy to see that such a measure is not non-parametrically identified. 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 suggests a number of solutions to this problem. While some papers rely of documenting dominated decisions or modelling things that directly affect utility, like switching costs (Handel 2012), other papers rely on restrictions from theory (Abaluck and Gruber 2012).

In our context, we can identify how standardization *changes* decision weights from one context to another; it does not allow us to identify context-free fundamental utility parameters. More formally, let δ_{jPre} be the decision index of a plan with characteristics $\mathbf{X}_j$ pre-standardization and δ_{jPost} be the decision index of a plan with the same characteristics, but post-standardization. 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 cannot 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 change in valuation post-standardization).

Identifying σ_{Post} requires some additional assumptions, which are likely to be valid in our context. First, 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, as the post-standardization plans were often modelled after pre-standardization plans. Second, there can be no heteroskedasticity that requires rescaling of the coefficients (see Train 2003). In our example, these assumptions are likely to be satisfied. However, our welfare calculations (below) will consider robustness checks that loosen these assumptions (at the cost of separate identification).

4.3 Estimates

We estimate a consumer utility function that allows for changes in decision-weights in 2010:

$$u_{ij} = \mathbf{X}_{ij}\boldsymbol{\beta}\left(\boldsymbol{\sigma}_{Post}1\left[2010\right]\right) - \alpha p_{ij}\left(\sigma_{Post}1\left[2010\right]\right) + \varepsilon_{ij},$$

allowing for separate changes in weights on all the elements in X_j : premiums, an indicators for tier and HDHP, and insurer brand. The results are in Table 3. We estimate conditional logit specifications in columns 1 and 2, which allow for heterogeneity in α based on age, but not further heterogeneity in decision weights. Then, in columns 3 and 4, we estimate a mixed logit specification, in which α is allowed 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. The estimated standard deviation on the premium coefficient is substantial and statistically significant, and so we prefer the mixed logit estimates in column 4.

In general, we find evidence that standardization increased price sensitivity only very slightly. In columns 1 and 2, we find that the premium coefficient gets more negative, a difference of about 5%. In column 4, the premium coefficient $\alpha's$ estimated mean and standard deviation is similar both pre- and post-standardization; in column 3 the mean is similar, but the standard deviation is somewhat larger post-standardization. The simpler conditional logit specifications find somewhat lower premium coefficients. You can also see relative changes in the valuation of tiers (bronze is the comparison category): HDHPs receive more negative weight post standardization, and the differences between silver/gold plans and bronze plans increases post-standardization.

Table 4 presents the results of a more structured estimation. We constrain the multiplicative change in decision weight σ to be the same with in groups of characteristics. Specification 1 gives the baseline model, in which decision weights are constrained to be the same pre- and post-standardization. Specification 2 shows that, holding the weight on brand constant, the premium is slightly more salient to consumers (a σ of 1.02), while financial characteristics,

including plan tier and the HDHP dummy, are much more salient to consumers (a σ of 2.02). The full model (specification 3) shows the effect standardization had on the value of brands, which represent both network breadth and firm reputation. The σ coefficient on brand is much lower than one (0.70), indicating brand quality became less important post-standardization Again, financial characteristics become more important, and premium has roughly the same impact on choices pre- and post- standardization.

The estimates in both Table 3 and Table 4 make an implicit normalization: the variance of the idiosyncratic error term ε_{ij} remains constant pre- and post-standardization. Appendix Table A.3 takes an alternative normalization, holding constant the vector β, and estimating how the variance of the error term changes; in this case, the variance of ε_{ij} declines from 1 to about 0.7 to 0.8.

The salience model can also be translated into consumer willingness-to-pay (WTP) for various product characteristics. For example, consider the WTP for a gold plan (as opposed to a bronze plan): in 2009, this can be calculated as the coefficient on gold divided by the price sensitivity. This calculation implies that the average consumer in 2009 is WTP \$81/month for a gold (rather than bronze) plan. By contrast, to obtain the same number for 2010, we need to multiply the gold coefficient by the σ for financial characteristics of 1.83 and divide by the new price sensitivity. This calculation implies that the average consumer in 2010 in 2010 is WTP \$149/month for a gold (rather than bronze) plan. This corresponds to an increase in the popularity of gold plans in the latter year.

