Journal of Health Economics 50 (2016) 71–85

Contents lists available at ScienceDirect

Journal of Health Economics

journal homepage: www.elsevier.com/locate/econbase

How product standardization affects choice: Evidence from the Massachusetts Health Insurance Exchange

Keith M. Marzilli Ericson *, Amanda Starc

Boston University, Boston, MA and Northwestern University, Evanston, IL

ARTICLE INFO

Article history:
Received 4 May 2016
Received in revised form 6 September 2016
Accepted 9 September 2016
Available online 21 September 2016

JEL classification:
D80
I11
L15

Keywords:
Health insurance
Standardization
Framing
Consumer choice

ABSTRACT

This paper examines the effect of choice architecture on Massachusetts' Health Insurance Exchange. A policy change standardized cost-sharing parameters of plans across insurers and altered information presentation. Post-change, consumers chose more generous plans and different brands, but were not more price-sensitive. We use a discrete choice model that allows the policy to affect how attributes are valued to decompose the policy's effects into a valuation effect and a product availability effect. The brand shifts are largely explained by the availability effect and the generosity shift by the valuation effect. A hypothetical choice experiment replicates our results and explores alternative counterfactuals.

1. Introduction

Effective competition relies on consumers making informed choices. Yet in many contexts, consumers face difficult trade-offs because they have difficulty comparing different products, are overwhelmed by large choice sets, or cannot observe important dimensions of product quality.[1] Market-makers and regulators often seek to help consumers with choice architecture that simplifies complex problems. For example, regulators may require firms to disclose certain types of information about their products, or may create certification schemes that indicate whether a product has met minimal levels of quality.[2] Policy-makers may "nudge" consumers into making different, and potentially better, choices by changing default options, the salience of different attributes, or other aspects of the decision interface.[3]

The complexity of insurance plans make the newly created health insurance exchanges (HIXs) an ideal context to examine choice architecture. We examine a natural experiment: how plan "standardization" on the Massachusetts Health Insurance Exchange (HIX) affected consumer choice. The Massachusetts exchange has taken an active role in shaping the individual insurance market, and "standardization" is the regulator's description of a policy change that took effect on January 1, 2010. This policy change did two things: standardized the financial cost-sharing characteristics of the plans on the exchange, and simplified the website's choice interface as a result. Before the change, 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). After the change, firms were still permitted to set prices and differentiate themselves by brand and physician/hospital network.[4] Critically, the networks remained the same and no other substantive changes were introduced at the same time, allowing us to cleanly identify the effect of changing the choice set and interface. After the policy, only seven distinct cost-sharing plan types were allowed. However, because firms were required to offer all seven designs, the choice menu actually expanded post-standardization, as each firm had previously offered fewer than seven different plans. Because of the new choice menu, the decision interface on the website was simplified and given structure: individuals

* Corresponding author at: Boston University, Boston, MA 02215.
E-mail addresses: kericson@bu.edu (K.M.M. Ericson), amanda.starc@kellogg.northwestern.edu (A. Starc).

[1] On comparison frictions in the Medicare Part D market, see Kling et al. (2012). On choice overload, see Iyengar and Lepper (2000).

[2] On disclosure of information, see Dranove and Jin (2010) regarding gas mileage on cars. On certification of quality, see Jin and Leslie (2003) on restaurant quality.

[3] For default options, see e.g. Choi et al. (2011); for salience, see Chetty et al. (2009) for a field experiment and Bordalo et al. (2012) on consumer choice.

[4] Starc (2014) 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 brand denotes both network breadth as well as less tangible characteristics, such as customer service and trustworthiness.

http://dx.doi.org/10.1016/j.jhealeco.2016.09.005
0167-6296/

could pick a plan design, and then an insurance firm, rather than sorting through a long list of all the insurers and all the plans.

We show reduced-form evidence that the policy change had a substantial effect on the brands and characteristics of plans chosen: consumers who enroll just before and just after the change look similar but make very different choices. Overall, the generosity of plans chosen rose after the policy change: there was a drop in the share of enrollees choosing bronze tier and high-deductible health plans, and average actuarial value of plans chosen rose 4.7 percentage points. 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.

The policy change could have affected plan choice through two major channels: an "availability" and a "valuation" effect. Choices may have changed because the mix of products available changed and the utility-maximizing choice differed between the old and new choice sets: the availability effect. However, the policy itself may have affected consumer decision-making and preferences may be context-dependent (Tversky and Simonson, 1993). We therefore allow for a valuation effect: a shift in decision weights (utility function parameters) as a result of the policy.[5] 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. We allow decision weights on various insurance attributes (deductible, brand, etc.) to vary pre- and post-standardization. In some specifications, we place additional structure on consumer choice, allowing the salience of bundles of product characteristics and/or optimization errors to vary across regimes. We find that decision weights differed significantly after the policy change: post-standardization, consumers placed 3–4 times more weight on cost-sharing characteristics. In addition, we find little evidence of increased price sensitivity: the change is always statistically insignificant, with small increases or decreases depending on specification. We decompose the total effect of the policy change on market shares into components and find that the valuation effect plays a major role in the shift toward more generous plans; changes in relative prices play a minor role.

We then 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. The experiment replicates the effect of observed policy change: consumers choose more generous plans as a result of the policy. In the counterfactual condition, we 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 the choice interface itself matters: it shifts choice and the reported importance of plan attributes independently of the standardization of plan attributes. The experiment shows that if standardization of the plan attributes had not been accompanied by a change in the choice interface, choices would have shifted away from the silver plans toward choices on either extreme (bronze/gold). We interpret this result as the interface helping consumers structure their choices: because the post-standardization choice menu contained many options that were difficult to sort through, individuals may have gravitated toward either the cheapest coverage or the most generous coverage in the absence of a structured choice interface.

[5] Here, standardization may alter decision utility because it changes the salience of product characteristics (e.g. as in Bordalo et al. (2012)), or shifts consumers' attention (DellaVigna, 2009).

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. Health insurance has many dimensions that are difficult for consumers to evaluate – coinsurance, copayments, deductibles, maximum out-of-pocket limits – which lead to consumer confusion (see Abaluck and Gruber, 2011 on Medicare Part D as an example); moreover, individuals misunderstand important aspects of insurance contracts (Bhargava et al., 2015; Handel and Kolstad, 2013).[6] Our paper adds to the growing behavioral public finance literature (e.g. on tax salience, see Finkelstein, 2009, Chetty et al., 2009) and speaks to a critical policy issue. 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 runs 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.

Our paper provides valuable evidence on how choice architecture can have an impact on market outcomes. The policy raised the weight consumers place on cost-sharing characteristics, which may be beneficial given that previous research suggests consumers underweight out-of-pocket costs relative to premiums (Abaluck and Gruber, 2011). By combining structural and experimental analyses, we highlight the mechanisms through which the policy change shifted consumer choice. Our results indicate that the standardization policy expanded choice without increasing choice frictions, providing an example of how choice architecture can improve consumer choice without limiting the number of choices.

The paper is organized as follows. Section 2 describes the policy change in Massachusetts and our data sources. Section 3 discusses the channels through which the policy change might have an effect. Section 4 provides the reduced form impact of "standardization," while Section 5 describes the discrete-choice model and presents the structural estimates. Section 6 conducts counterfactual analyses and discusses welfare implications, and Section 7 describes the hypothetical choice experiment. Section 8 concludes.

2. The Massachusetts HIX

2.1. History and existing literature

The Massachusetts HIX was established by the state's 2006 health reform.[7] We examine the 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. The Massachusetts reform was widely seen as a success: uninsurance rates fell

[6] While most existing health insurance markets do not have standardized plan types, Medigap (Medicare Supplemental Insurance) is an exception. Suggestive evidence from interviews with program administrators indicates that Medigap's standardization reduced consumer confusion (Fox et al., 2003; also see Rice and Thomas, 1992). We are unaware of any work examining the effect of the Medigap standardization on price competition or consumer choice among brands, though Finkelstein (2004) finds that the introduction of minimum standards in the Medigap market reduced the fraction of the population holding such insurance. A well-identified analysis of the Medigap standardization is difficult since many regulations changed simultaneously.

[7] The HIX operated from 2007 to 2013; beginning in 2014 it operated as an ACA exchange, with slightly different regulation.

