Estimating the Heterogeneous Welfare Effects of Choice Architecture: An Application to the Medicare Prescription Drug Insurance Market

By JONATHAN D. KETCHAM, NICOLAI V. KUMINOFF AND CHRISTOPHER A. POWERS*

We develop a structural model for estimating the welfare effects of policies that alter the design of differentiated product markets when some consumers may be misinformed about product characteristics. We use the model to analyze three proposals to simplify Medicare markets for prescription drug insurance: (1) reducing the number of plans, (2) providing personalized information, and (3) changing defaults so consumers are reassigned to cheaper plans. First we combine national administrative and survey data to determine which consumers appear to make informed enrollment decisions. Then we analyze the welfare effects of each proposal, using the revealed preferences of informed consumers to proxy for the concealed preferences of misinformed consumers. Results suggest that the menu reduction would harm most consumers whereas personalized information and reassignment would benefit most consumers. Each policy produces large gains and losses for small groups of consumers, but no policy changes average consumer welfare by more than 19% of average expenditures.

Preliminary Draft for Seminar Presentation

February 2016

* Ketcham: Arizona State University, Department of Marketing, and Box 874106, Tempe, AZ 85287-4106 (e-mail: Ketcham@asu.edu). Kuminoff: Arizona State University, Dept. of Economics and NBER, Tempe, AZ 85287 (e-mail: kuminoff@asu.edu). Powers: U.S. Department of Health and Human Services, Centers for Medicare and Medicaid Services, Center for Strategic Planning / DDSG, 7500 Security Boulevard, Mailstop C3-24-07, Baltimore, MD 21244 (e-mail: Christopher.Powers@cms.hhs.gov). Ketcham and Kuminoff's research was supported by a grant from the National Institute for Health Care Management (NIHCM) Research and Educational Foundation. The findings do not necessarily represent the views of the NIHCM Research and Education Foundation. We are grateful for insights from Gautam Gowrisankaran, Alvin Murphy, Sean Nicholson, Jaren Pope, Dan Silverman, Meghan Skira, V. Kerry Smith, and seminar audiences at the AEA/ASSA Annual Meeting, the Congressional Budget Office, Health and Human Services Office of the Assistant Secretary for Planning and Evaluation, the ASU Health Economics Conference, the Annual Health Economics Conference, the Quantitative Marketing and Economics Conference, Brigham Young University, University of Arizona, University of Southern California, the University of Miami, Cornell, Vanderbilt, and Yale.

One of the frontiers in empirical microeconomics is to assess the equity and efficiency of policies that modify *choice architecture*—the design of a market environment in which people make decisions (Thaler and Sunstein 2008). Examples of policies designed to simplify choice architecture include restricting the number of options consumers can choose from, providing people with personalized information via decision support tools, and altering their default options. Understanding how choice architecture affects consumer welfare is increasingly important for public policy. Over the past few years the United Kingdom, the United States, the World Bank and several other government organizations have created "nudge units" that aim to use insights from behavioral economics and psychology to refine their programs. For example, in September 2015 President Obama issued an executive order directing the newly created US Social and Behavioral Sciences Team to help federal agencies review their programs and look for ways to improve consumer welfare by simplifying choice architecture.[1]

Policies designed to simply choice architecture are often hypothesized to benefit consumers who are misinformed about their options and to potentially harm those who are informed (Camerer et al. 2003). Yet, using revealed preference analysis to estimate the distribution of welfare effects presents several empirical challenges. First, the researcher must identify the decision makers and determine which of their decisions are based on misinformation. This is complicated by the fact that important financial decisions may be made in part or in whole by a spouse, a relative, or a paid advisor. Second, a method is needed to infer the preferences of misinformed consumers. Third, researchers must predict how a counterfactual policy would affect consumer behavior and market prices. Finally, all of this information must be aggregated into theoretically consistent welfare measures.

We develop a framework for addressing these challenges in differentiated product markets and use it to evaluate the distribution of welfare effects of recent proposals to simplify markets for prescription drug insurance in Medicare. Our research builds on recent empirical studies that have refined standard revealed preference methods to address heterogeneity in information (e.g. Allcott and Kessler 2015, Allcott and Taubinsky 2015, Bernheim et al. 2015, Handel 2013, Handel and Kolstad 2015). These studies aim to recover preferences and welfare by leveraging field experiments, laboratory experiments, quasi-natural experiments, and surveys to distinguish be-

[1] Section 1(b)(iii) of executive order 13707 directs U.S. agencies to "identify programs that offer choices and carefully consider how the presentation and structure of those choices, including the order, number, and arrangement of options, can most effectively promote public welfare, as appropriate, giving particular consideration to the selection and setting of default options."

tween active and passive choices made by consumers who are assumed to differ in their knowledge of markets. Their applications utilize data from online sample frames or workers at a few firms who chose among a small number of options. We extend this literature in several ways and provide the first national analysis of proposals to simplify choice architecture in a high-stakes differentiated product market that is both subsidized and regulated by the government.

Medicare Part D created a nationwide government-designed, taxpayer-subsidized series of markets for standalone prescription drug insurance plans (PDP). In 2013, the PDP markets enrolled 23 million people with federal outlays of $65 billion (US Department of Health and Human Services 2014). Due to concerns about consumer confusion, researchers and federal agencies have proposed several reforms designed to simplify choice architecture (e.g. McFadden 2006, Thaler and Sunstein 2008, Federal Register 2014). These include restricting the number of insurance plans that can be sold, providing consumers with customized information about their options, altering default rules and reassigning people to plans. Similar policies have been implemented in employer-sponsored retirement savings programs and proposed for the health insurance exchanges implemented under the Patient Protection and Affordable Care Act (ACA).

We assess the heterogeneous benefits and costs of these prospective policies by drawing on a novel combination of administrative records and survey data on consumers' knowledge of the market, their enrollment decisions, and the financial consequences of those decisions for a national sample of the non-poor Medicare PDP enrollee population from 2006-2010. We worked with CMS to link the annual responses given by people who participated in their longitudinal Medicare Current Beneficiary Survey (MCBS) to administrative records on the insurance enrollment decisions made by those individuals in Part D, as well as the universe of their drug claims, their demographic characteristics, and their evolving chronic medical conditions.

Although the survey and administrative data sets have been analyzed separately by prior studies, we believe this is the first time they have been linked by researchers. Linking the two data sets is important for our purposes because the MCBS tests respondents' knowledge of key market institutions and asks about the effort they exerted to search for information. Equally important, the MCBS allows us to determine whether each enrollee made health insurance decisions on their own or had help from somebody else. This information allows us to disentangle several potentially important mechanisms that are typically conflated as part of unobserved heterogeneity in models of consumer decision making, substantially improving our ability to apply

revealed preference logic to investigate how knowledge and decision making relate to consumer demographics. Given the large amounts of money at stake and the age range of the eligible population, it is unsurprising to find that 38% of enrollees do not make health insurance decision on their own. Our MCBS data show that enrollees are more likely to get help choosing a plan if they are older, sicker, lower income, less educated, less internet savvy, or diagnosed with depression or dementia (including Alzheimer's disease).

We use the linked data to isolate a subset of enrollment decisions that we suspect will not reveal the consumer's preferences in a standard econometric model of insurance plan choice because the decision maker appears to be misinformed about the market and her beliefs cannot be fully observed. In contrast, we rely on the conventional assumption of full information in the absence of evidence to the contrary. Specifically if an enrollee correctly answers an MCBS question testing her knowledge of the market <u>and</u> chooses a plan that can be justified as maximizing a utility function satisfying standard axioms of consumer preference theory, then we treat her decision as providing information about her preferences. We use the nonparametric GARP-like test from Ketcham, Kuminoff, and Powers (2015) to identify choices that cannot be explained as maximizing utility under full information. As in Chetty et al. (2015) and Handel and Kolstad (2015) we distinguish between active and passive choice processes and we recognize that people may incur search costs and hassle costs from switching insurance plans. We find that 56% of enrollees answer the knowledge question correctly and make choices that obey standard axioms of consumer theory. The probability of being in this informed group increases with education and with the effort that people exert to learn about the market. The probability decreases as people age, as they are diagnosed with cognitive illnesses, as their drug expenditures increase, and as they rely on other people to help them make health insurance decisions.

We then estimate multinomial logit models of the annual enrollment process separately for the apparently informed and misinformed choices.[2] If taken literally, the model for misinformed choices would imply that those consumers are risk-loving and have an average willingness to pay to avoid switching out of their status quo insurance brands of nearly $3660.[3] The parameters for

[2] We distinguish between active and passive choices, and model inertia, but our framework is static in that consumers do not forecast future changes in the menu of insurance plans and their prices. This approach is consistent with prior studies (e.g. Handel 2013, Handel and Kolstad 2015). A static framework seems appropriate here because it is difficult for consumers to forecast their own future health shocks, let alone the future health shocks and enrollment decisions of other consumers together with the implications for market prices in a fluid regulatory environment.
[3] Inertia can arise in health insurance markets from at least three mechanisms with potentially differing implications for the welfare effects of counterfactual policies. Inertia may reflect consumers' preferences for latent features of their preferred insurers and plans (Ketcham, Kuminoff, and Powers 2015); it may reflect hassle costs of switching plans (Handel and Kolstad 2015); and it may reflect a variety of psychological biases such as status quo bias and procrastination (Kling et al. 2012). We explore the implications of various interpretations in our empirical analysis.

informed choices, on the other hand, imply far less inertia and risk aversion consistent with the prior literature (e.g. Cohen and Einav 2007, Handel 2013, Handel and Kolstad 2015).[4] Following the approach to welfare analysis proposed by Bernheim and Rangel (2009), we infer the preferences for consumers in the misinformed group from the choices made by consumers in the informed group conditional on health status, drug use, and a rich set of demographic characteristics. The key assumption is that information is uncorrelated with preferences after conditioning on demographics, health, and drug use. Then we build on Small and Rosen (1981) to adapt the standard multinomial logit welfare framework to allow for the possibility that misinformed consumers could be made better off by a policy that provides information or even reduces the number of choices.[5] We use the computational approach from Bayer and Timmins (2005) to allow equilibrium plan premiums to adjust following a policy in response to consumer sorting and adverse selection. This framework allows us to consider how the welfare effects of prospective policies targeting choice architecture vary across people with different drug needs, demographics, and choice processes.