5 Counterfactuals

5.1 Introduction

In order to conduct counterfactual simulations, we model firm prices in two ways: 1) using a hedonic pricing model, and 2) using an equilibrium pricing model. In the hedonic model, we estimate the premium attached to plans' attributes both pre and post-standardization. For

instance, we estimate that the list price of a silver plan is $126 more expensive than a bronze plan pre-standardization, and an $114 more expensive post-standardization. Thus, when we simulate firms' pre-standardization pricing levels on post-standardization plans, we simply apply the estimated pre-standardization hedonic pricing model to the post-standardization plans. The full model is described in the online appendix.

In the equilibrium pricing model, we assume the observed prices are optimal in each year, given the demand that firms face in that year. We then use our demand estimates to infer what firms' cost of providing each plan is. Then, when we simulate firms' pre-standardization pricing levels on post-standardization plans, we use our implied cost for the post-standardization plans, and derive the optimal price for those plans if the firm was to face pre-standardization demand curves.

There are advantages and disadvantage of each pricing model. The disadvantage of the equilibrium pricing model is that if costs in general rise from 2009 to 2010, we will incorporate that in our simulations, so that the exact same plan could be priced differently. The advantage of the equilibrium pricing model is that, because demand differs pre- and post-standardization, it will models how firms would have actually set prices if they introduced these plans in the other time period. The hedonic pricing model makes no assumptions about costs or firm policies, but also fails to capture the impact of strategic interaction.

5.2 Impact of Standardization on Choice

We first run a counterfactual experiment that attempts to disentangle the supply- and demand- side forces that explain changes in market shares across plans. We focus on the number of consumers choosing bronze policies. Using the model estimated in Table 4 (specification 3), we simulate the percentage of consumers choosing bronze plans pre and post-standardization, using both old and new 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 price vectors. In the second two columns, we use the post-

standardization decision weights ($\boldsymbol{\beta\sigma}_{Post}$) and choice set, again under each price vector.

Table 5 shows the results. The change in premiums contribute little to the change in choices, however the shift in decision weights makes a large difference. In the first two columns, we see that under either set of prices, consumers purchase bronze plans approximately 64% of the time. The second column shows that, under either set of prices, consumers purchase bronze plans only 55% of the time. The roughly 10% fall in the market share of bronze plans is due to demand-side factors, rather than firm pricing or the choice set.

5.3 The Effect of Standardization on Welfare

Did standardization on the HIX improve welfare and make consumers better off? We present our estimates of the change in consumer welfare in Table **??**. In order to assess welfare, we need a welfare criterion. We rely on revealed preference, but present welfare using two different welfare criteria, U_{pre} and U_{post}, since we estimate different preferences pre- and post-standardization. Fortunately, these different welfare criteria give similar results; in the welfare framework of Bernheim and Rangel (2012), one cannot rank the inconsistent choices.

In our evaluation of welfare, we put aside the any effect of increasing the number of options in the choice set. Recall that there are actually more plans post-standardization than pre-standardization. Our welfare analyses hold the number of plans fixed, choosing the 25 most popular plans in 2010 to compare to the 25 plans in 2009. Thus, we potentially underestimate the positive effects of standardization on welfare.

For each welfare criterion $w \in \{pre, post\}$, we measure equivalent variation using the standard formula of Nevo (2001) and McFadden (1999), which is given by:

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

where $\delta_j = \alpha_w p_j + X_j \beta_w$ is the estimated mean utility of plan j, which can 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, since they will differ depending on the welfare criterion used

Row 1 presents our estimate of the total effect of standardization on welfare, which includes a shift in menu, premiums, and consumer decision weights. Using either welfare criterion, we find that standardization improved welfare by 2-3% percent of premiums (or about $100/enrollee per year).