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, 2012), examined pricing regulation in the presence of imperfect competition (Ericson and Starc, 2015b), and examined consumer valuation of network breadth (Ericson and Starc, 2015a). In Massachusetts, insurers can only price on age, family size, and location; the oldest consumer in a plan can only be charged twice the rate of the youngest consumer. See Ericson and Starc (2013) for more detail on the HIX, and Dafny et al. (2014) for a general description of the effects of competition in exchanges nationally.

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.[8] The Connector Authority running the HIX 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 Series, 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. The Connector Authority saw the standardization policy 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".[9]

The policy change applied to all sales of plans offering coverage beginning January 2010.[10] Prior to standardization, consumers had to evaluate complex financial contingencies for each plan. For example, the Tufts bronze plan had a $2000 deductible, $40 copayments for provider office visits, and full coverage for allergy shots after the deductible had been met. By contrast, the Neighborhood bronze plan had a $2000 deductible, $25 copayments for provider office visits, and full coverage for allergy shots as part of an office visit regardless of whether or not the deductible had been met. The policy change eliminated these discrepancies and 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 the policy change 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. The standardization policy also unbundled the decision making process into one decision about a contract design, followed by a decision about an insurer.

The standardization policy 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.

Standardization did not change the networks within a carrier (a major source of product differentiation), nor did it change the set of carriers offering plans in the exchange.[11] There is no evidence of shifts in marketing activities during this period. Bolstering this point, while we observe major shifts in brand purchased on the HIX, no brand shifts are observed in a market unaffected by the policy change (the subsidized Commonwealth Choice program).[12] No other regulatory changes accompanied standardization,[13] nor were there changes in the consumer population in this narrow time frame.

Fig. 1 shows the plan designs available before and after standardization for each insurer, focusing on each 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 indicate when an insurer's plan design differed slightly from the standardized version. In effect, the standardization policy aligned all the plan options on the grid and filled in the holes in the grid.

2.3. Data and analysis sample

Our dataset is transaction-level data (purchase, cancellation, and payments) from the unsubsidized market (Commonwealth Choice) from the beginning of the HIX's existence in July 2007 until July 2010. There are large spikes in initial enrollment during the first month of the HIX's existence as well as just before the individual mandate took effect in December 2007, with a steady-state enrollment of approximately 1000 households per month.

Our main analyses focus on the subset of the data for which we have detailed price information: Nov. 2009–Feb. 2010. Because our data from the HIX are 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 HIX website using a Perl script. 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.

We exclusively use choices by consumers aged 27–64 entering the exchange for the first time and purchasing individual coverage. We focus on purchasers of individual coverage rather than family coverage since most purchased policies are for individual coverage. We exclude individuals under 27 since they are eligible for a separate, young-adult market. We only use the choices of consumers entering the HIX for the first time, so that inertia does not contaminate our estimate of the effect of the policy change. (While

[8] In the Massachusetts HIX, bronze plans generally had actuarial value within the 60–70% range, silver plans range between 70–90%, and gold plans have actuarial values above 90%. The ACA's definition of tiers differs slightly.

[9] Nancy Turnbull, quoted in Toolkit Series (2011).

[10] That is, consumers buying year-long plans effective in Dec. 2009 did not purchase from a standardized menu, but consumers buying plans effective in Jan. 2010 did.

[11] Major changes in provider networks for this market over the 2009–2013 period were confined to the introduction of limited network products. (Later in 2010, plans could introduce limited network products on the HIX.) See Ericson and Starc (2015a) and testimony from each firm in Massachusetts Health Policy Commission (2013).

[12] See aggregate data in the Massachusetts Health Connector (2010).

[13] During this time period, consumers could enroll in any month. Later in 2011, the HIX shifted to more limited open enrollment periods.

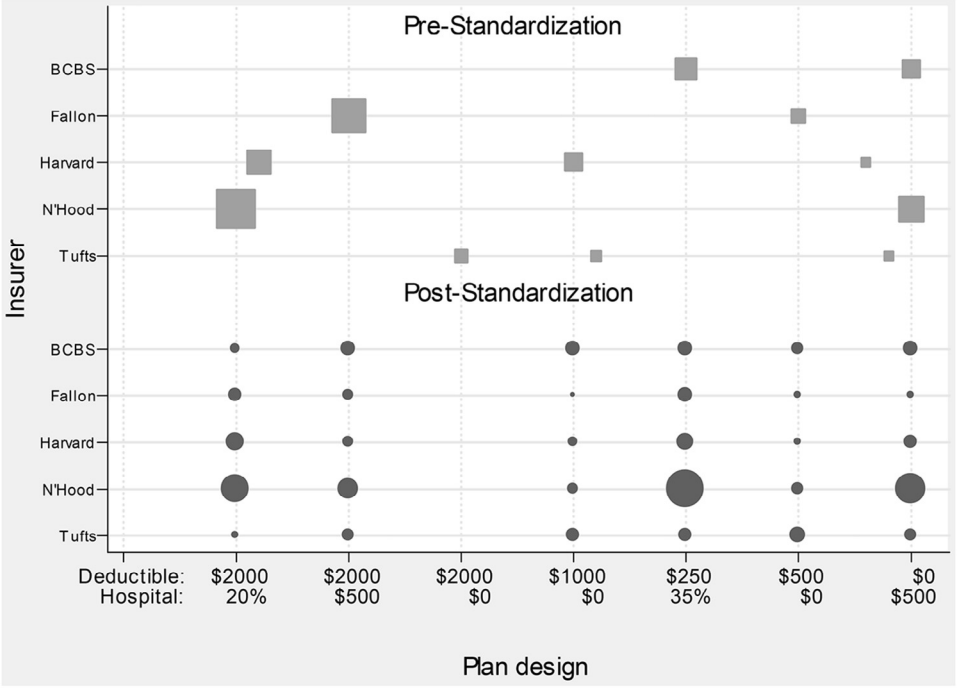

Fig. 1. Plan design and market share before and after standardization. *Notes:* The *x*-axis indicates plan design parameters. All post-standardization plans line up with one of the listed designs, while pre-standardization plans are offset from the vertical line when plans differ slightly in parameters. Size of markers indicates relative market share. Only bronze and silver plans are shown above. Throughout all of the figures, we exclude the insurer Health New England, which does not offer plans in all geographic areas. Plans may vary on financial characteristics beyond deductible and hospital copay pre-standardization. Sample: Analysis Sample.

repeat enrollees are interesting, we don't observe many repeat enrollees due to the level of churn in the HIX. Repeat enrollees are also likely to display inertia in choice, and face a different decision environment.[14]) We include each individual only once in our analysis, leaving us with 982 choices pre-standardization and 1336 choices post-standardization in our Analysis Sample.

While it is not possible to link claims data to the choices of individuals on this HIX, we can provide some insight on the distribution of out-of-pocket costs faced by enrollees in the insurance plans available. First, 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,[15] along with the implied mean and standard deviation of out-of-pocket costs for each plan. We then use the Massachusetts All-Payer Claims (APCD) database to augment our analysis. The APCD has medical claims for Massachusetts healthcare payers, including employer-sponsored plans, Medicaid, small group plans, and the individual markets (both purchased on and off of the HIX). We select a sample of individuals in insurance plans that were purchased on the individual market, for single person coverage. While we cannot identify the particular plans individuals chose, we can estimate the distribution of medical spending in this population, and use that to describe mean and variance of out of pocket costs in plans. (Further details are in the Appendix.)

3. Policy mechanisms

The policy change could have affected 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 allowing for alternative decision processes is necessary to capture the valuation effect.

3.1. Availability effect

Plans are differentiated by brand as well as cost-sharing design. Fig. 1 shows that not all brands offered all the cost-sharing designs pre-standardization. For instance, pre-standardization, only Neighborhood Health Plan offered the Bronze Low plan and only Blue Cross Blue Shield (BCBS) offered the Bronze High plan. The policy change "filled in the grid," requiring Neighborhood to offer Bronze High and BCBS to offer Bronze Low. Consumers can then take advantage of new insurer and plan design combinations. The resulting choice depends on consumers' valuation of each insurer's network.