In the first policy experiment, we simulate the federal government's recent proposal to limit each insurer to sell no more than two plans per market (Federal Register 2014). In the second experiment, we calibrate our model to replicate a field experiment conducted by Kling et al. (2012) in which Part D enrollees were sent personalized letters with information on the amount of money they could expect to save by switching to their lowest-cost plans. In the third experiment, we simulate the federal government's proposal to alter the default to automatically reassign people to low-cost plans (Health and Human Services 2014). We find that all three policies have winners and losers. Our results suggest that the CMS proposal to limit the number of plans would reduce welfare for at least two thirds of consumers and effectively operate as an income transfer from consumers and taxpayers to insurers. Further, the insurers have strong incentives for regulatory capture as they could increase the size of transfer payments substantially if the regulations permit them to choose which plans to retain. In contrast, we find that providing personalized information would benefit more than two thirds of consumers and would reduce government spending. Similarly, assigning people to low-cost default plans would reduce government ex-

[4] For example, our estimates imply that consumers in the informed group would be indifferent between a 50-50 bet of winning $100 and losing between $91.6 and $97.3. Variation within this range is explained, in part, by our finding that risk aversion is increasing with education.

[5] Our approach is equally applicable to neoclassical and behavioral models of consumer decision making. Some consumers may optimally choose not to become fully informed because it is costly to collect information and conduct transactions (Stigler and Becker 1977); others may be influenced by psychological biases (Kahnemann, Wakker, and Sarin 1997).

penditures and benefit more than two thirds of consumers under the assumption that it is costless for them to opt out. The average consumer would be better off from this policy as long as annualized opt out costs are less than $73.

All of these findings are stable across a range of different approaches to identifying which choices reveal consumers' preferences. Likewise, our results are robust to extreme assumptions about the underlying causes of consumer inertia and about the extent to which each policy would nudge consumers to adjust their behavior. In particular, the share of consumers who benefit from each policy and the breakeven opt out cost of default assignment are bounds on ranges that we obtain by running our model under extreme "most effective" and "least effective" assumptions about the effects of policies targeting choice architecture. At one extreme, we assume that the policies cause misinformed consumers to change their behavior to mirror observationally similar consumers who are informed. This scenario also assumes that the inertia we observe for informed consumers is irrelevant for evaluating their change in welfare because inertia is caused by psychological biases. At the opposite extreme, we assume that the government policies would not change consumer behavior and that inertia by informed consumers reflects the hassle cost of switching plans and/or their utility from latent but welfare relevant features of their preferred insurance plans. We view the fact that our policy conclusions are robust to these two extreme cases as informative because reality probably lies somewhere in between.

The remainder of the paper is organized as follows. Section I provides relevant background on Medicare Part D and prior studies of consumer decision making. Section II describes our data and Section III explains how we used it to identify choices that we suspect are based on misinformation. Section IV presents our parametric model of drug plan choice and Section V uses it to derive measures of consumer welfare for policies that modify choice architecture. Section VI presents results from logit models of drug plan choice for informed and misinformed consumers. Section VII presents our main results from the counterfactual policy experiments and section VIII summarizes robustness checks. Concluding remarks are provided in Section IX.

I. Medicare Part D

People typically become eligible for Medicare benefits in the US when they turn 65. In 2006 Medicare Part D extended these benefits to include prescription drug insurance sold through standalone PDPs. Prior to the ACA, Part D was the largest expansion of public insurance pro-

grams since the start of Medicare. A novel and controversial feature of Part D is that it created a quasi-private marketplace for delivering insurance, serving as a precursor to the markets created by the ACA. Part D created 34 spatially delineated markets within which the average enrollee chose among 50 drug plans sold by 20 private insurers. Subject to CMS approval, private insurers can sell multiple PDPs in each market. The default for new or uninsured Medicare beneficiaries is to be uninsured.[6] After an enrollee chooses a plan she is automatically reassigned to that same plan the following year unless she chooses to switch to a different one during the annual open enrollment window. Enrollees pay monthly premiums as well as out of pocket (OOP) costs for the drugs they purchase and taxpayers subsidize the total costs of non-poor enrollees by an average of 75.5%.

PDPs differ in terms of premiums, OOP costs of specific drugs, and measures of quality such as customer service, access to pharmacy networks, the ability to obtain drugs by mail order, and the prevalence and stringency of prior authorization requirements.[7] The novelty of the market together with the complexity of the product led many analysts to speculate that consumers would not make informed choices. Liebman and Zeckhauser (2008) summarize this concern when they write that: "Health insurance is too complicated a product for most consumers to purchase intelligently and it is unlikely that most individuals will make sensible decisions when confronted with these choices." Some analysts suggested that Medicare Part D is a prime candidate for libertarian paternalism (e.g. McFadden 2006, Thaler and Sunstein 2008). Likewise the government has expressed a desire to simplify health insurance markets and nudge enrollees toward cheaper plans. In 2014, for example, CMS proposed limiting insurers to selling no more than two plans per region, which would reduce the average consumer's number of choices by about 20% (Federal Register 2014). The US Department of Health and Human Services also announced that it is considering revising the design of the federal health insurance exchanges to automatically reassign people to low-cost plans unless they choose to opt out (Health and Human Services. 2014). The welfare effects of these types of policies depends on several factors including consumers' preferences for PDP attributes, the cost of switching plans, and how the policies affect consumers' decision processes and outcomes.

Several prior studies have investigated the role of information and consumer behavior in Med-

[6] Enrollees who qualify for low-income subsidies are autoenrolled to certain plans, but we exclude them from our empirical analysis.
[7] Many insurers require consumers to have prior authorization from a doctor in order to obtain certain drugs, but the stringency of these requirements differs from insurer to insurer.

icare Part D. Over the first five years of the program, the average enrollee could have reduced their annual expenditures (premium + out of pocket) by $341, which is equivalent to 25% of average expenditures (Ketcham, Lucarelli and Powers 2015), yet the implications for consumer decision making remain ambiguous. When enrollees are surveyed about their experiences in Part D most report being satisfied with the plans they chose (Heiss, McFadden and Winter 2010, Kling et al. 2012). Furthermore, Ketcham, Kuminoff and Powers (2015) demonstrate that most of the people who could have saved money by switching plan had chosen plans that were either superior in some measure of quality or provided greater protection from negative health shocks. Hence, one explanation for why people leave money on the table is that they are making informed decisions to pay for quality and risk protection. On the other hand, when Kling et al. (2012) asked 406 Wisconsin enrollees how much they thought they could save by switching plans, most respondents underestimated the true figure. Kling et al. also found that sending enrollees a letter with personalized information about their potential savings increased the rate at which enrollees switched plans by 11.5 percentage points. Overall, the existing evidence suggests that some consumers are misinformed, but others may be choosing to pay more for plans with higher quality and/or greater risk protection.

II. Linking Administrative Records to Enrollee Surveys

We collaborated with CMS to link administrative records on PDP enrollees to their responses in the Medicare Current Beneficiary Survey (MCBS). This is the first time the two data sets have been linked for research. Our data collection process began by using CMS administrative records on beneficiaries' basic demographic characteristics, their prescription drug claims, the set of PDPs available to them, and their actual plan choices over the first five years of Part D. Then we used their drug claims to calculate what each enrollee would have spent had they purchased the same bundle of drugs under each alternative PDP in their choice set. This was done by combining their actual claims with the cost calculator developed in Ketcham, Lucarelli and Powers (2015).[8] Next we used administrative data from CMS's Chronic Condition Data Warehouse to determine if and when each individual had 16 different medical conditions. This includes dementia and Alzheimer's disease, which are associated with diminished cognitive performance

[8] We have confirmed the calculator's accuracy by finding a correlation of .92-.98 each year between the OOP prescription drug costs calculated for the actual plan and the actual OOP cost observed in the administrative data.

(Agarwal et al. 2009).[9] Like prior studies of PDP choice we limit our analysis to enrollees who chose a standalone PDP and did not receive a low-income subsidy.[10]

TABLE 1—SUMMARY STATISTICS FOR THE MCBS-ADMINISTRATIVE SAMPLE

	2006	2007	2008	2009	2010
number of enrollees	1,908	2,711	3,143	3,617	4,250
Medicare Current Beneficiary Survey					
high school graduate (%)	78	78	79	80	80
college graduate (%)	20	20	21	23	24
income>$25k (%)	52	52	52	56	56
currently working (%)	14	14	13	13	13
married (%)	56	53	54	55	54
has living children (%)	93	93	93	93	92
uses the internet (%)	31	32	34	37	38
searched for CMS info: internet (%)	21	22	25	27	28
searched for CMS info: 1-800-Medicare (%)	28	24	19	15	12
makes own health insurance decisions (%)	63	61	62	63	62
gets help making insurance decisions (%)	27	27	26	26	27
insurance decisions made by proxy (%)	10	12	12	11	11
CMS Administrative Data					
mean age	77	77	78	78	78
female (%)	62	63	62	63	63
white (%)	93	92	93	93	93
dementia including Alzheimer's (%)	6	8	9	10	12
depression (%)	8	8	10	11	11
mean number of drug claims	28	34	36	35	36
mean number of available plans	43	56	55	50	47
mean number of available brands	20	24	23	23	20
has a default plan (%)	0	84	91	86	84
switches out of the default plan (%)	0	8	12	11	9
active enrollment decisions (%)	100	23	20	23	21
mean premium ($)	330	359	406	475	509
mean out-of-pocket costs ($)	690	844	864	903	925
mean potential savings, ex post ($)	460	349	294	330	336

Note: The table reports means for key variables for the sample of Medicare Part D enrollees found in both the MCBS and cost calculator samples in the given year. See the text for details.