We then decompose increase in welfare into components using counterfactual simulations. First we examine the effect of prices for plans, which differed pre and post standardization. We want to set aside the effect of cost increases that result from either medical inflation or from changes in firms' pricing power (markups). Thus, in Row 2 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 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: at total increase of 10% using hedonic pricing (about $400/enrollee-year) or 5% using equilibrium pricing (about $200/enrollee-year). 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 only 25% as large (equilibrium pricing) or 40-60% as large (hedonic pricing). From this result, we conclude that "nudges" that affect consumer behavior are also likely to affect equilibrium outcomes as firms respond to changing consumer demand.

5.4 Experiment

The standardization on the Massachusetts HIX involved 2 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[9], 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 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

[9]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.

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 in not the levels chosen in our experiments, 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. 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

H4: Decrease the market share of Harvard Pilgrim

Table 7 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 (2012) show that choices 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 four 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 shares of Fallon and Harvard Pilgrim each drop by 4 percentage points. (All these differences are significant with $p < 0.01$, except Harvard Pilgrim which has $p < 0.10$.) Appendix Table

XX verifies these results using a regression framework; controlling for demographics alters point estimates of differences only slightly, but improves precision. A surprising finding is that experimental participants were slightly less likely to choose Tufts in the Post-Stdz. condition, while HIX enrollees were slightly more likely to choose Tufts; this may be an artefact of the high rate of preference for Tufts among experimental participants.

These results are encouraging, as they show that hypothetical choice experiments can 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 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; this interface would not have been possible without the concurrent change in the choice menu.

5.5 Discussion

Given that, in many scenarios, standardization improves outcomes for consumers and firms can capture some of the newly generated surplus, there is an open question of why they did not offer such an assortment of plans initially. We believe that this is due to the central role of information provision in shaping consumer choice. However, other explanations for insurer behavior are equally plausible. The first, firm learning, argues that the market simply may not be in equilibrium yet. 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. Demand will obviously affect pricing (see Ericson and Starc 2012) as well as product offerings. Furthermore, as we explore below, firms may also need to know about the nature of selection. Finally, given that this may be a relatively small proportion of a insurer's book of business, they may not have a huge incentive to perfect their offerings in this particular market.

Second, regulation has a large impact on the offerings of firms in many insurance markets, and especially in this insurance market. In some sense, this is a paper about improved regulation via information provision. Initially, firms were told to submit just two bids for bronze plans. They were to be of the same basic design, with one having prescription drug coverage and one omitting prescription drug coverage (the latter were later abandoned). Underlying this logic was the idea that consumers, contrary to economists, may actually prefer less choice, so that they don't have to compare "apples to lava" (cite: Toolkit Series). Given this view, the standardization may have improved welfare by introducing additional choice while providing additional decision support tools that allowed them to express their preferences. Therefore, it may be the interaction of the choice set and the information provision that led to gains in consumer surplus.

A third possibility is that selection led to the product assortment in the pre-standardization period. It is possible that a low-deductible Neighborhood bronze plan would have attracted a very high cost subset of the population and would not have been profitable. However, it then seems interesting that BCBS would offer such a plan, given that sicker consumers would be drawn to their more extensive network and stronger brand name. Regardless, this highlights the fact that one-shot deviations may or may not be profitable. As a case-in-point, we estimate that the Neighborhood Bronze High plan, which captures a 16% market share in 2010, would capture roughly half that number of consumers, or an 8% market share, under the 2009 preferences.

Finally, we note that the policy change was not without risk. In the bottom panel of Table 5, we see that if the insurers were forced to offer only one plan of each medal -

bronze, silver, and gold - the consumers are made worse off under the 2010 prices regardless of their preferences. Therefore, 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 Conclusion

We analyze a real-world change in "choice architecture" and examine its effect on consumers. In 2010, the Massachusetts Connector standardized the policies available to consumers and changed the way information was presented. We argue that the change altered the decision weights consumers attached to multiple product characteristics. In this market, consumers trade-off financial and network generosity off against premiums. The change in "choice architecture" made financial generosity more salient and network generosity less salient to consumers. As a result, post-standardization consumers chose more financially generous plans (shifting away from HDHP), while simultaneously opting for insurers with narrower networks. We show that this is due to demand-side, rather than supply-side, factors.