Consider the choice between Bronze Low and Bronze High. We use APCD data to estimate the distribution of medical spending for an exchange-like population from Massachusetts (details in the Appendix; results are similar if the MEPS is used to construct the distribution). For someone who faced the population distribution of costs, we simulate the mean and variance out-of-pocket costs for both the Bronze Low plan, which we implement as having a $2000 deductible, 20% coinsurance (a simplification), and a $5000

[14] We observe 278 enrollees from the previous year who renew coverage in Jan./Feb. 2010. We are unable to determine whether they faced an automatic reenrollment default (and if so, what plan they were defaulted into). While it is not feasible to estimate choice models on this sample with any precision, we have tested whether the choices of renewing enrollees shifted to different tiers post-standardization. We cannot reject the null hypothesis that distribution across tier chosen initially and at reenrollment is the same (Fisher's exact test, $p = 0.578$). Moreover, in the repeat choosers the probability of choosing bronze had a statistically insignificant increase from 59% to 63% post-standardization, while in the new choosers the probability of choosing bronze declined statistically significantly post-standardization (61% vs. 56%).

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

out-of-pocket maximum, and the Bronze High plan, which we implement as having a $250 deductible, 35% coinsurance (a simplification), and a $5000 out-of-pocket maximum. We add premiums to the out-of-pocket costs to get total annual expected costs. For this discussion, we note that risk aversion does not differentially affect the utility from these two plans as the out-of-pocket maximums are the same and the variance of out-of-pocket costs are quite similar: $1356 and $1224 respectively.

Neighborhood is an inexpensive carrier and BCBS is traditionally a more expensive carrier, so the pre-standardization price gap between the available Bronze Low and Bronze High plans was quite large. For our example consumer, the Neighborhood Bronze Low plan premium is $262 a month while the BCBS Bronze High plan is $381 a month.[16] These monthly premiums, combined with simulated out-of-pocket costs, lead to a total annual expected cost of $4729 for the Bronze Low plan and $5753 for the Bronze High plan. Absent preferences over networks, the Bronze Low plan is a better financial deal except in cases of extreme small scale risk aversion, and the consumer will only choose the BCBS plan if they highly value the larger network or stronger brand name.

Post-standardization, consumers also have the choice of a Neighborhood Bronze High plan, which has substantially lower premiums ($273/month). Now, the lowest cost Bronze High plan has a total annual expected cost of only $4406, albeit with a potentially smaller network. The broader range of choice may lead some consumers to purchase the Neighborhood Bronze High plan, which was not available pre-standardization.

3.2. Valuation effect

The policy change could also lead to a shift in how consumers value plan attributes, which we term the "valuation effect." The valuation effect could arise from many different channels. The valuation effect may arise because preferences are context-dependent (Tversky and Simonson, 1993), or because the standardization policy makes it easier to observe and/or compare plan characteristics. It is difficult to disentangle and separately identify these mechanisms, since they all have the implication that the decision utility function changes.[17] The standardization policy could shift the decision utility function because it alters which heuristics individuals use (see e.g. Ericson and Starc, 2012), because it changes how consumers allocate attention (DellaVigna, 2009) or changes the salience of product characteristics (e.g. as in Bordalo et al., 2012). Alternatively, the change could make it easier to observe certain plan characteristics and thus increase the weight consumers place on them. For instance, if consumers value some dimension (e.g. "quality" vs. price), but observe it with more noise pre-standardization than post-standardization, they will optimally place more weight on that dimension post-standardization.

Plan standardization could also lower 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 Hortacsu 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).[18] In our context, "standardization" does not turn plans into homogenous products – they are still differentiated on network/brand – but does reduce differentiation on the cost-sharing dimension. 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 than a $500 copayment per hospital admission?"). We anticipate that the policy change reduced the difficulty in assessing relative plan generosity.

Finally, standardization policy may have reduced "choice overload." The choice overload literature finds that – in some cases – increasing the number of options available makes consumers worse off. A number of studies document a reduced willingness to choose or reduced satisfaction with choice when faced with a larger set of options, beginning with Iyengar and Lepper (2000).[19] There is a debate about the conditions under which more options lead to choice overload (see Scheibehenne et al., 2010 and responses in that issue). Because the standardization policy provides more structure to the choice process and more comparability between options, it can reduce the cognitive load associated with considering a large number of options. In our context, we find that standardization resulted in consumers being presented with more choices without increasing optimization frictions in the sense of the variance of idiosyncratic preference shocks. Idiosyncratic shocks in a multinomial logit model can be interpreted as unobserved preference shocks, but can also result from noise in consumers' valuations of the underlying good. Of course, by reducing optimization frictions, standardization may have made it easier for consumers to consider and compare a greater number of products from the available choice set. Consumers may have found some of these additional products a better match for their needs; we attribute this to the valuation effect, noting that the actual availability of plans available to this consumer did not change.

4. Reduced-form evidence

4.1. Plans chosen pre- and post-standardization

The reduced-form, pre- and post-comparison shows that the policy change had a significant effect on the level of insurance generosity chosen, as well as on the choice of insurer (brand/network). Throughout our sample period, the networks within a carrier (a major source of product differentiation) remained constant, as did the set of carriers offering plans in the exchange. Table 1 shows that the demographics of enrollees (age, gender, race and income) are similar and not statistically significantly different pre- and post-comparison.

Fig. 1 graphically shows each insurance plan's design and market share before and after standardization, and Fig. 2 gives precise numbers. It is immediately apparent that the standardization policy

[16] We perform this calculation for a 46 years old facing a premium of $262 a month for the Neighborhood Bronze Low plan with the $2000 deductible, $381 a month for the BCBS of Massachusetts plan, and $273 a month for the Neighborhood Bronze High plan with a $250 deductible. To obtain the total annual expected cost of the plan, we simulate expenditures using the APCD data, calculate average out of pocket costs, and add this number to the monthly premium. For the example in the text only, we simplify the plan parameters (which include prescription copays, doctor visit copays, emergency room copays, etc.) to deductible, coinsurance and out-of-pocket maximum. We note that our measures of variance are across individuals in the population rather than within an individual.

[17] Note that the decision utility function may be distinct from welfare-relevant utility. For example, Kahneman et al. (1997) distinguish "decision utility" – the function that rationalizes observed choices – from "experienced utility", the hedonic flow from actual consumption. Standardization may change the decision weights, but is unlikely to change the hedonic flow from insurance plans.

[18] In the health insurance setting, Dafny and Dranove (2008) 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.

[19] In the health plan realm, Bundorf and Szrek (2010), Hanoch et al. (2009), and Besedes et al. (2012) show experimentally that decision making difficulty grows with choice set size. Similarly, Frank and Lamiraud (2009) examine health insurance markets in Switzerland, and find that as the number of choices offered to individuals grows, their willingness to switch plans for a given gain declines. In the investment realm, Iyengar et al. (2004) find that the number of funds in a 401(k) plan is associated with lower participation rates, and Iyengar and Kamenica (2010) find – in both field data and randomized experiments – that when employees face more options in their 401(k), they gravitate toward simpler options.

Table 1
Average enrollee and chosen plan characteristics.

	2009 (Pre)	2010 (Post)	
Actuarial value	77.8	82.5	$p < 0.001$
Expected (Plan OOP cost-sharing)	1129	870	$p < 0.001$
Std. dev. (Plan OOP cost-sharing)	803	752	$p < 0.001$
Monthly premium paid	$374	$389	$p = 0.02$
Demographic characteristics:			
Enrollee age	42.5	43.3	$p = 0.12$
Fraction male	0.48	0.47	$p = 0.54$
Enrollee zipcode: Per capita income	28011	28295	$p = 0.50$
Enrollee zipcode: Fraction white	0.85	0.86	$p = 0.16$
N	982	1336	

Notes: Two sample tests of proportions (binary variables) or t-tests (continuous variables). Sample: Analysis Sample. Actuarial value and mean/standard deviation of out-of-pocket cost-sharing calculated using HHS guidelines for the ACA exchanges. A plan's Expected (Plan OOP cost-sharing) is the mean amount spent OOP by a representative population if they chose that plan, and Std. dev. (Plan OOP cost-sharing) is the standard deviation of the amount spent OOP in the plan. The table presents the average Expected (Plan OOP cost-sharing) and Std. dev. (Plan OOP cost-sharing) across individuals.

leads to a large increase in Neighborhood Health Plan's market share (38.7% to 49.3%), and a large decrease in Fallon's market share (21.1% to 7.6%); both differences are highly statistically significant ($p < 0.001$).[20] We find a marginally significant upward trend in the probability of choosing Neighborhood in the six months preceding standardization. However, as seen in Table A.1, the post-standardization indicator is much larger in magnitude and highly significant. Other insurers also saw shifts in market share to a lesser degree.