[9] In addition to an indicator for Alzheimer's or dementia, we observe diagnoses of depression, acute myocardial infarction, atrial fibrillation, cancer, cataracts, heart failure, chronic kidney disease, chronic obstructive pulmonary disease, diabetes, glaucoma, hip/pelvic fractures, ischemic heart disease, osteoporosis, and strokes/transient ischemic attack.

[10] We exclude those receiving "low income subsidies" because they are autoenrolled into plans, they receive larger premium subsidies, and their copayments are much more uniform across plans. Hence while interesting to study for policy, they are less relevant for our evaluation of prospective policies designed to alter choice architecture. Despite excluding them, our sample has similar income levels to the national average of people age 65 and above. In our sample 54% of households have annual income over $25,000 (weighted 2006-2010 dollars), compared with 63% (constant 2010 dollars) based on all householders 65 and older in the 2010 Census American Community Survey.

Finally, we worked with CMS to link the administrative data with supplementary information on PDP enrollees who also participated in the MCBS from 2005-2011. The MCBS is a national rotating panel questionnaire that began in 1991 and is administered to approximately 16,000 people annually.[11] It collects information about Medicare beneficiaries and their use of health care services. Each participant is interviewed up to three times per year for four consecutive years, regardless of whether they stay at the same address or move into and out of long term care facilities. Importantly for our purposes, participants are asked a series of questions designed to test whether they understand key features of the PDP market. These knowledge questions are explained below. The MCBS also asks participants if and how they searched for information about Medicare services and it provides richer data on enrollee demographics than the CMS administrative files. This includes variables describing income, education, marital status, employment status, and enrollees' use of the internet. Also of particular value for our study, the MCBS indicates whether a proxy responded to the survey, including the knowledge questions, and whether the beneficiary makes health insurance decisions on her own, with help from someone else, or whether someone else makes decisions for her.

Our linked sample includes 5,269 individuals who made 15,629 annual enrollment decisions between 2006 and 2010.[12] Table 1 reports means of the key variables. The typical enrollee is a retired high school graduate with living children. Approximately 22% are college graduates, 54% are married, and 54% have annual pre-tax household incomes over $25,000. Only 35% report that they ever personally use the internet to get information of any kind. However, among those who do use the internet most have used it to search for information on Medicare programs (25%). Another 18% searched for information by calling 1-800-Medicare.

The average beneficiary's total expenditures on premiums and out of pocket costs increased from $1,203 in 2007 to $1,434 in 2010.[13] This is a significant share of income given that 45% of

[11] A potential limitation of working with the MCBS sample is that it is not designed to be nationally representative without weighting, and selecting the appropriate weights is complicated by panel rotation and by our exclusive focus on respondents who participated in the PDP market. For example, by design the MCBS does not attempt to sample individuals from 3 out of the 34 PDP regions: 1(Maine and New Hampshire), 20 (Mississippi), and 31 (Idaho and Utah). To assess whether using unweighted MCBS data might compromise the external validity of our results, we compared the unweighted demographics of the average enrollee in our linked sample with a random 20% sample of all Part D enrollees from CMS's administrative files. Table A1 shows that the average enrollee in our linked sample is 1 to 2 years older. Otherwise, the two samples are virtually identical in terms of race, gender, rates of dementia and depression, number of PDP brands and plans available, expenditures on plan premiums and OOP costs, and the maximum amount of money that the average enrollee could have been saved by enrolling in a different plan. Given the strong similarity in demographics and PDP expenditures between the two samples, we expect that our findings from the linked MCBS-administrative sample can be generalized to the broader population of non-poor Part D enrollees.

[12] Approximately half of all survey respondents purchased a standalone PDP during the first five years of the program. Respondents who do not purchase a standalone PDP can instead obtain prescription drug insurance through an employer sponsored plan or a Medicare Advantage plan.

[13] The figure for 2006 is $1,020. It is smaller because during the inaugural year of the program open enrollment extended through May. More than half the enrollees in our sample were not enrolled for the full year. If we limit the sample to full-year enrollees to make it more comparable

beneficiaries have household incomes below $25,000. The data also reveal that by the end of our study period significant fractions of enrollees had been diagnosed with dementia (12%) and depression (11%). Given the relatively large amount of money at stake, the age range of the eligible population and the prevalence of cognitive illnesses it is unsurprising to find that 38% of enrollees did not make health insurance decisions on their own: 27% had help and 11% relied on a proxy to make the decision for them. Table 2 shows that beneficiaries who get help are likely to be older, sicker, lower income, less educated, and less internet savvy than beneficiaries who made decisions on their own. Those getting help are also more likely to have been diagnosed with depression or dementia. All of these differences are amplified when we compare beneficiaries who make their own health insurance decisions to those who rely on proxies to make decisions for them.

TABLE 2—CHARACTERISTICS OF PEOPLE WHO MAKE THEIR OWN DECISIONS OR GET HELP

	Who makes health insurance decisions?		
	Beneficiary	Beneficiary gets help	Proxy
number of enrollment decisions	9,739	4,157	1,733
high school graduate (%)	83	76	62
college graduate (%)	25	19	14
income>$25k (%)	56	52	47
uses the internet (%)	39	33	19
mean age	77	78	80
dementia including Alzheimer's (%)	5	10	30
depression (%)	9	11	15
mean number of drug claims	32	37	41
mean premium ($)	431	431	443
mean out-of-pocket costs ($)	908	1,063	1,303
mean potential savings ($)	334	351	380

Note: The table reports means for key variables for the sample of Medicare Part D enrollees found in both the MCBS and cost calculator samples from 2006-2010. See the text for details.

Similar to Chetty et al. (2015) we define an enrollment choice as *active* if one of the following statements is true: (1) the person is new to the market and must select a plan in order to obtain health insurance, (2) the old plan the person selected the prior year was discontinued, or (3) the old plan was still available but the person chose to switch to a new plan during open enrollment. If none of these statements is true, then the enrollee took no action during open enrollment and

to later years, then the mean annual consumer expenditure is $1,373.

CMS automatically reassigned her to the plan she chose last year—her default plan—in which case we code her choice as *passive*. Table 1 shows that after the inaugural enrollment cycle in 2006 between 20% and 23% of the people in our sample made active enrollment decisions each year. Approximately half of these are people who chose to switch out of their default plans and half are new enrollees or people whose old plans were discontinued.

III. Identifying Enrollment Decisions Suspected to be Based on Misinformation

Only 8% of all enrollment decisions made between 2006 and 2010 resulted in the enrollee minimizing ex post expenditures. The last row of Table 1 shows that in 2006 the average enrollee could have saved $460 in terms of lower premiums and OOP costs by choosing a different plan.[14] This is equivalent to reducing total PDP expenditures by 45%. Potential savings declined to $349 in 2007 (or 29% of expenditures) and remained similar thereafter. The last row of Table 2 shows that potential savings are *larger* for beneficiaries who get help or rely on proxies. Why are people leaving so much money on the table? We hypothesize that the answers differ from person to person. Some may be making informed decisions to purchase more expensive plans because those plans provide more risk protection and higher quality. Others may not fully understand how the market works or may be underestimating their potential savings. We must distinguish between these two groups in order to evaluate the welfare effects of prospective policies.

For the informed group we can apply standard revealed preference logic to infer their preferences for cost reduction, risk protection, and quality from their enrollment decisions (e.g. using a logit model of PDP choice). We cannot use the same approach to recover preferences for the misinformed group. Standard revealed preference logic does not apply to them because their beliefs about the market appear to contradict the objective information that we observe (Ackerberg et al. 2007, Train 2009). With this in mind, we adapt two features of Bernheim and Rangel's (2009) proposed approach to revealed preference analysis in the presence of partially latent heterogeneity in beliefs.[15] First, we use theory and data to identify enrollment decisions that we suspect may not reveal consumers' preferences for PDP attributes. We label these choices as *suspect*, using Bernheim and Rangel's terminology. Second, for the subset of consumers making suspect choices, we calibrate their preference relations using proxy measures derived from the

[14] This average falls below the $520 figure reported by Ketcham, Lucarelli and Powers (2015) based on CMS's 20% sample of 2006 full year enrollees because our average also includes people who only enrolled for part of the year. The primary reason for part-year enrollment in 2006 was the fact that the initial open enrollment period was extended through May (Heiss, McFadden, and Winter 2010).

[15] Partially latent heterogeneity in beliefs is an example of what Bernheim and Rangel refer to as "ancillary conditions" on decision making.

behavior of observationally equivalent consumers who we observe making *non-suspect* choices. Thus, we implement Bernheim and Rangel's proposal to respect consumer sovereignty and the standard logic of revealed preference analysis unless theory and data suggest the standard approach will fail to reveal consumers' preferences.

A. Potential Indicators of Suspect Choices

We follow the prior literature on Part D by assuming that consumer i's utility from drug plan j in year t depends on the mean and variance of her potential expenditures under that plan. Expenditures equal the plan premium, p_{jt}, plus out of pocket costs, $oop_{jt}(x_i t)$, of an exogenously given vector of drug quantities, x_{it}. Utility also depends on measures of plan quality, q_{ijt}, that reflect the time and effort required for an individual to obtain her eligible benefits under the plan.