Consistent with the behavioral economics literature, we find that well-designed "choice architecture" generates benefits for consumers. Depending on the exact specification, consumer welfare increases 4-10% due to the policy change. However, firms are able to capture some of the surplus associated with the change. As a result, we argue that potential choice architects take firm behavior into account when setting policy.

Table 1: Basic Summary Statistics

	2009	2010
Bronze	0.6135	0.5547
Silver	0.3012	0.3397
Gold	0.0853	0.1056
HDHP	0.4493	0.2858
BCBS	0.1513	0.1573
Fallon	0.2166	0.0775
Harvard Pilgrim	0.1892	0.1344
HNE	0.0471	0.0237
Neighborhood	0.3408	0.4911
Tufts	0.0550	0.1167

Table 2: Hedonic Regressions, List Prices

	(1)	(2)	(3)	(4)
Silver		-141.0***	-146.6***	-146.6***
		(0.426)	(0.535)	(0.534)
Bronze		-267.2***	-275.0***	-272.3***
		(0.450)	(0.597)	(0.831)
2010 Dummy	-14.71***	21.98***	7.268***	7.266***
	(0.735)	(0.320)	(0.759)	(0.757)
Silver*2010			16.37***	16.40***
			(0.887)	(0.885)
Bronze*2010			19.66***	27.62***
			(0.923)	(1.185)
HDHP				-4.093***
				(0.893)
HDHP*2010				-11.75***
				(1.141)
Constant	338.7***	538.0***	542.8***	542.5***
	(0.830)	(0.560)	(0.604)	(0.607)
Fixed Effect		age cat, carrier	age cat, carrier	age cat, carrier
N: Person*Plan	106,940	106,940	106,940	106,940
R-squared	0.514	0.912	0.912	0.912

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

	(1)	(2)	(3)	(4)
	Conditional Logit		Mixed Logit	
Premium*2009	-0.0259***	-0.0274***		
	(0.00126)	(0.00131)		
Change in Premium Coeff, PostStd.	-0.00122**	-0.00119**		
	(0.000554)	(0.000580)		
Premium*2009, Mean			-3.484***	-3.298***
			(0.0490)	(0.0489)
Premium*2010, Mean			-3.442***	-3.326***
			(0.0473)	(0.0466)
Premium*2009, SD			0.189***	0.255***
			(0.0207)	(0.0211)
Premium*2010, SD			0.277***	0.269***
			(0.0246)	(0.0229)
Premium*Age	0.000287***	0.000303***	0.000316***	0.000359***
	(1.79e-05)	(1.84e-05)	(2.13e-05)	(2.45e-05)
Silver*2009	0.362***	0.395***	0.778***	0.856***
	(0.0850)	(0.0927)	(0.101)	(0.102)
Gold*2009	1.350***	1.469***	1.885***	2.151***
	(0.155)	(0.158)	(0.170)	(0.175)
Silver*2010	1.119***	0.689***	1.565***	1.226***
	(0.0907)	(0.0947)	(0.104)	(0.109)
Gold*2010	2.531***	2.168***	2.942***	2.744***
	(0.155)	(0.156)	(0.172)	(0.176)
HDHP*2009		-0.0850		-0.549***
		(0.0690)		(0.0902)
HDHP*2010		-0.832***		-0.900***
		(0.0736)		(0.0740)
N Person*Plan	106,940	106,940	106,940	106,940