Turning to the generosity of health plan, the policy change 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[21]): 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. The mean actuarial value of chosen plans rises from 78% pre-standardization to 83% post-standardization. Similar to the calculation of actuarial value, we can also calculate each plan's expected out-of-pocket (OOP) cost-sharing and standard deviation of OOP cost-sharing that would be faced by a representative population. Both the average expected OOP cost-sharing and average standard deviation of OOP cost-sharing decline post-standardization. Despite these differences in choices, observable demographics are similar across the two time periods.

4.2. Difference-in-difference, trends, and placebo

Our preferred reduced-form estimates come from a comparison of the months immediately prior to and following standardization described above, as it is the period with the most detailed data. Our enrollee age-mix is similar pre- and post-standardization (Table 1). We find no evidence of trends in choice in the six months prior to standardization. Fig. 3 presents graphical evidence for the absence of trends for brand and tier (bronze and silver shown; remaining enrollees are in the gold tier, as there is no platinum tier in this time period). Table A.1 shows corresponding regressions. The pre-trends are all statistically insignificant with the exception of Neighborhood Health Plan, which is significant at the 10% level. Neighborhood Health Plan is also a major player in the subsidized market (see Shepard, 2016 and Jaffe and Shepard, 2016), and changes affecting that market during this time period may have spillovers on brand choice.

We also conduct a difference-in-difference analysis to show that year-end or seasonal effects do not drive our results. Table 2 presents three columns of regressions. Column 1 conducts the simple pre–post comparisons by running regressions of the form $y_i = \beta_0 + \beta_1 PostStd_i$, where *PostStd* is 1 if the observation is post-standardization, and 0 otherwise. Each row runs this regression for a different outcome variable y and displays the coefficient on *PostStd*. The difference-in-difference estimates are contained in column 2, and use data from the analysis sample as well as from the year prior to standardization (Nov.–Dec. 2008 and Jan.–Feb. 2009). These regressions are of the form $y_i = \beta_0 + \beta_1 PostStd_i + \beta_2 JanFeb_i + \beta_3 AnalysisSample$, where *JanFeb* is 1 if the observation is from January or February and 0 otherwise, and *AnalysisSample* is 1 if the observation is from the paper's Analysis Sample (Nov.–Dec. 2009 and Jan.–Feb. 2010). The coefficient β_1 is the difference-in-difference estimate.

The difference-in-difference estimates in column 2 of Table 2 are very similar to the simple pre–post comparison seen in column 1; in fact, the difference-in-difference estimates show a slightly larger effect of the standardization policy. 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 the policy.

Finally, we conduct a placebo test using data from the year prior to standardization (Nov.–Dec. 2008 and Jan.–Feb. 2009), as this time period did not see any major changes to the HIX. Table 2, column 3, shows regressions of the form $y_i = \beta_0 + \beta_1 JanFeb_i$ and presents the coefficient on *JanFeb*. It 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. Moreover, there were 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.

5. Standardization and consumer choice

To distinguish the effects of availability, due to the changes in the choice set, and valuation, largely due to information presentation, we estimate a series of discrete choice models flexible enough to capture a variety of decision processes described in Section 3.

5.1. Discrete choice model and identification

To further explore the effect of the standardization policy, we estimate a discrete choice model that allows context to affect decision-making. Because we are fundamentally interested in how consumers choose plans, we infer valuation of insurance attributes

[20] The shift in brand choice is consistent with standardization leading consumers to choose more "generous" plans, conditional on price. (We explore the sources of this shift in later sections.) As shown in Ericson and Starc (2015a), Neighborhood has a fairly broad network, while Fallon is offering a narrow network plan during this period. Neighborhood is also fairly inexpensive, as it also has a Medicaid managed care product, and therefore has negotiated lower rates with providers than some of the commercial carriers.

[21] We define HDHPs following the tax code as plans with at least a $1200 individual deductible ($1150 in 2009).

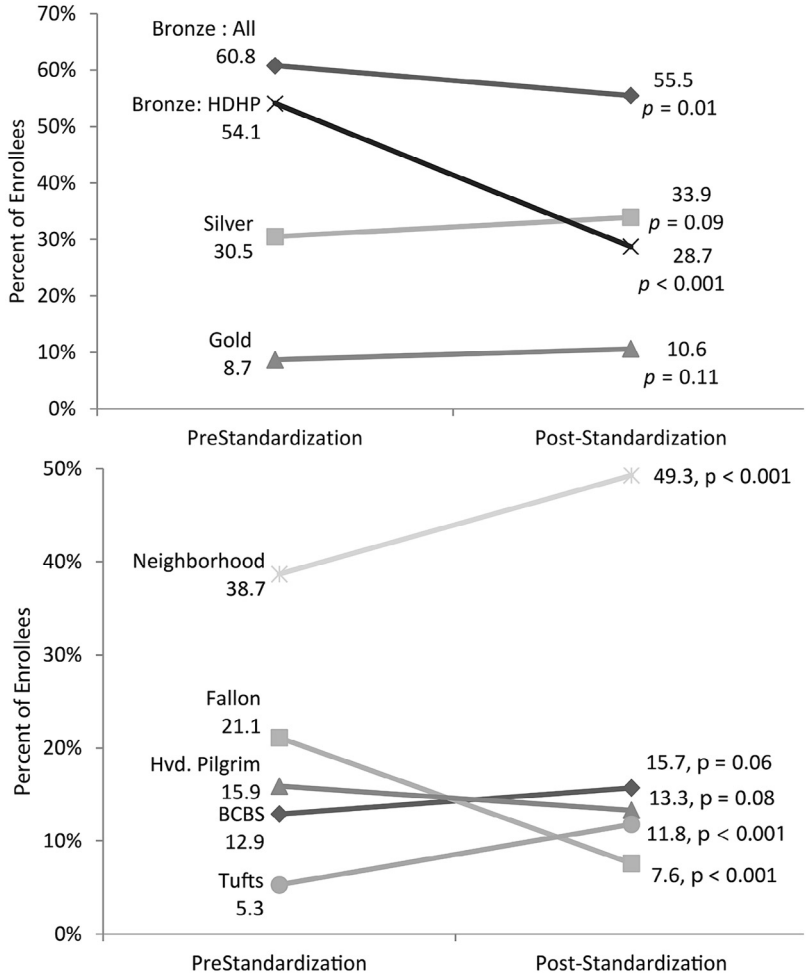

Fig. 2. Enrollee plan choice, pre- and post-standardization. *Notes:* Sample: Analysis Sample. Percentages do not sum to one due to rounding. HDHP is a subset of bronze plans.

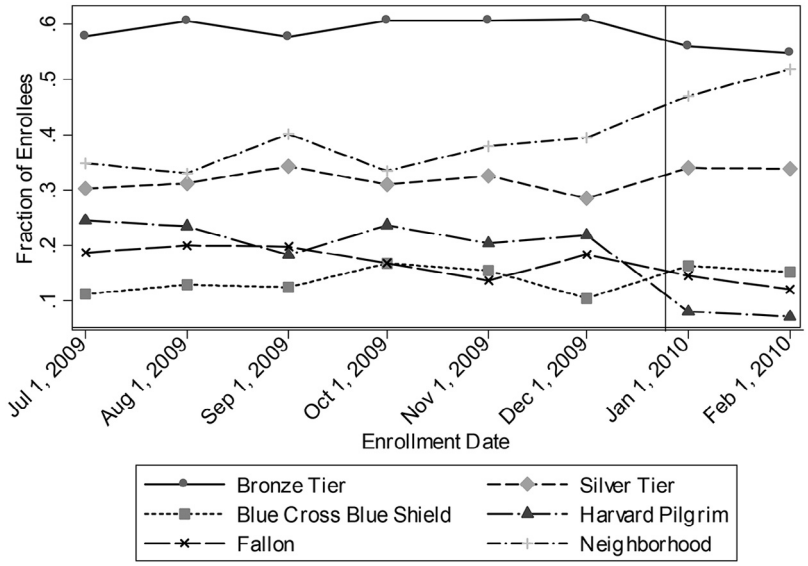

Fig. 3. No evidence of time trends. *Notes:* Shows fraction of new enrollees in each tier and in each of the four largest brands, by month. Sample: Individuals purchasing single coverage, enrolling for the first time in the HIX from July 2009 through February 2010.