Our first indicator of suspect choices is derived by applying Ketcham, Kuminoff, and Powers' (2014) nonparametric test for whether consumers making active enrollment decisions are choosing plans that cannot be rationalized as maximizing a well behaved utility function under full information. To simplify notation we denote total costs as $c_{ijt} = p_{jt} + oop_{ijt}$. We assume that consumers are risk averse and have preference orderings that are complete, transitive, and strongly monotonic over expected cost savings, risk reduction, and quality.[16] Under this assumption, a fully informed utility maximizing consumer will never choose a plan, j, that is dominated by another, k, in the sense that the following four conditions hold simultaneously.

$$E(c_{ikt}) \leq E(c_{ijt}). \tag{1.a}$$

$$var(c_{ikt}) \leq var(c_{ijt}). \tag{1.b}$$

$$q_{ijt} \leq q_{ikt}. \tag{1.c}$$

$$At\ least\ one\ of\ the\ inequalities\ is\ strict. \tag{1.d}$$

In words, an informed utility maximizing consumer will never choose a plan that has higher cost, higher variance, and lower quality than some feasible alternative. We refer to active plan choices that satisfy (1.a)-(1.d) as being *dominated*. In theory, a consumer may choose a dominated plan if she is risk loving, if she dislikes quality, if she has a negative marginal utility of income, or if she is misinformed about her options. We believe that misinformation is the most plausible of these

[16] Completeness says that consumers can compare any two plans. Transitivity says that if plan A is preferred to plan B, and plan B is preferred to plan C, then plan A must be preferred to plan C. Strong monotonicity says that, all else constant, consumers prefer plans with more of any positive attribute.

explanations. Hence, if we observe a consumer choosing a dominated plan then we label that choice as "suspect". We suspect that the consumer is misinformed and, therefore, that her enrollment decision will not reveal her preferences for health insurance.

On the other hand, consumers who violate at least one of the four conditions are necessarily choosing plans on what Lancaster (1966) called the "efficiency frontier" in attribute space. Every plan on the frontier can be rationalized as maximizing some utility function that satisfies completeness, transitivity, strong monotonicity, and risk aversion under the assumption of full information. For example, an informed risk averse consumer may maximize utility by choosing a more expensive and lower quality plan that better insures her against negative health shocks.

To test whether enrollees chose dominated plans we define PDP cost, variance, and quality using standard methods from the literature on modeling PDP choice (Abaluck and Gruber 2011, Ketcham, Kuminoff, and Powers 2015). First we assume that informed utility-maximizing consumers have unbiased expectations of their drug needs for the upcoming year: $E(c_{ijt}) = c_{ijt}$.[17] Then we use the cost calculator from Ketcham, Lucarelli, and Powers (2015) to define c_{ijt} for every available plan based on consumer *i*'s actual drug claims in year *t*.

Measuring $var(c_{ijt})$ is complicated by the fact that we only observe consumer *i* under one realization from her distribution of potential health states in year *t*. We address this challenge by calculating $var(c_{ijt})$ from the distribution of expenditures that would have been made under plan *j* in year *t* by the set of individuals who looked similar to consumer *i* in year *t-1* in terms of prescription drug claims. More precisely, we use CMS's random 20% sample of all PDP enrollees to assign each individual in the MCBS sample to 1 of 1000 cells defined by the deciles to which she belonged in the national distributions of the prior year's total drug spending, total days' supply of branded drugs, and total days' supply of generic drugs. Then we calculate $var(c_{ijt})$ for the distribution of drugs consumed by everyone in consumer *i*'s cell.

Consumers may also have heterogeneous preferences over PDP quality. Plans differ in their pharmacy networks, customer service, ease of obtaining drugs by mail order, and various aspects of formulary coverage not captured by mean and variance of ex post costs, such as the prevalence of prior authorization (PA) requirements. PA requirements for certain drugs may be unattractive to consumers who believe they have a high likelihood of purchasing those drugs and ir-

[17] Our econometric estimates and policy conclusions are robust to assuming that consumers are myopic: $E(c_{ijt}) = c_{ijt-1}$. This is unsurprising since individual prescription drug use is strongly persistent over time.

relevant to consumers who do not. Likewise, consumers differ in their proximity to in-network pharmacies. These factors vary across insurance brands and consumers but not across plans within a brand. Therefore we use brand indicators as a proxy measure of horizontally differentiated quality, in addition to a vertical index of average plan quality developed by CMS.[18] This means that in order for a chosen plan to be dominated the enrollee must have been able to choose another plan offered by the same insurer that would have lowered both the mean and variance of her drug expenditures.

The first row of Table 3 shows that between 15% and 23% of consumers chose dominated plans each year. All other consumers made PDP choices that can be rationalized as maximizing some utility function satisfying risk aversion, completeness, transitivity, and strong monotonicity under our assumptions about how to measure mean and variance of cost. The default assumption of consumer sovereignty would dictate that we label all of the choices as *non-suspect.* However, this approach has the potential for type II error, as enrollees with misinformation could have chosen undominated plans. We investigate this possibility by using the MCBS to develop an additional suspect choice indicator.

TABLE 3—POTENTIAL INDICATORS OF SUSPECT CHOICES

	Percent of enrollees					
	2006	2007	2008	2009	2010	2007-2010
choosing a dominated plan	19	23	18	15	16	18
wrong answer to MCBS knowledge question	44	29	31	28	28	29
union of the two indicators	55	45	43	38	39	41

Note: The table reports the share of choices triggering each indicator, by year. The MCBS knowledge question asks whether the enrollee's out of pocket costs are the same under every available drug plan. The correct answer is coded as yes for enrollees who filed drug claims in both the prior and current years if their out of pocket costs did in fact vary across plans in both years. The last row reports the share of enrollees satisfying the criteria in either of the first two rows. See the text for additional details.

Our second indicator comes from a question on the MCBS that is designed to test enrollees' knowledge of the Part D program. Respondents are asked to state whether the following sentence is true or false. *Your OOP costs are the same in all Medicare prescription drug plans*. For people with no drug claims, the statement is true. For virtually all people with drug claims the statement is false due to variation in formularies, deductibles, negotiated drug prices and other plan design attributes. This variation is economically important. The average beneficiary's OOP costs for her

[18] A 2006 MedPac survey asked beneficiaries about the factors affecting their initial choice of PDP. 90% of respondents stated that company reputation was "important" or "very important" to their choice and 84% gave the having a preferred pharmacy in the plan's network (84%)

chosen bundle of drugs vary by more than $1,100 over the plans available to her. Misunderstanding this crucial feature of the market could cause enrollees to spend far more than they would have otherwise.[19]

We use each person's actual drug claims to determine the correct answer to the MCBS question. Because respondents may be unsure about which enrollment year the question is referring to, we code a person's answer for year *t* as correct if the answer she gave in that year is correct for either year *t* or year *t-1*. Row 2 of Table 3 shows that a substantial share of respondents gave the wrong answer—44% in the first year of the program and between 28% and 31% in each subsequent year. The sharp reduction between 2006 and 2007 is consistent with prior evidence on learning in Medicare PDP markets (Ketcham, Lucarelli, and Powers 2015, Ketcham et al. 2012).

The dominated choice and knowledge question indicators are complementary. Twenty-five percent of our sample gave the wrong answer to the MCBS knowledge question but did not choose a dominated plan. The average person who gave the wrong answer could have saved 14% more by switching to a different plan than the average person who gave the right answer.[20] Our main analysis is based on defining suspect choices using the union of the two indicators. The last row of Table 3 shows that this group includes 41% of enrollment decisions made between 2007 and 2010. We explore the sensitivity of our estimates to a variety of alternative definitions for suspect choices and find that our policy conclusions are robust. This includes defining suspect choices based exclusively on dominated plans; using a more inclusive definition that adds every enrollee who could have reduced her expenditures by more than 50% or by more than 33%; and to assuming that consumers are myopic in the sense that they expect their drug use in the upcoming year to be identical to their drug use in the prior year. These and other robustness checks are described in section VII and the supplemental appendix.

B. Who is More Likely to Make Suspect Choices?

To develop intuition for the potential mechanisms driving the probability of making a suspect

[19] The MCBS asks five other questions that test knowledge of Part D, but they are less relevant for forecasting individual drug expenditures. They are (correct answer in parentheses): "All Medicare prescription drug plans cover the same list of prescription drugs" (false); "Everyone in Medicare has at least two Medicare prescription drug plans to choose from" (true); "Everyone with Medicare can choose to enroll in the voluntary Medicare prescription drug coverage regardless of their income or health. "(true); "Medicare prescription drug plans can change the price of prescription drugs only once per year."(false); "Generally, once you join a Medicare prescription drug plan, you can only change to another plan during the 'Open Enrollment period' each year." (true); "If you have limited income and resources, you may get extra help to cover prescription drugs for little or no cost to you."(true). Howell, Wolff and Herring (2012) provide further analysis of the MCBS knowledge questions.

[20] Table A2 shows that when we focus on individuals who made active enrollment decisions without help from others answering the knowledge question incorrectly is associated with a 2.1% increase in the probability of choosing a dominated plan after conditioning on demographics and year and region indicators (p=.172). It is also associated with a $24 increase in the amount of money the enrollee could have saved by switching to a different plan offered by the same insurer (p=.011).

choice, we estimate linear probability models in which the dependent variable, S_{it}, is an indicator for whether person i in CMS region r chose a dominated plan and/or answered the MCBS knowledge question incorrectly in the year t enrollment cycle,

$$S_{irt} = \kappa + \lambda d_{irt} + \phi_r + \rho_t + e_{it}. \tag{2}$$

On the right of the equality d_{irt} is a vector of individual characteristics defined from the variables in Table 1, some of which change over time, and ρ_t and ϕ_r are indicators for enrollment year and CMS region. These indicators capture variation in the complexity of choice architecture across space and time. For example, in the first year of the program the number of available plans per region ranged from 27 to 52. The number of plans also changed over time, increasing noticeably between 2006 and 2007. This variation allows us to test the choice overload hypothesis that consumers are less likely to make informed decisions as the number of options grows.