Table 4: Estimated σ Change in Decision Weights

	Spec. 1		Spec. 2		Spec. 3	
	Est.	S.E.	Est.	S.E.	Est.	S.E.
Changes in decision weights:						
σ, Premium			1.0237	0.0003	0.9978	0.0000
σ, Financial Characteristics			2.0179	0.0123	1.8344	0.0088
σ, Brand					0.6969	0.0051
Pre-Std. decision weights::						
Premium	-0.0306	0.0000	-0.0276	0.0000	-0.0281	0.0000
Premium*Age	0.000338	1.73E-11	0.000304	1.82E-11	0.000306	1.88E-11
Silver	0.5432	0.0033	0.3192	0.0037	0.3476	0.0036
Gold	1.8512	0.0066	1.1334	0.0083	1.2263	0.0082
HDHP	-0.4121	0.0025	-0.3310	0.0028	-0.3338	0.0026

Table 5: Counterfactual Plan Enrollment: Alternative Decision Weights and Prices

Choice Set and Decision Weight	2009	2009	2010	2010	2009	2010
Prices	2010	2009	2010	2009	Observed	Observed
Bronze	0.69	0.63	0.59	0.52	0.61	0.55
Silver	0.23	0.25	0.26	0.29	0.30	0.34
Gold	0.09	0.12	0.15	0.19	0.09	0.11
HDHP	0.53	0.46	0.44	0.36	0.45	0.29
BCBS	0.16	0.17	0.15	0.16	0.15	0.16
Fallon	0.11	0.14	0.13	0.15	0.22	0.08
Harvard Pilgrim	0.10	0.11	0.11	0.11	0.19	0.13
HNE	0.07	0.06	0.06	0.05	0.05	0.02
Neighborhood	0.52	0.46	0.49	0.44	0.34	0.49
Tufts	0.05	0.06	0.07	0.08	0.05	0.12

Table 6: Welfare

Welfare Criterion	2009	2010
Constant Sized Choice Set (25 Plans)		
Total Welfare Change, in $/month	9.74	7.28
holding 2009 prices constant (eqm.)	16.13	17.98
holding 2009 prices constant (hedonics)	37.42	30.89
Full Choice Set		
Total Welfare Change, in $/month	15.58	14.90
holding 2009 prices constant (eqm.)	21.10	26.56
holding 2009 prices constant (hedonics)	52.25	51.63
Year	2009	2010
Mean Premium Paid, in $/month	352.45	386.70

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

	Treatment			2009	2010
	Pre	Post	Post-Alt	Observed	Observed
Bronze	33%	30%	40%	61%	55%
... Bronze HDHP	29%	13%	27%	45%	29%
Silver	41%	43%	28%	30%	34%
Gold	26%	26%	32%	9%	11%
Premium Paid	$ 472	$ 437	$ 439	$352	$387
Blue Cross	16%	18%	18%	15%	16%
Fallon	5%	1%	6%	22%	8%
Harvard Pilgrim	10%	6%	6%	5%	2%
Neighborhood	43%	59%	63%	34%	49%
Tufts	26%	16%	8%	5%	12%
N Participants	299	307	304	1354	2080

Web Appendix

A.1 Appendix Tables and Figures

Table A.1: Enrollment-Weighted Hedonics

	(1)	(2)	(3)	(4)
Silver		-108.8***	-115.0***	-115.8***
		(2.373)	(2.988)	(2.956)
Bronze		-226.0***	-233.4***	-248.7***
		(2.256)	(2.832)	(3.405)
2010 Dummy	15.01***	15.52***	-1.225	-2.064
	(3.277)	(1.384)	(4.262)	(4.215)
Silver*2010			16.33***	17.43***
			(4.872)	(4.820)
Bronze*2010			20.15***	45.12***
			(4.645)	(5.253)
HDHP				21.76***
				(2.797)
HDHP*2010				-39.70***
				(4.090)
Constant	264.3***	524.9***	532.0***	538.9***
	(3.274)	(2.809)	(3.267)	(3.345)
Fixed Effect		age cat, carrier	age cat, carrier	age cat, carrier
Observations	4,156	4,156	4,156	4,156
R-squared	0.542	0.923	0.923	0.925