Table 2
Difference-in-difference estimates.

	Pre- vs. post-std.		Difference in difference		2008 vs. 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	

Notes: 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- vs. post-std. is the paper's Analysis Sample (Nov. 2009–Feb. 2010). Sample for Placebo is Nov. 2008–Feb. 2009 and otherwise uses the same sample selection criteria (single coverage, first-time enrollees). The difference-in-difference regressions use both samples. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

from their choices.[22] Consumers on the HIX face a discrete choice problem. Products are differentiated based on network and brand, so that a bronze plan from one insurer may be valued differently from a bronze plan from another insurer. In our model, consumers attach decision weights to various plan characteristics, with consumer i's decision utility of plan j given by $u_{ij} = \mathbf{X}_{ij}\beta + \alpha p_{ij} + \varepsilon_{ij}$, where X_{ij} is a vector of plan attributes and p_{ij} is the premium of the plan. Given the assumption that the error term ε_{ij} is i.i.d. extreme value, the probability that consumer i purchases product j is given by the standard logit probabilities. In our approach, we abstract from estimating risk aversion to focus instead on the trade-offs consumers make between product characteristics.

Estimates of price sensitivity are difficult to identify because unobserved plan characteristics may be correlated with price. To accurately identify price sensitivity α, we follow a similar identification strategy that we used in our previous work (Ericson and Starc, 2015a). Our identification strategy makes use of the coarseness in how firms set prices during this time period. Firms do not set a different price for each age: they price in 5-year age bins, so that the premium is constant within the 5-year age bin, but jumps at round-numbered ages ending in 0 or 5 (30, 35, 40, etc.). As a result of this coarse pricing, similar consumers face very different vectors of premiums. While the underlying preferences of a 39-year-old and a 40-year-old will likely be very similar, they will face different premiums for the same plan. To implement this, we allow for an age trend in price-sensitivity that evolves continuously. Then, discontinuities in mean utilities at round-numbered ages are solely attributable to discontinuous changes in premiums, which then allows us to back out α. For example, let δ_{j30} be the mean utility of product j for consumers who are age 30, and δ_{29j} be the mean utility of product j for consumers who are age 29. Mean utilities are implied by market shares. In the absence of age trends, the price coefficient can be simply written as: $\alpha = E_j\left[\frac{\delta_{j30} - \delta_{j29}}{p_{j30} - p_{j29}}\right]$. Of course, we allow for age trends in our empirical specification.

We give a detailed defense of the validity of this identification strategy in Ericson and Starc (2015a), but here note a few relevant facts. The primary concern about identification is that the discontinuity in price setting results from a discontinuity in preferences at these ages. However, firms indicated that they priced coarsely for simplicity, and indeed, later shifted to continuous pricing by age. Moreover, in contexts where prices do not change by age (i.e. employer-sponsored insurance), probability of take-up and plan choice conditional on take-up do not show discontinuities by age, indicating that preferences do evolve continuously.

Identifying the valuation effect 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 conditional on the covariates included in the model. Importantly, we control for actuarial value, or the percentage of the average consumer's expenditure that would be covered by the plan to control for the financial generosity of plans. We believe that dummies for tier and carrier, combined with this measure of generosity, accurately capture all plan characteristics across our entire time period. Tier and AV describe the financial generosity, while carrier captures the network. An examination of plan documents indicates that hospital networks (searchable on the Connector's website) were stable across our time period. Furthermore, the post-standardization plans were modeled after pre-standardization plans. Finally, there can be no differential heteroskedasticity across regimes that requires rescaling of the coefficients (see Train, 2003). In the presence of such heteroskedasticity, the models are only identified up to the constant that rescales all of the deterministic parts of utility.

We explore this assumption in our optimization frictions model, in which we specifically model the heteroskedasticity by allowing the variance of the idiosyncratic error term to change post-standardization. It is likely that any changes in the variance are due to change in consumer decisions post-standardization; as noted in Section 2.2, no other regulatory changes accompanied standardization, nor were there changes in the consumer population in this narrow time frame, as shown in Table 1.

5.2. Results

We seek to explore whether and how consumer valuations differ pre- and post-standardization. To capture this 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.[23]

[22] The alternative method of forming consumers' subjective expected distribution of out-of-pocket costs under each plan, as well as 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, 2011; Barseghyan et al., 2013; Sydnor, 2010). Third, claims data are unavailable for this population. Fourth, plans are still differentiated based on network and brand.

[23] While we estimate the model using our preferred interpretation of product attribute salience and optimization frictions, the theoretical model in Section 3.2 could instead be implemented following Petrin and Train (2010) by allowing for a normally distributed mean zero error, uncorrelated with the idiosyncratic error, with a coefficient that varies with year.

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–α_i]		–3.140***
		(0.0662)
Premium*2010, Mean [ln–α_i]		–3.223***
		(0.0632)
Premium*2009, SD [ln–α_i]		0.340***
		(0.0301)
Premium*2010, SD [ln–α_i]		0.288***
		(0.0250)
Insurer fixed effect	Yes	Yes
N person	2318	2318
N person-plan	70,577	70,577

Notes: 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 lognormally in the population. Sample: Analysis Sample. *** p < 0.01, ** p < 0.05, * p < 0.1

In Table 3, we separately estimate decision weights pre- and post-standardization.[24] Here, with little structure, we allow each decision weight to shift post-standardization. Thus, we estimate $u_{ij} = \mathbf{X}_{ij}(\beta + \mathbf{1}_{Post}\Delta_\beta) + (\alpha + 1_{Post}\Delta_\alpha)p_{ij} + \varepsilon_{ij}$, where α and vector $\boldsymbol{\beta}$ are pre-standardization decision weights and $\mathbf{1}_{Post}$ is a post-standardization indicator. The additive shift in each decision weight post-standardization is given by the vector Δ_β (each element may differ) and Δ_α.

We estimate two different specifications. 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. Next, Column 2 presents a mixed logit specification that also 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 of the premium coefficient is substantial and statistically significant. Given the constraints of the conditional logit model, we want to allow for flexible patterns of choice using random coefficients.

Both specifications show little evidence that the standardization policy 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-standardization. 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. Both of these models include the actuarial value of the plan as a control, and so the coefficients on tier and HDHP should be interpreted as the additional weight consumers place on these characteristics holding constant actuarial value. Table A.2 shows that a similar pattern results even if actuarial value is not included. Table A.3 shows that if the effect of actuarial value is allowed to vary pre- and post-standardization, we also see an increased weight on financial characteristics of plans.

We perform two additional robustness checks in the Appendix. The first, described in Table A.4 limits our analysis to bronze plans. Our data provide detailed information about purchases within the exchange, but little information about the rest of the insurance market. Ericson and Starc (2015b) describe alternative sources of coverage. Here, we also show that our results are not sensitive to the specification of the outside option. In Table A.4 we treat all silver and gold plans as the outside good and show that the pattern of decision weights is similar to the main results. Consumers place more negative weight on both high deductibles and the inside good (bronze plans). Finally, Table A.5 estimates price coefficients pre- and post-standardization within 10-year age bands, which more closely matches the main specifications in Ericson and Starc (2015b). We replicate the main findings in our earlier paper and estimate the change in price sensitivity post-standardization. Consistent with our main estimates, we find no evidence of an increase in price sensitivity post-standardization. If anything, these estimates indicate a slight decrease in price sensitivity post-standardization for consumers around age forty.

5.3. Structured change in attribute valuation

In order to interpret and generalize our findings, we add additional structure for five models in Table 4. These models build on our discrete choice framework and have similar practical interpretations despite slightly different conceptual interpretations. The models vary in their normalizations, restrictions on how decision weights can shift, and assumptions on the error term. Just as in Table 3, each model strongly rejects the hypothesis that the decision utility is constant across years.