The first column of Table 4 reports results for our full sample of 15,629 enrollment decisions. The omitted demographic indicators define the reference enrollee as a 65 to 69 year old unmarried and retired white male with no high school diploma. The estimated coefficients imply that obtaining a high school degree is associated with a 3.7 percentage point reduction in the probability of making a suspect choice and this differential increases to 8.4 percentage points for enrollees with a college degree. The probability is higher for nonwhites (+5.0) which might proxy for unobserved differences in income, education, and English proficiency, and it is higher for people who get help making insurance decisions (+5.3) which might proxy for agency problems or reflect the fact that individuals who expect to get help may have less incentive to become informed. We also see lower probabilities for enrollees who report using the internet to search for information about Medicare programs (-6.0) and who reporting called 1-800-Medicare for information (-5.7). The later result is consistent with Kling et al.'s (2012) secret shopper audit of the Medicare help line in which actors calling the number for information found that customer service representatives consistently identified low-cost plans based on the actors' fictional drug needs.

Looking at the administrative variables, we see the probability of making a suspect choice increasing in age, consistent with prior evidence on the decline in cognitive performance for individuals over 65 (Agarwal et al. 2009, Tymula et al. 2013). The predicted probability is approximately 5 percentage points higher for enrollees in their late 70's and approximately 12

percentage points higher for enrollees in their late 80's. This is after controlling separately the effects of diagnosed cognitive illnesses normally associated with aging, namely dementia (+5.5) and depression (+4.2), and conditioning on the increased complexity of decisionmaking associated with greater drug needs via a measure of total drug claims (+5.4 for a one standard deviation increase in claims).

TABLE 4—ASSOCIATION BETWEEN SUSPECT CHOICES AND DEMOGRAPHICS

	Dominated plan or wrong answer to knowledge question		Suspect choice		Suspect choice	
	2006 - 2010		2006 - 2010		2007 - 2010	
high school graduate	-0.037	[0.015]**	-0.004	[0.022]	-0.003	[0.022]
college graduate	-0.047	[0.014]***	-0.062	[0.022]***	-0.058	[0.022]***
income>$25k	-0.016	[0.012]	-0.017	[0.018]	-0.020	[0.019]
currently working	0.027	[0.017]	0.005	[0.025]	0.005	[0.025]
married	-0.019	[0.013]	-0.015	[0.019]	-0.018	[0.020]
has living children	-0.010	[0.021]	0.001	[0.032]	-0.001	[0.032]
uses the internet	-0.016	[0.014]	-0.010	[0.021]	-0.006	[0.021]
searched for CMS info: internet	-0.060	[0.014]***	-0.073	[0.021]***	-0.064	[0.021]***
searched for CMS info: 1-800-Medicare	-0.057	[0.012]***	-0.044	[0.019]**	-0.051	[0.020]**
has help making insurance decisions	0.053	[0.012]***	0.028	[0.016]*	0.024	[0.017]
number of available plans (standardized)	0.004	[0.015]	0.001	[0.013]	0.011	[0.015]
female	0.003	[0.012]	0.022	[0.019]	0.021	[0.019]
nonwhite	0.050	[0.023]**	0.135	[0.034]***	0.129	[0.035]***
age: 70-74	0.017	[0.014]	0.047	[0.020]**	0.055	[0.022]**
age: 75-79	0.052	[0.016]***	0.062	[0.025]**	0.068	[0.026]***
age: 80-84	0.061	[0.017]***	0.092	[0.026]***	0.103	[0.027]***
age: over 84	0.117	[0.019]***	0.136	[0.028]***	0.148	[0.029]***
dementia including Alzheimer's	0.055	[0.018]***	0.047	[0.025]*	0.043	[0.025]*
depression	0.042	[0.016]***	0.037	[0.021]*	0.029	[0.022]
number of drug claims (standardized)	0.054	[0.006]***	0.034	[0.008]***	0.042	[0.008]***
number of plan choices	15,629		11,739		9,831	
number of enrollees	5,269		3,607		3,511	
mean of the dependent variable	0.43		0.46		0.44	
R-squared	0.0768		0.075		0.074	

Note: The table reports coefficients and standard errors from linear probability models of individual enrollment decisions. In the first column the dependent variable equals one if the decision maker chose a dominated plan or answered the MCBS knowledge question incorrectly. The next two columns drop observations where the enrollee elected to remain in their default plan and we do not observe their response to the MCBS survey in the year in which they originally chose that plan, as described further in the text. All explanatory variables are binary expect for the number of available plans and the number of drug claims, both of which are standardized. The omitted indicator variables define the baseline enrollee as a 65 to 69 year old white male who did not finish high school, has income below $25k, does not get help making insurance decisions, has not searched for CMS information using the internet or 1-800-Medicare, has the mean number of drug claims, and has not been diagnosed with dementia or depression. All regressions include indicators for enrollment year and region. Robust standard errors are clustered by enrollee. *,**, and *** indicate that the p-value is less than 0.1, 0.05, and 0.01 respectively.

In comparison we find that income, gender, marital status, and the existence of children have small and statistically insignificant effects. We also obtain a precisely estimated zero on the number of available plans. Because the OLS model includes fixed effects for years and regions, the coefficient on the number of plans is identified by the within-region changes over time in the number of available plans. The estimated coefficient provides evidence against the hypothesis that choice overload causes suspect choices.

The middle column of Table 4 shows results from repeating the estimation after dropping 3,890 passive enrollment decisions for which we do not observe the enrollee's knowledge at the time they made their preceding active choice (i.e. their preceding active choice occurred before they joined the MCBS). Dropping these observations allows us to strengthen the link between active decision making and the way that we define choices as being *suspect* or *non-suspect*. Specifically, we define a choice as suspect if and only if at least one of the following statements is true (i) the decision maker actively enrolled in a dominated plan; (ii) the decision maker passively reenrolled in a plan that was dominated when she actively chose it during a prior enrollment cycle; (iii) the decision maker answered the MCBS knowledge question incorrectly; or (iv) the decision maker passively reenrolled in a plan that she actively chose during a year in which she answered the knowledge question incorrectly. Hence, during years in which decision makers answer the MCBS knowledge question correctly and choose to passively remain in their default plans we defer to their most recent active choice when coding their subsequent passive decisions as suspect or non-suspect. It is also important to reiterate here that the decision maker refers to the person who answered the MCBS survey and made the enrollment decision. In about 10% of our sample this person is a proxy for the Medicare beneficiary such as a spouse or child (Table 1).

The last column of Table 4 shows results for the linear probability model of suspect choices after dropping enrollment decisions in 2006. We exclude 2006 from our main analysis because of the sharp improvement in responses to the MCBS knowledge question in 2007. Because the consumer population appears to have been less informed during the inaugural year of the program, their choices during that year may be less informative for analyzing prospective policies. That said, we show in the supplemental appendix that including 2006 does not change our main findings (Table A8).

The general pattern of results from the linear probability model in column 1 is robust to the

data cuts that we use to refine our sample and sharpen our definition of suspect choices. In columns 2 and 3 the probability of making a suspect choice declines as education increases and as people exert effort to learn about the market using the internet or 1-800-Medicare. The probability increases as people age, as they are diagnosed with cognitive illnesses, and as their drug spending increases. These results are consistent with the hypothesis that information is costly to acquire (Stigler and Becker 1977) and that decision making costs vary systematically with age and human capital (Agarwal et al. 2009, Tymula et al. 2013). Increases in the magnitudes of the coefficients on age and nonwhite in the last two columns are driven by the way that we defer to prior active decisions to code some subsequent passive reenrollments as suspect or non-suspect. This is a signal that inertia may be greater for older and nonwhite beneficiaries, and we test this hypothesis within our parametric model of decision making.

IV. A Parametric Model of Decision Making with Heterogeneity in Beliefs

To evaluate the welfare effects of prospective polices we must first select a parametric approximation to utility. The main novelty of our approach to is to allow for heterogeneity in beliefs about plan attributes. We focus on identifying parameters that describe how plan attributes affect PDP choice and then use our indicators for *suspect* and *non-suspect* choices to guide how we interpret those parameters.

A. Initial Enrollment Decision

When a beneficiary first enters the market in year 0 she must actively choose a plan to obtain insurance. She will choose the plan that maximizes her utility, conditional on her beliefs about plan attributes. We approximate this process with a simple linear model similar to the ones used by Abaluck and Gruber (2011), Kling et al. (2012) and Ketcham, Kuminoff, and Powers (2015),

$$U_{ij0} = \alpha_{it}\acute{c}_{ij0} + \beta_{it}\acute{\sigma}^2_{ij0} + \gamma_{it}\acute{q}_{ij0} + \epsilon_{ij0}. \tag{3}$$

$\acute{c}_{ij0}$ denotes the amount that person *i* expects to spend under plan *j* in terms of the premium plus out of pocket costs for prescription drugs, $\acute{\sigma}^2_{ij0}$ is the variance of out of pocket costs, $\acute{q}_{ij0}$ is a vector of quality attributes, and ϵ_{ij0} is an idiosyncratic person-plan specific taste shock. The accents indicate that the variables reflect person *i*'s subjective beliefs about plan attributes at the time of her enrollment decision. Heterogeneity in beliefs is discussed below. Beneficiaries may

also have heterogeneous marginal rates of substitution between expected cost, variance, and quality. We model this heterogeneity as a linear function of observable demographic characteristics, some of which may evolve over time: $\alpha_{it} = \alpha_0 + \alpha_1 d_{it}$, and similarly for β_{it} and γ_{it}. Finally, there may be a utility cost of time and effort required to learn about a plan and enroll in it. We assume that this cost is constant across plans so that it cancels out of between-plan comparisons and can therefore be suppressed in (3).