Table A.2: Standardization's Effect on Decision Weights, Robustness Checks

	(1)	(2)	(3)	(4)
Premium*2009, Mean	-4.382***	-4.048***	-3.298***	-3.253***
	(0.0657)	(0.0626)	(0.0489)	(0.0517)
Premium*2010, Mean	-4.326***	-4.153***	-3.326***	-3.427***
	(0.0634)	(0.0583)	(0.0466)	(0.0702)
Premium*2009, SD	0.652***	0.737***	0.255***	0.241***
	(0.0472)	(0.0385)	(0.0211)	(0.0201)
Premium*2010, SD	0.744***	0.744***	0.269***	0.305***
	(0.0555)	(0.0514)	(0.0229)	(0.0334)
Premium*Age			0.000359***	
			(2.45e-05)	
Premium*Age*2009				0.000393***
				(2.92e-05)
Premium*Age*2010				0.000289***
				(4.01e-05)
Silver*2009	0.333***	0.425***	0.856***	0.871***
	(0.101)	(0.102)	(0.102)	(0.102)
Gold*2009	0.884***	1.243***	2.151***	2.187***
	(0.168)	(0.176)	(0.175)	(0.176)
Silver*2010	1.212***	0.952***	1.226***	1.203***
	(0.105)	(0.111)	(0.109)	(0.109)
Gold*2010	2.242***	2.133***	2.744***	2.697***
	(0.168)	(0.176)	(0.176)	(0.178)
HDHP*2009		-0.640***	-0.549***	-0.562***
		(0.0894)	(0.0902)	(0.0899)
HDHP*2010		-0.882***	-0.900***	-0.900***
		(0.0745)	(0.0740)	(0.0740)
Observations	106,940	106,940	106,940	106,940

Table A.3: Estimation of Decision Weight Parameters

	Est.	S.E.	Est.	S.E.
Inverse of Var(e_ij)	0.78159	0.04128	0.73279	0.049227
Premium, mean	-0.03322	0.00167	-7.4033	4.4769
Premium, s.d.			0.41246	0.18494
Silver	0.59652	0.096927	1.2713	0.12258
Gold	2.0179	0.17418	2.9375	0.21126
HDHP	-0.44193	0.05583	-0.89428	0.073234
Premium*Age	0.000365	1.74E-05	0.000442	3.67E-05

Table A.4: Welfare, Assuming Optimal Pricing

DW	CS	WC	2010 Hed.	2009 Hed.	2010 Eqm.	2009 Eqm.
2009	2009	2009	-164.876	-183.701	-198.213	-183.701
2009	2009	2010	-136.522	-145.741	-157.873	-145.741
2010	2009	2009	-152.568	-166.839	-180.065	-166.839
2010	2009	2010	-149.437	-163.266	-176.738	-163.266
2010	2010	2010	-138.464	-127.757	-138.464	-114.849
2010	2010	2009	-173.966	-167.575	-173.966	-146.277
2009	2010	2010	-156.009	-147.201	-156.009	-128.779
2009	2010	2009	-155.787	-147.531	-155.787	-131.809

Table A.5: Counterfactuals, Equilibium Pricing Assumption

	DW	2009	2009	2010	2010
	Prices	2010	2009	2010	2009
Bronze		0.628027	0.640462	0.521049	0.539125
Silver		0.245906	0.263839	0.280628	0.302027
Gold		0.126067	0.095699	0.198323	0.158848
HDHP		0.487185	0.44938	0.390213	0.353737
BCBS		0.23457	0.134062	0.228743	0.124928
Fallon		0.064022	0.179307	0.074762	0.198084
Harvard Pilgrim		0.075315	0.122409	0.080389	0.125803
HNE		0.039422	0.070542	0.032534	0.062694
Neighborhood		0.527366	0.42992	0.503725	0.404317
Tufts		0.059305	0.063759	0.079847	0.084173

Table A.6: Welfare, Equilibrium Pricing Assumption

Welfare Criterion	2009	2010
Total Welfare Change, in $/month	9.735336	7.276533
holding 2009 prices constant	16.12641	17.98362
Year	2009	2010
Mean Premium Paid, in $/month	352.4527	386.6988