We estimate a multiplicative change in decision weight σ, which is constrained to be the same within groups of characteristics: for brand σ_{brand}, for financial cost-sharing characteristics $\sigma_{cost-sharing}$ (tier, actuarial value), for premiums $\sigma_{premium}$, and/or for the error term σ_ε. A general framework for the pre- and post-standardization decision indices can be given by:

$$u_{ijPre} = \mathbf{B}_j\beta_b + \mathbf{F}_j\beta_c + \alpha_i p_{ij} + \varepsilon_{ij}$$
$$u_{ijPost} = \mathbf{B}_j\beta_b\sigma_{brand} + \mathbf{F}_j\beta_c\sigma_{cost-sharing} + \alpha_i p_{ij}\sigma_{premium} + \sigma_\varepsilon\varepsilon_{ij}$$

where $\mathbf{B}_j$ is a matrix of brand indicator variables and $\mathbf{F}_j$ is a matrix with tier indicator variables and actuarial value. Note that we can identify how the policy *changes* decision weights from one context to another, but of course cannot identify context-free fundamental utility

[24] 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.

Table 4
Structured change in decisionweights.

Panel A:						
Model 1			Model2		Model 3	
σ, Error term	0.920	(0.058)	0.898	(0.052)	1	
σ, Cost share	3.54	(0.516)	3.97	(0.536)	4.43	(0.382)
σ, Premium	1		1		1.11	(0.097)
σ, Brand	1		0.869	(0.146)	0.968	(0.155)
Pre-standardization utility index						
Premium	–0.0218	(0.003)	–0.0211	(0.002)	–0.0211	(0.003)
Premium*age	2.25E-04	(3.44E-05)	2.17E-04	(2.89E-05)	2.17E-04	(3.73E-05)
Tier: Silver	0.107	(0.077)	0.0903	(0.048)	0.0902	(0.086)
Tier: Gold	0.356	(0.212)	0.304	(0.129)	0.303	(0.199)
Actuarial value	1.079	(0.451)	0.943	(0.221)	0.942	(0.107)
Fixed effects	Insurer		Insurer		Insurer	

Panel B:							
Model4			Model 5				
σ, Error term	1		Upper nest: choice of tier			Bronze nest	
σ, Cost share	4.54	(0.664)	ρ, Pre	0.276	(0.047)	Premium	–0.0528	(0.003)
σ, Premium	0.986	(0.005)	ρ, Post	0.141	(0.042)	Silver nest		
σ, Brand	0.971	(0.167)	Tier: Silver	0.519	(0.077)	Premium	–0.0422	(0.004)
Distribution of α_i:			Tier: Gold	1.16	(0.076)	Gold nest		
Mean [ln–α_i]	–3.82	(0.081)				Premium	–0.0179	(0.008)
SD [ln–α_i]	0.0228	(0.008)	Lower nests also include premium*age, insurer f.e., and actuarial value					

Notes: Bootstrapped standard errors in parentheses. Model 1 follows Section 3.2, normalizing premium and brand coefficients to be the same pre- and post-standardization. Model 2 normalizes the premium coefficients to be the same pre- and post-, while Model 3 normalizes the error term. Model 4 uses the same normalization as Model 3, but allows additional heterogeneity in price sensitivity α_i. Model 5 presents a nested logit model. AV is dropped when estimating the gold nest of Model 5 due to lack of variation in AV among gold plans. The σ parameters should be interpreted relative to 1, the normalized value for pre-standardization. Actuarial value (AV) is measured on a 0 to 1 scale. $\sigma_{CostShare}$ multiplies tier and AV, while $\sigma_{Premium}$ multiples the premium and the age*premium coefficients. Sample: Analysis Sample. N enrollees = 2318 for all specifications.

parameters.[25] We simply treat σ as the multiplicative change in decision weights post-standardization. When $\sigma > 1$, the decision weight placed on that characteristic increased post-standardization, and when $\sigma < 1$, the weight decreased. Each model must normalize one σ term to be one, as the coefficients are interpreted in relative terms.[26]

In Model 1 we allow standardization to affect the variance of the idiosyncratic error term post-standardization and the weight placed on cost-sharing characteristics, but not the relative weight on other attributes; thus, σ_{brand} and $\sigma_{premium}$ are normalized to 1. We let the weight on cost-sharing characteristics vary as well to capture the valuation effect from standardization enabling clearer comparisons of plan design. The results show that the variance of ε_{ij} is slightly lower post-standardization ($\sigma_\varepsilon = 0.92$), consistent with choices being less "noisy" post-standardization. That is, consumers with identical demographics are more likely to pick the same plan post-standardization than pre-standardization. This can be interpreted as fewer idiosyncratic shocks to preferences or less noise in evaluating product characteristics. Moreover, consumers place more weight on tiers and deductibles post-standardization, even controlling for the actuarial value of the plan. In this model, we estimate that the financial cost-sharing characteristics become much more important after the policy change ($\sigma_{cost-share} = 3.54$), consistent with the valuation effect described in Section 3. Cost-sharing parameters receive more weight in consumer choice *holding constant* the percentage of expenses covered by insurance.

In Model 2 we relax the restriction that the weight on brand preferences did not change post-standardization. Just as in Model 1, the variance of ε_{ij} is smaller and weight on cost-sharing is larger post-standardization. We estimate that the weight placed on brand preferences declined post-standardization ($\sigma_{brand} = 0.87$) but this parameter is measured relatively imprecisely. As the model needs a normalization, we continue to set $\sigma_{premium} = 1$.

We normalize the variance of the error term in Models 3 and 4. Model 3 is a conditional logit model that is identical to Model 2 except for the renormalization, while Model 4 is a mixed logit model that allows for unobserved consumer heterogeneity in premium sensitivity. Both models show a very similar pattern: σ_p is about 1 (and statistically indistinguishable from 1), indicating that the policy did not increase (or decrease) price sensitivity. While Model 3 shows a slight increase in price sensitivity, Model 4 shows a slight decrease once individual heterogeneity is allowed for. This is consistent with the results in Table 3, which found a small, insignificant increase in price sensitivity before consumer heterogeneity was allowed for, but found less price sensitivity (mean and median α closer to zero) in the mixed logit specification. The σ_{brand} is again slightly lower than and indistinguishable from 1. Finally, both specifications show that cost-sharing characteristics become much more important post-standardization ($\sigma_{cost-sharing} > 4$).

In Model 5, we allow the substitutability of products to vary pre- and post-standardization in a nested logit model. After the standardization policy took effect, 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 model, product standardization may lead consumers to view products as closer substitutes and strengthen the degree of price competition as a result.

In the nested logit model, the utility of a plan u_{ij} is decomposed into X_{ij}, 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,

[25] In a model in which context affects the decision weight placed on product characteristics, underlying utility is not easily identified (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 on documenting dominated decisions or modeling things that directly affect utility, like switching costs (Handel, 2013), other papers rely on restrictions from theory (Abaluck and Gruber, 2011).

[26] Suppose $\sigma_\varepsilon < 1$ but that the econometrician did not let σ_ε vary (i.e. imposing that optimization frictions are the same pre- and post-standardization). If the other coefficients are allowed to vary, the estimated coefficients post-standardization will be multiplied by $1/\sigma_\varepsilon$, inflating all the coefficients proportionally.

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.

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}.

The dissimilarity parameter ρ can play a similar role to the σ parameters described above if it is allowed to vary across years. 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 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.

Model 5 presents the nested logit specification. The dissimilarity parameter pre-standardization is much higher than the dissimilarity parameter post-standardization (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 $\sigma_{cost-sharing}$ 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 cost-sharing salience parameter; the decline in the dissimilarity parameter post-standardization implies an increase in $\sigma_{cost-sharing}$, as the larger the ratio of ρ_{Pre}/ρ_{Post}, the greater is the importance of tier. In addition, the nested model implies a higher degree of substitutability between plans within a tier post-standardization. We argue that the standardization policy 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.

All five models estimated in Table 4 tell a consistent story: the policy change altered consumer decision making. Our results show a reduction in idiosyncratic errors, increased weight on cost-sharing parameters, and greater willingness to substitute between carrier brands. By contrast, we find little evidence of increased price sensitivity. Overall, our results show that information presentation can have a large impact on consumer choice. In the following sections, we further explore the availability and valuation effects.