B. Subsequent Enrollment Decisions

After an enrollee chooses a plan in year 0 she is automatically reassigned to that plan in year 1 unless she actively chooses to switch to a different plan during open enrollment.[21] As before, we assume that it is costly to make an active decision. Because no effort is required for the consumer to reenroll in her default plan the passive decision to do so has a relatively higher payoff:

$$U_{ij1} = \alpha_{it}\acute{c}_{ij1} + \beta_{it}\acute{\sigma}^2_{ij1} + \gamma_{it}\acute{q}_{ij1} + \eta_{it}\Delta\acute{B}_{ij1} + \delta_{it}\Delta\acute{P}_{ij1} + \epsilon_{ij1}. \tag{4}$$

We use two terms to capture the utility cost of actively switching plans: $\Delta\acute{P}_{ijt}$ is an indicator for whether plan *j* is a non-default plan sold by the same insurer as the default plan, and $\Delta\acute{B}_{ijt}$ is an indicator for whether plan *j* is a non-default plan sold by a different insurer. The decision process follows the same structure as (4) in subsequent enrollment years *2*,…,*T*. The disutility of switching plans is captured by the switching parameters, $\eta_{it} = \eta_0 + \eta_1 d_{it}$ and $\delta_{it} = \delta_0 + \delta_1 d_{it}$, which summarize how inertia varies with consumer demographics. We address the issue of how to interpret inertia when we discuss welfare measurement in Section V, and we explore the policy implications of these interpretations in Section VII.

C. Heterogeneity in Information

We model heterogeneity in information by allowing suspect and non-suspect choices to be driven by different beliefs about PDPs. Non-suspect choices are assumed to be informed in the sense that decision makers' beliefs about plan attributes coincide with the objective measures we have collected. Put differently, we respect consumer sovereignty and invoke the standard assumption of full information in the absence of evidence to the contrary. In contrast, we do not

[21] Plans are occasionally discontinued, which can force people to make an active choice. In such case, we can revert to equation (3) to model the new enrollment decision.

observe the beliefs about plan attributes that led to suspect choices. While the non-suspect (*n*) and suspect (*s*) groups may have different beliefs about plan attributes, we assume that they maximize utility functions characterized by the same underlying preference parameters, after conditioning on demographics and prescription drug utilization.

$$U_{ijt}^{n} = \alpha_{it}c_{ijt} + \beta_{it}\sigma_{ijt}^{2} + \gamma_{it}q_{jt} + \eta_{it}\Delta B_{ijt} + \delta_{it}\Delta P_{ijt} + \epsilon_{ijt}. \tag{5}$$

$$U_{ijt}^{s} = \alpha_{it}\acute{c}_{ijt} + \beta_{it}\acute{\sigma}_{ijt}^{2} + \gamma_{it}\acute{q}_{ijt} + \eta_{it}\Delta\acute{B}_{ijt} + \delta_{it}\Delta\acute{P}_{ijt} + \epsilon_{ijt}. \tag{6}$$

We dropped the accents in (5) to indicate that we are using objective measures of plan attributes for the non-suspect group. Their expected PDP costs are defined as $c_{ijt} = p_{jt} + E\left[oop_{ijt}\right]$, their type-specific variance is defined as $\sigma_{ijt}^{2} = var\left(oop_{ijt}\right)$, and q_{jt} is a vector containing the CMS quality index and brand indicators. All variables are calculated using the techniques developed in prior studies of PDP choice as described in III.A.

Because we do not observe the beliefs of people making suspect choices, we cannot expect to identify their preferences from their observed behavior. To see this notice that if we replace the subjective beliefs in (6) with objective measures of plan attributes then, in general, we must also allow the values of the preference parameters and the error term to change in order to maintain the same utility ranking of plans:

$$U_{ijt}^{s} = \acute{\alpha}_{it}c_{ijt} + \acute{\beta}_{it}\sigma_{ijt}^{2} + \acute{\gamma}_{it}q_{ijt} + \acute{\eta}_{it}\Delta B_{ijt} + \acute{\delta}_{it}\Delta P_{ijt} + \acute{\epsilon}_{ijt}. \tag{7}$$

For example, if people make suspect choices because they have downward biased expectations about their OOP costs at the time they choose a plan (i.e. $c_{ijt} > \acute{c}_{ijt}$) then we would expect $\alpha_{it} < \acute{\alpha}_{it}$. If they answer the MCBS knowledge question incorrectly because they mistakenly believe that their PDP choice will have no effect on their out of pocket costs, then we would expect $\beta_{it} < \acute{\beta}_{it}$. Likewise, if they have downward biased beliefs about their potential savings from switching plans, then we would expect $\eta_{it} < \acute{\eta}_{it}$ and $\delta_{it} < \acute{\delta}_{it}$.

To facilitate estimation we make the standard assumption that the person-plan specific taste shocks in (5) and (7) are *iid* draws from type I extreme value distributions. However, notice that the variance may differ between the suspect and non-suspect groups. This is because the idiosyncratic shocks in (7) will absorb any residual utility differences needed to maintain the preference

ordering over plans when we move from (6) to (7). Therefore, when we follow the standard approach to normalizing the model variance to equal $\pi^2/6$, the coefficients estimated for the suspect group will be scaled by the ratio of the group-specific variances (Train 2009). After making this normalization, we can rewrite the estimating equation for the suspect group (*s*) as

$$U_{ijt}^{s} = \alpha_{it}^{s} c_{ijt} + \beta_{it}^{s} \sigma_{ijt}^{2} + \gamma_{it}^{s} q_{ijt} + \eta_{it}^{s} \Delta B_{ijt} + \delta_{it}^{s} \Delta P_{ijt} + \epsilon_{ijt}, \tag{8}$$

where $\alpha_{it}^{s} = \acute{\alpha}_{it}\sqrt{var(\epsilon_{ijt})/var(\acute{\epsilon}_{ijt})}$ and similarly for β_{it}^{s}, γ_{it}^{s}, η_{it}^{s}, and δ_{it}^{s}. Our econometric model identifies the parameters of (5) and (8).

D. Identification

Equations (3)-(4) illustrate how the model parameters can be identified from data on suspect and non-suspect enrollment decisions made by each group. Consider the non-suspect group. Given the assumed parametric form for utility and the distributional assumption on ϵ_{ijt}, we can use a multinomial logit model of initial enrollment decisions (3) to identify the parameters that describe how marginal rates of substitution between cost, variance, and quality vary with beneficiary demographics, $\alpha_0, \alpha_1, \beta_0, \beta_1, \gamma_0, \gamma_1$. Then we can use a model of their subsequent enrollment decisions (4) to identify the inertia parameters, $\eta_0, \eta_1, \delta_0, \delta_1$, via the rates at which individuals actively switched out of the plans they initially chose. In practice, we pool the data from all enrollment decisions and estimate the parameters simultaneously using the specification in (5). The same arguments can be made to identify the parameters of (8) for the suspect group. Prior studies have analyzed the properties of this model and underlying identification arguments in detail (c.f. Abaluck and Gruber 2011 and Ketcham, Kuminoff, and Powers 2015). The novelty of our approach is estimating separate parameters for suspect and non-suspect groups and investigating how their marginal rates of substitution between cost, risk reduction, and quality vary with demographics. The ability to differentiate their decision processes is critical to assessing who would win and who would lose from prospective policies affecting choice architecture.

V. Welfare Effects of Modifying Choice Architecture

When some decision makers are misinformed, reforms that modify choice architecture by reducing information costs and/or simplifying the choice process can, in principle, increase some

consumers' welfare. Consider a policy implemented between periods 0 and 1 that changes the set of available plans from J to K. Consumer welfare may be directly affected through three channels. First, the policy may change the menu of options by adding choices, removing choices, and regulating their costs or quality. Second, the policy may change how consumers or firms make decisions, e.g. by lowering the cost of information in a way that reduces the disutility of switching plans. Finally, if the policy induces consumers and firms to adjust their behavior then those adjustments may feed back into the levels of endogenous attributes (e.g. premiums) through the equilibrium sorting process.

A. Non-Suspect Group

The expected change in welfare for people in the non-suspect group (n) can be derived by integrating over ϵ_{ijt} in the standard expression for consumer surplus to generate the log sum ratio from Small and Rosen (1981).

$$\Delta E[CV_i^n] = \frac{1}{\alpha_{it}^n}\left\{ln\frac{\sum_{k\in K}\left[exp\left(V_{ik}^{n1}\right)\right]}{\sum_{j\in J}\left[exp\left(V_{ij}^{n0}\right)\right]}\right\}, \tag{9}$$

where V_{ij}^{n0} and V_{ik}^{n1} denote the observed part of the utility function in (5) evaluated for PDPs j and k before and after the policy. The temporal subscript is suppressed for brevity such that $V_{ij}^{n0} = V_{ijt}^{n0}(\theta^n, d_{it}) = U_{ijt}^{n0} - \epsilon_{ijt}$, where $\theta^n = [\alpha^n, \beta^n, \gamma^n, \eta^n, \delta^n]$ and each letter is a vector of parameters describing how preferences vary with individual demographics within the non-suspect group such that $\alpha^n = [\alpha_0^n, \alpha_1^n]$ for example.

B. Suspect Group

Welfare calculation is more involved for those making suspect choices. The observed part of (8) determines how PDP attributes affect their enrollment decisions, but their ex post realized utility from those decisions is determined by (5). This follows from our assumption that the suspect and non-suspect groups share the same underlying preference parameters. Therefore, a single plan's contribution to expected utility is defined by integrating over the product of (5) and the probability of choosing that plan based on (8). Aggregating over the PDP menu prior to the policy yields the following general expression

$$(10) \qquad E[U_i^{s0}] = \sum_{j \in J} \int_{-\infty}^{\infty} \left(V_{ij}^{n0} + \epsilon_{ij}\right) F_j\left(V_{ij}^{s0} - V_{i1}^{s0} + \epsilon_{ij}, \dots, V_{ij}^{s0} - V_{iK}^{s0} + \epsilon_{ij}\right) d\epsilon_{ij},$$

where $F_j(\cdot)$ is the derivative of the joint CDF of the idiosyncratic taste shocks with respect to ϵ_{ij}. Subtracting this expression from the post-policy measure of expected utility, dividing by the marginal utility of income, and integrating over the idiosyncratic taste shocks yields an expression for welfare that was first derived by Leggett (2002) as a way to describe decision making under misinformation.

$$(11) \qquad \Delta E[CV_i^s] = \frac{1}{\alpha_{it}^n} \left\{ ln \frac{\sum_{k \in K}[exp(V_{ik}^{s1})]}{\sum_{j \in J}\left[exp\left(V_{ij}^{s0}\right)\right]} + \sum_{k \in K}\left[\psi_{ik}^{s1}\left(V_{ik}^{n1} - V_{ik}^{s1}\right)\right] - \sum_{j \in J}\left[\psi_{ij}^{s0}\left(V_{ij}^{n0} - V_{ij}^{s0}\right)\right] \right\},$$

where $V_{ij}^{s0} = V_{ijt}^{s0}(\theta^s) = U_{ijt}^{s0} - \epsilon_{ijt}$, $\theta^s = [\alpha^s, \beta^s, \gamma^s, \eta^s, \delta^s]$, and ψ_{ij} is the logit probability of choosing plan j so that $\psi_{ij}^{s0} = exp\left(V_{ij}^{s0}\right) / \sum_{m \in J}[exp(V_{im}^{s0})]$.

The first term inside braces in (11) is the standard log sum ratio evaluated at θ^s. It provides a biased measure of welfare when $\theta^s \neq \theta$ because suspect choices are based on misinformation. The second and third terms adjust the log sum ratio to account for the welfare implications of the difference between θ^s and θ for each choice, weighted by the predicted probability of making that choice before and after the policy. In the special case where $\theta^s = \theta$, equation (11) reduces to the standard welfare measure in (9).

C. The Welfare Treatment of Inertia

Equations (9) and (11) treat the inertia parameters estimated for the non-suspect group as being directly relevant to consumer welfare. This is consistent with interpreting inertia as some combination of latent preferences and hassle costs of switching plans. However, Kling et al. (2002) argue that inertia is more likely to reflect consumers' biased expectations about their potential savings from switching plans along with other psychological mechanisms such as status quo bias, procrastination, and limited attention. These mechanisms would have no direct effect on consumers' welfare; they only affect welfare indirectly by lowering the rate at which consumers switch plans. Unfortunately, the data do not allow us to determine the importance of psychological bias relative to latent preferences and switching costs. We address this by calculating bounds on welfare for two extreme cases, similar to Handel (2013). In the first case, inertia is as-

sumed to be entirely welfare relevant (as in equations (9) and (11)) and in the second case it is assumed to be entirely irrelevant.

To calculate the change in expected welfare when inertia is irrelevant we replace equations (9) and (11) with (9') and (11').

$$(9') \qquad \Delta E[CV_i^n] = \frac{1}{\alpha_{it}^n}\left\{ln\frac{\sum_{k\in K}[exp(V_{ik}^{n1})]}{\sum_{j\in J}[exp(V_{ij}^{n0})]} + \sum_{k\in K}[\psi_{ik}^{n1}(V_{ik}^{n^*1} - V_{ik}^{n1})] - \sum_{j\in J}[\psi_{ij}^{n0}(V_{ij}^{n^*0} - V_{ij}^{n0})]\right\}.$$

$$(11') \qquad \Delta E[CV_i^s] = \frac{1}{\alpha_{it}^n}\left\{ln\frac{\sum_{k\in K}[exp(V_{ik}^{s1})]}{\sum_{j\in J}[exp(V_{ij}^{s0})]} + \sum_{k\in K}[\psi_{ik}^{s1}(V_{ik}^{n^*1} - V_{ik}^{s1})] - \sum_{j\in J}[\psi_{ij}^{s0}(V_{ij}^{n^*0} - V_{ij}^{s0})]\right\}.$$

These equations differ from (9) and (11) in that $V_{ik}^{n^*1} = V_{ik}^{n1} - \eta_{it}^n \Delta B_{ijt} - \delta_{it}^n \Delta P_{ijt}$ and $V_{ij}^{n^*0} = V_{ik}^{n0} - \eta_{it}^n \Delta B_{ijt} - \delta_{it}^n \Delta P_{ijt}$. That is, we set the welfare-relevant share of the inertia parameters to zero so that inertia affects decision making by consumers in the suspect and non-suspect groups, but it has no direct effect on the welfare of consumers in either group.

D. The Policy's Effect on Consumer Behavior

We must also take a stance on whether a counterfactual policy that modifies choice architecture would induce consumers to behave differently than they do in our data and whether these changes equally affect consumers making suspect and non-suspect choices. In principle, a policy designed to simplify the choice process could induce decision makers in the suspect group to update their beliefs about the market and behave more like decision makers in the non-suspect group. Or it could have no effect at all. In the absence of empirical evidence, we again take a bounding approach and consider two extreme scenarios. One scenario assumes that the policy has no effect on behavior; the other assumes that the policy induces consumers in the suspect group to behave identically to those in the non-suspect group, conditional on observed demographics and prescription drug utilization. The second case simply requires that we replace V_{ik}^{s1} with V_{ik}^{n1} and ψ_{ik}^{s1} with ψ_{ik}^{n1} in equations (11) and (11').

E. Discussion

The conceptual logic underlying our welfare framework is well established. When it is costly to acquire information, to make a decision, or to negotiate a transaction some consumers may optimally choose not to become fully informed (Stigler and Becker 1977). Some consumers may

also be influenced by psychological biases (Kahnemann, Wakker, and Sarin 1997). Several prior studies have proposed theoretical and empirical approaches to inferring welfare when some consumers' choices cannot be interpreted as revealing their preferences due to incomplete information or biases (e.g. Leggett 2002, Bernheim and Rangel 2009, Fleurbaey and Schokkaert 2013). Our framework exemplifies Bernheim and Rangel's proposal to use preference relations estimated from non-suspect choices as proxies for people making suspect choices.[22] Our approach to defining suspect choices is grounded by the axioms of consumer theory and by evidence on whether decision makers are informed about a key feature of the market. Further, our approach to bounding welfare avoids the need to take a stance on whether inertia is driven by search costs or psychological biases.

To the best of our knowledge, our study is the first to use such an approach to evaluate the welfare effects of proposals to simplify choice architecture in a high-stakes differentiated product market that is both subsidized and regulated by the federal government.[23] Notably, our framework recognizes that modifying choice architecture may simultaneously create winners and losers. Consider the partial equilibrium welfare effects of a policy that eliminates consumer choice by assigning each consumer to a plan. Nobody can be made better off from such a policy under the conventional assumption that every decision maker is fully informed (e.g. Lucarelli, Prince, and Simon 2012). At the opposite extreme, nobody can be made worse off when the policy is imposed by government agents assumed to be omniscient, benevolent, and incorruptible (e.g. Abaluck and Gruber 2011). Our approach allows for a middle ground between these extremes. Equation (9) and its analogs recognize that informed consumers can be made worse off from restrictions on their ability to choose for themselves. Equation (11) and its analogs recognize that consumers who are misinformed may benefit from restrictions on choice if they would otherwise choose an inferior plan. In addition to examining distributional implications, analysts can aggregate the gains and losses over informed and uninformed consumers to yield a criterion for policy evaluation consistent with the concept of asymmetric paternalism (Camerer et al. 2003).

Our framework also highlights the various pieces of information needed to evaluate a pro-

[22] Alternatively, to use the language popularized by Kahnemann, Wakker, and Sarin (1997), one can think of $V_{ij}^{n}(\theta)$ as an approximation to the "hedonic utility" derived by consuming a good and $V_{ij}^{s}(\theta^{s})$ as an approximation to the "decision utility" function that is maximized by people who are less than fully informed about their choices.

[23] Perhaps the closest comparison is to Allcott and Taubinsky's (2015) recent field experiment in which shoppers in a single hardware store were randomized to different information treatments regarding the energy efficiency of certain types of light bulbs. They used the results to implement a version of Bernheim and Rangel's (2009) logic to evaluate welfare effects of EPA's restrictions on energy inefficient bulbs.

spective policy. First we must estimate the parameters describing how suspect and non-suspect choice probabilities vary with plan attributes, θ^n and θ^s, in order to calibrate ψ_{ij}^{s0}, V_{ij}^{n0}, V_{ij}^{s0}, and $V_{ik}^{n^*0}$. Then we must map each prospective policy into the parameters and endogenous plan attributes in order to calibrate ψ_{ij}^{s1}, V_{ij}^{n1}, V_{ij}^{s1}, and $V_{ik}^{n^*1}$.

VI. Multinomial Logit Estimation

We started with fully flexible models that interacted cost, variance, quality, and inertia with a rich set of demographic variables from Table 1. However, we found that most of the interactions had small and statistically insignificant effects on marginal rates of substitution (Table A3).[24] This led us to adopt the more parsimonious specification shown in Table 5 as the basis for our policy experiments.

The first column of Table 5 reports results for a naïve model that pools data on suspect and non-suspect choices. The main effects have the expected signs and are statistically different from zero, with the exception of variance. Its insignificant coefficient mirrors the finding from Abaluck and Gruber (2011) and Ketcham, Kuminoff and Powers (2015) that naive models of PDP choice imply that the typical enrollee does not consider risk protection when making health insurance decisions. Interacting variance with the MCBS indicator for college graduates suggests that the 22% of enrollees with college degrees tend to be more risk averse.