6. Counterfactuals and welfare

6.1. Counterfactual simulations

Given our estimates of decision weights and their change post-standardization (the valuation effect), we turn to counterfactual exercises that allow us to decompose the changes in market shares into availability effect, price effects, and valuation effects. 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 Model 3 estimated in Table 4, we simulate the choices of consumers under a variety of conditions. We separately change the choice set, the decision weights, and the prices. In order to run counterfactuals with prices, we use a hedonic pricing model described in Appendix A.2. 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.[27]

Fig. 4 reports the results of the counterfactuals, plotting the percent of enrollees in each brand or tier by year.[28] The upper left panel shows actual 2009 and 2010 choices. The lower left panel shows the 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; it also contributes to the decrease in Fallon's market share and the market share of HDHP plans.

The valuation effect is shown in the upper right panel. This panel holds constant prices at 2010 levels and uses the 2010 choice set, but simulates choice under pre-standardization decision weights (β) and post-standardization weights ($\sigma\beta$). We see that the valuation effect leads enrollees to shift away from HDHP and bronze plans, and also contributes to the reduction in Fallon's market share. 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 an increase in HDHP and bronze plans.

Decomposing the valuation and availability effects is important: our simulations in which we shift the choice set alone (availability effect) also show that more enrollees would have 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 cost-sharing parameters, such as tier). Similarly, product availability disproportionately increases Neighborhood's market share. Our simulations predict that the market share of Neighborhood plans would be an extremely high 58% 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.

To summarize, demand-side factors and the change in plan offerings due to regulation, rather than firm pricing, are largely responsible for the shifts in tier chosen. 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, rather than the valuation effect.

6.2. Welfare

In our choice sets, there are no dominated plans or choices that cannot be rationalized with utility maximization. Furthermore, normative welfare calculations in models where choices deviate from full rationality or are context dependent are challenging (Bernheim and Rangel, 2009; Cutler and Zeckhauser, 1998). All of our analysis so far has focused on the impact of the policy on choices without discussing the impact of these decisions on consumer welfare. In this section, we discuss possible welfare consequences of the policy change.[29]

[27] Also, as detailed in Ericson and Starc (2015b), an equilibrium model of pricing requires incorporating modified community rating regulation and is outside the scope of this paper.

[28] Appendix Table A.6 provides more detailed counterfactuals in tabular form.

[29] We focus on welfare claims that depend on neither claims data (which necessarily precludes a discussion of selection) nor additional assumptions about consumer rationality or foresight. This would require additional assumptions and linked claims data. For example, a number of papers in the literature would look for changes in overall costs ex post (Abaluck and Gruber, 2011), assuming that consumers had rational expectations or were looking for expenditure minimizing plans holding variance

Fig. 4. Counterfactual simulations of 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. Counterfactuals use the hedonic pricing model and decision weights from Table 4's Model 3.

There is evidence that the restriction of choice in employer-sponsored plans leads to welfare losses (Dafny et al., 2013), while too much choice can be overwhelming and demotivating (Iyengar and Lepper, 2000). This seems to present a challenge to policy makers, including the HIX. Additional choice need not necessarily improve consumer welfare. The fact that some plans, such as the Neighborhood Bronze High plan, are now a " better deal" made consumers better off, but this didn't need to be the case. If consumers face optimization frictions or search costs, additional choice could harm consumers.

However, if a policy change increases choice without creating " choice overload", it can increase welfare. Models 1 and 2 indicate that there is no choice overload here. The policy substantially increased choice, as seen in Fig. 1. At the same time, our results suggest that the intervention helped consumers express their preferences: the smaller variance of the error term can be interpreted as a reduction in optimization errors post-standardization. Moreover, we see from Models 2–4 that consumers place substantially more weight on cost-sharing characteristics post-standardization. While this is not necessarily welfare enhancing, a number of empirical papers (e.g. Abaluck and Gruber, 2011) have argued that consumers underweight cost-sharing characteristics relative to premiums. Therefore, these results are suggestive of an increase in welfare due to the standardization policy.

Given additional assumptions, we can quantify the welfare gain from the change in product mix and optimization frictions. The revealed preference metrics imply a welfare gain of approximately $25/month, depending on the exact assumption. This is consistent with the $323 in potential annual savings from the Neighborhood Bronze High plan described in Section 3.2. Additional derivations, along with the required assumptions, are available in the Appendix. We note that this calculation assumes no moral hazard; to the extent that more generous coverage generates deadweight loss, our welfare gains are an overestimate.

6.3. Discussion

Given that the new plans are popular, why did not firms offer such an assortment of plans initially? At least part of the explanation is that the standardization policy 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. 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.

Our focus is on consumer decision-making, rather than firm behavior. This is largely driven by the institutional environment and

constant. We cannot do an exact calculation for at least two reasons. First, we do not have linked claims data. This does not affect any of our earlier analysis, which describes the impact of standardization on choice. Second, this analysis shows a large impact of information presentation on consumer choice. Therefore, it seems problematic to assume rational expectations while modeling how small changes dramatically impact choice.

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. In addition, we focused only on first-time choosers with no history in the HIX. We are interested in how the standardization policy affects consumer choice; individuals who are inertial and do not make a new choice are, in effect, not exposed to the new policy. 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, 2015b 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 this policy on (adverse) selection across plans.[30] Changes 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.

7. Experiment

The standardization policy on the Massachusetts HIX involved two changes. First – and most important – 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,[31] 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-Stdz." 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 of 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 are 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 a northeastern state (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. Using the change in actual choices on the HIX (Fig. 2 and Table 1), we make predictions about how choices should change 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 change in the choice set and interface 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): that change in the choice set and interface should decrease the fraction of bronze plans chosen, increase the fraction silver plans chosen, and increase the market share of Tufts Health Plan.

Table 5 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 HIX; Ericson and Starc (2012) 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 being that Tufts is relatively more popular among the experimental participants. (Note our experiment's participants were intentionally shown the plan menu for a geographic region in which the smallest insurer, Health New England, was not offered.)

The experiment's results confirm the validity of our design analyzing the actual HIX data, providing evidence that the observed shift in choices was due to the standardization policy rather than some other factor (e.g. a shift in enrollee composition). The experiment verifies 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

Table 5
Experiment: the effect of choice menu and interface.

	Experiment			Observed in HIX	
	Pre	Post	Alt-Post.	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

Notes: Compares choices of participants in the experiment, by condition, alongside observed HIX choices from Analysis Sample.

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

[31] 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.

Table 6
Experiment: importance of plan characteristics by condition.

	Tier	Hospital	MaxOOP	Deduct.	Brand	Premium	Dr. visit
Alt-Post.	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). Constant is mean level in Pre-Stdz. condition. Coefficients on indicator variables for other conditions show changes in importance relative to Pre-Stdz. condition. Sample: Experiment Participants. *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$

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 in 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 artifact 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 Table 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 (Ericson and Kessler, 2013; Kesternich et al., 2013; Krueger and Kuziemko, 2013). The experiment's results confirm 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 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 6's bottom row gives the mean importance rating for each attribute in the Pre-Stdz. condition. 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: the coefficient on an indicator for the Post-Stdz. condition shows that tier increased in importance. The 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 Alt-Post. 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 premium 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 a significant increase in price sensitivity post-standardization.

8. Conclusion

Choice architecture on HIXs matters. The standardization policy led to different choices made on the exchange. Our results show it is important to decompose the change into availability and valuation effects, which we do using a discrete choice model that allows context to affect preferences in structured ways. More generous plans were chosen as a result of the policy, which was a result of consumers placing increased weight on cost-sharing parameters. Different brands were chosen, largely due to changes in what plans were available as a result of the policy, highlighting the stake firms have in choice architecture. The standardization policy expanded choice, but simplifications in the choice interface enabled by standardization helped consumers structure their choices. Our results indicate that despite increasing choice, the policy did not increase choice frictions, in the sense of idiosyncratic shocks to preferences. Our study highlights the potential for regulators to help consumers when making choices from a complicated set of products. Simple shifts both in choice menu and information presentation can have large impacts, shifting choices and improving (or reducing) consumer welfare. However, standardization and other choice architecture interventions face a trade-off: while standardization of plan attributes may help consumers choose in the short-run, it may also limit product innovation in the longer run.

Acknowledgements

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.

Appendix: Supplementary material

Supplementary data to this article can be found online at doi:10.1016/j.jhealeco.2016.09.005.

References

Abaluck, J.T., Gruber, J., 2011. Choice inconsistencies among the elderly: evidence from plan choice in the Medicare part D program. The American Economic Review 101 (4), 1180–1210.
Barseghyan, L., Molinari, F., O'Donoghue, T., Teitelbaum, J.C., 2013. The nature of risk preferences: evidence from insurance choices. The American Economic Review 103 (6), 2499–2529.
Bernheim, B.D., Rangel, A., 2009. Beyond revealed preference: choice-theoretic foundations for behavioral welfare economics. The Quarterly Journal of Economics 124 (1), 51–104.
Besedes, T., Deck, C., Sarangi, S., Shor, M., 2012. Age effects and heuristics in decision making. Review of Economics and Statistics 94 (2), 580–595.
Bhargava, S., Loewenstein, G., Sydnor, J., 2015. Do individuals make sensible health insurance decisions? Evidence from a menu with dominated options. NBER Working Paper 21160.
Bordalo, P., Gennaioli, N., Shleifer, A., 2012. Salience theory of consumer choice. The Journal of Political Economy 121 (5), 803–843.
Bundorf, M.K., Szrek, H., 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., Rebitzer, J., Taylor, L., Votruba, M., 2011. Unhealthy insurance markets: search frictions and the cost and quality of health insurance. The American Economic Review 100 (4), 1842–1871.
Chetty, R., Looney, A., Kroft, K., 2009. Salience and taxation: theory and evidence. The American Economic Review 99 (4), 1145–1177.
Choi, J.J., Laibson, D., Madrian, B.C., 2011. 100 bills on the sidewalk: suboptimal investment in 401(k) plans. Review of Economics and Statistics 93 (3), 748–763.
Cutler, D., Zeckhauser, R., 1998. Adverse selection in health insurance. Frontiers in Health Policy Research 1 (1), 1–31.
Dafny, L., Dranove, D., 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., Ho, K., Varela, M., 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.
Dafny, L., Gruber, J., Ody, C., 2014. More insurers lower premiums: evidence from initial pricing in the health insurance marketplaces. NBER Working Paper 20140.
DellaVigna, S., 2009. Psychology and economics: evidence from the field. The Journal of Economic Literature 47 (2), 315–372.
Dranove, D., Jin, G.Z., 2010. Quality disclosure and certification: theory and practice. The Journal of Economic Literature 48 (4), 935–963.
Ericson, K.M., Starc, A., 2012. Heuristics and heterogeneity in health insurance exchanges: evidence from the Massachusetts connector. The American Economic Review 102 (3), 493–497.
Ericson, K.M.M., Kessler, J., 2013. The articulation effect of government policy: health insurance mandates versus taxes. NBER Working Paper 18913.
Ericson, K.M.M., Starc, A., 2013. Designing and regulating health insurance exchanges: lessons from Massachusetts. Inquiry: A Journal of Medical Care Organization, Provision and Financing 49 (4), 327–338. 2012–2013 Winter.
Ericson, K.M.M., Starc, A., 2015a. Measuring consumer valuation of limited provider networks. The American Economic Review 105 (5), 115–119. <https://www.aeaweb.org/articles.php?doi=10.1257/aer.p20151082>.
Ericson, K.M.M., Starc, A., 2015b. Pricing regulation and imperfect competition on the Massachusetts Health Insurance Exchange. Review of Economics and Statistics 97 (3), 667–682. <http://dx.doi.org/10.1162/REST_a_00514>.
Finkelstein, A., 2004. Minimum standards, insurance regulation and adverse selection: evidence from the Medigap market. Journal of Public Economics 88 (12), 2515–2547.
Finkelstein, A., 2009. E-ztax: tax salience and tax rates. The Quarterly Journal of Economics 124 (3), 969–1010.
Fox, P.D., Snyder, R.E., Rice, T., 2003. Medigap reform legislation of 1990: a 10-year review. Health Care Financing Review 24 (3), 121–138.
Frank, R.G., Lamiraud, K., 2009. Choice, price competition and complexity in markets for health insurance. Journal of Economic Behavior & Organization 71 (2), 550–562.
Handel, B., 2013. Adverse selection and inertia in health insurance markets: when nudging hurts. The American Economic Review 103 (7), 2643–2682.
Handel, B.R., Kolstad, J.T., 2013. Health insurance for humans: information frictions, plan choice, and consumer welfare. NBER Working Paper 19373.
Hanoch, Y., Rice, T., Cummings, J., Wood, S., 2009. How much choice is too much? The case of the Medicare prescription drug benefit. Health Services Research 44 (4), 1157–1168.
Hortacsu, A., Syverson, C., 2004. Search costs, product differentiation, and welfare effects of entry: a case study of SP 500 index funds. The Quarterly Journal of Economics 119 (4), 403–456.
Iyengar, S., Huberman, G., Jiang, G., 2004. How much choice is too much: determinants of individual contributions in 401k retirement plans. Pension Design and Structure: New Lessons from Behavioral Finance 6, 88–95.
Iyengar, S.S., Kamenica, E., 2010. Choice proliferation, simplicity seeking, and asset allocation. Journal of Public Economics 94 (7), 530–539.
Iyengar, S.S., Lepper, M.R., 2000. When choice is demotivating: can one desire too much of a good thing? Journal of Personality and Social Psychology 79 (6), 995.
Jaffe, S., Shepard, M., 2016. Price-linked subsidies and health insurance markups. Working Paper.
Jin, G.Z., Leslie, P., 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., Sorensen, A.T., 2006. Information and consumer choice: the value of publicized health plan ratings. Journal of Health Economics 25 (2), 248–275.
Kahneman, D., Wakker, P.P., Sarin, R., 1997. Back to Bentham? Explorations of experienced utility. The Quarterly Journal of Economics 112 (2), 375–406.
Kesternich, I., Heiss, F., McFadden, D., Winter, J., 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 32 (6), 1313–1324.
Kling, J.R., Mullainathan, S., Shafir, E., Vermeulen, L.C., Wrobel, M.V., 2012. Comparison friction: experimental evidence from Medicare drug plans. The Quarterly Journal of Economics 127 (1), 199–235.
Kolstad, J., Kowalski, A., 2012. The impact of health care reform on hospital and preventative care. Journal of Public Economics 96 (11), 909–929.
Krueger, A., Kuziemko, I., 2013. The demand for health insurance among uninsured Americans: results of a survey experiment and implications for policy. Journal of Health Economics 32 (5), 780–793.
Massachusetts Health Connector. 2010. Annual report. Online at <https://www.mahealthconnector.info/portal/binary/com.epicentric.contentmanagement.servlet.ContentDeliveryServlet/Health> (accessed 01.10.13.).
Massachusetts Health Policy Commission. 2013. Annual cost trends hearing. Online at <http://www.mass.gov/anf/budget-taxes-and-procurement/oversight-agencies/health-policy-commission/annual-cost-trends-hearing/2013/testimony-and-presentations/pre-filed-testimony-from-witnesses.html> (accessed 01.10.13.).
Petrin, A., Train, K., 2010. A control function approach to endogeneity in consumer choice models. Journal of Marketing Research 47 (1), 3–13.
Rice, T., Thomas, K., 1992. Evaluating the new Medigap standardization regulations. Health Affairs 11 (1), 194–207.
Scheibehenne, B., Greifeneder, R., Todd, P.M., 2010. Can there ever be too many options? A meta-analytic review of choice overload. Journal of Consumer Research 37 (3), 409–425.
Shepard, M., 2016. Hospital network competition and adverse selection: evidence from the Massachusetts Health Insurance Exchange. Working Paper.
Sorensen, A.T., 2000. Equilibrium price dispersion in retail markets for prescription drugs. The Journal of Political Economy 108 (4), 833–850.
Starc, A., 2014. Insurer pricing and consumer welfare: evidence from Medigap. The Rand Journal of Economics 45 (1), 198–220.
Sydnor, J., 2010. (Over)Insuring modest risks. American Economic Journal: Applied Economics 2 (4), 177–199.
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., Simonson, I., 1993. Context-dependent preferences. Management Science 39 (10), 1179–1189.