Columns 2 and 3 repeat the estimation for non-suspect and suspect choices alone. Comparing main effects across the three columns reveals that the insignificant coefficient on variance in the pooled model is simply driven by aggregating over suspect and non-suspect choices. Taken literally, the coefficient on variance for the suspect group would imply that they are risk loving. In contrast, the non-suspect group is risk averse and their implied degree of risk aversion is consistent with findings from prior studies (e.g. Cohen and Einav 2007, Handel 2013, Handel and Kolstad 2015). For example, our results imply that enrollees in the non-suspect group would be indifferent between a 50-50 bet of wining \$100 and losing between \$94.2 and \$96.3; and indifferent between a 50-50 bet of winning \$1,000 and losing between \$665.4 and \$738.9.[25] Further, the monetary value of inertia—defined by dividing the switching indicators by the expected cost

[24] A notable negative result from the full model is that the subsets of enrollees who do and do not get help making health insurance decisions make choices that imply virtually identical marginal rates of substitution between PDD cost, variance, and quality. The main difference between the two groups is that those who get help exhibit less inertia as seen in Table 5.

[25] These calculations are based on the fact that our specification for utility provides a 1st order approximation to a CARA model. Our calculations are additional discussion are provided in Table A4 and associated discussion in the supplemental appendix.

coefficient—is nearly three times larger for the suspect group.

TABLE 5—LOGIT MODELS OF PRESCRIPTION DRUG PLAN CHOICE

	All Choices		Non-Suspect choices		Suspect choices	
expected cost	-0.256	[0.016]***	-0.396	[0.028]***	-0.163	[0.018]***
variance	-0.210	[0.151]	-1.531	[0.297]***	1.118	[0.251]***
quality (CMS index)	0.244	[0.072]***	0.138	[0.103]	0.284	[0.109]***
within-brand switch	-3.341	[0.105]***	-3.215	[0.149]***	-3.470	[0.152]***
between-brand switch	-5.317	[0.087]***	-5.188	[0.123]***	-5.588	[0.124]***
cost x 1{ bottom tercile of claims }	-0.134	[0.028]***	-0.166	[0.041]***	-0.062	[0.036]*
cost x 1{ top tercile of claims }	0.085	[0.019]***	0.132	[0.035]***	0.041	[0.021]**
cost x 1{ sought CMS info }	-0.056	[0.019]***	-0.066	[0.033]**	0.004	[0.022]
variance x 1{ college graduate }	-0.565	[0.304]*	-0.903	[0.632]	0.309	[0.466]
quality x 1{ income > $25k }	0.233	[0.084]***	0.289	[0.113]**	0.082	[0.139]
quality x 1{ sought CMS info }	0.239	[0.089]***	0.173	[0.117]	0.343	[0.151]**
switch within brand x standardized age	-0.179	[0.069]***	-0.203	[0.096]**	-0.101	[0.102]
switch within brand x 1{ income > $25k }	-0.336	[0.125]***	-0.306	[0.173]*	-0.406	[0.186]**
switch within brand x 1{ help }	0.272	[0.122]**	0.096	[0.180]	0.454	[0.182]**
switch within brand x 1{ sought CMS info }	0.083	[0.131]	0.252	[0.168]	-0.232	[0.220]
switch within brand x 1{ nonwhite }	-0.711	[0.283]**	-0.474	[0.341]	-0.849	[0.446]*
switch brand x standardized age	-0.096	[0.049]*	-0.096	[0.070]	-0.043	[0.070]
switch brand x 1{ income > $25k }	-0.289	[0.097]***	-0.335	[0.138]**	-0.320	[0.142]**
switch brand x 1{ help }	0.251	[0.096]***	0.243	[0.135]*	0.287	[0.140]**
switch brand x 1{ sought CMS info }	0.306	[0.094]***	0.277	[0.129]**	0.290	[0.146]**
switch brand x 1{ nonwhite }	-0.650	[0.211]***	-0.675	[0.335]**	-0.449	[0.286]
pseudo R^2	0.67		0.66		0.71	
number of enrollment decisions	9,831		5,465		4,366	
number of enrollees	3,511		2,166		1,675	

Note: The table reports parameter estimates from logit models estimated from data on all choices; from non-suspect choices only; and from suspect choices only. All models include indicators for insurers. Robust standard errors are clustered by enrollee. *,**, and *** indicate that the p-value is less than 0.1, 0.05, and 0.01 respectively.

Focusing on non-suspect choices in column 2, we see that the interaction terms are mostly consistent with intuition. The reference individual is white and 78 years old with no college degree and annual income below $25,000, is in the middle tercile of the distribution of total drug claims, does not get help making enrollment decisions, and did not use the internet or 1-800-Medicare to search for information about CMS programs. Interactions between cost and indicators for whether the beneficiary is in the top or bottom terciles of the claims distribution imply

that the marginal utility of income declines as people become sicker. People who have previously taken the time to search for information about Medicare programs on the internet or by calling 1-800-Medicare tend to be more sensitive to price and to have stronger preferences for CMS's index of overall plan quality which is based, in part, on customer satisfaction. Preferences for plan quality are also higher among higher income enrollees. One explanation is that the opportunity cost of time is increasing in income and that choosing a higher quality plan reduces the time and effort required to interact with the insurer.

Inertia tends to be lower for people who get help choosing a plan and who previously searched for information about CMS programs, whereas it tends to be higher for people who are older, nonwhite, and who have higher incomes. The effect of income could again be due to heterogeneity in the opportunity cost of time. The directions of these relationships are mostly consistent across the suspect and non-suspect groups, but the monetary implications are much larger for the suspect group. Our results imply that the average non-suspect enrollee would have to be paid $809 to hold their utility constant if they were randomly reassigned to a different plan offered by the same insurer or $1,290 if they were reassigned to a plan offered by a different insurer. Comparable figures for the suspect group are $2,398 and $3,660. The fact that we see greater inertia for between-brand switches compared to within-brand switches is consistent with the hypothesis that the inertia parameters pick up latent preferences and differences in hassle costs. All else constant, between-brand switches are likely to require more time and effort than within-brand switches as different plans within a same brand have the same formularies, pharmacy networks, customer service, prior authorization requirements and so on. In contrast, different brands almost always differ along all of these dimensions, so that switching brands may require new prior authorization requests, transferring prescriptions to new pharmacies, and becoming familiar with the new brand's formulary and customer service systems. That said, it could also be possible that psychological biases are greater for between-brand switches.

A potential concern with our approach to distinguishing between suspect and non-suspect choices is that it could be "overfitting" the data and consequently yielding less accurate predictions for how consumers will respond to prospective policies. We assess our model's predictive power by using validation tests similar to the ones designed by Keane and Wolpin (2007) and Galiani, Murphy, and Pantano (2015). The idea is to compare the out of sample predictive power of our preferred model (i.e. distinguishing between suspect and non-suspect choices) with the na-

ïve pooled model. Our test is powered by the largest year-to-year change in the PDP choice set that occurred during our study period. Between 2008 and 2009 the number of plans available fell by 10%. We first use data from 2008 to estimate the naïve and preferred models and then use each set of estimates to make out of sample predictions for how consumers would adapt to their new choice sets in 2009.[26] Table A5 shows that our preferred model yields more accurate predictions for the share of consumers choosing dominated plans; the share of consumers choosing the least expensive plans offered by their preferred brands; mean expenditures; the average amount that consumers choosing dominated plans could save by switching; and the share of consumers choosing to switch plans. The pooled model does a better job of predicting one data moment by one percentage point—the share of consumers choosing plans with gap coverage. Overall, this exercise suggests that distinguishing between suspect and non-suspect choices improves the logit model's predictive power.

Finally, as an indirect test on our maintained hypothesis that people in the suspect and non-suspect groups share the same underlying utility parameters, after conditioning on demographics and prescription drug use, we leverage the panel structure of our data to repeat the estimation for four mutually exclusive sets of enrollment decisions: (1) choices made by enrollees who always make suspect choices (n=3,749); (2) suspect choices made by enrollees who sometimes make non-suspect choices (n=617); (3) non-suspect choices made by enrollees who sometimes make suspect choices (n=759); and (4) choices made by enrollees who always make non-suspect choices (n=4,706). The results, shown in Tables A6-A7, reveal that the estimated marginal rates of substitution between cost, variance, and quality are similar between groups 1 and 2, and between groups 3 and 4, despite some reduction in statistical significance. In other words, when people who switch between the suspect and non-suspect groups make non-suspect choices they behave in similar ways to the people who always make non-suspect choices. This provides some limited support for our approach to using non-suspect preference parameters to predict welfare effects for people in the suspect group.

VII. Evaluating Prospective Policies Designed to Simplify Choice Architecture

A. Preliminaries

[26] We exclude indicators for insurance brand because some new insurers joined the market in 2009 so we are unable to estimate indicators for them in 2008.

i. Solving for Endogenous Premiums

Firms may choose to adjust plan premiums in response to changes in choice architecture and Handel's (2013) simulation results suggest that the resulting welfare effects may be important due to adverse selection. We allow for this by assuming that firms will correctly anticipate how consumers will adjust their behavior and then reset their premiums to maintain the net revenue per enrollee that they earned prior to the policy. This is equivalent to assuming that CMS would accompany any change in choice architecture by maintaining the same plan approval and oversight processes that yielded the net revenues observed under the status quo.

For the baseline equilibrium we define the expected net revenue per enrollee in plan *k* as

$$\pi_k^0 = \frac{p_k^0}{.255} - z_k - \frac{1}{N}\sum_i \psi_{ik}^0 (g_{ik} - oop_{ik}). \tag{12}$$

Premiums are divided by 0.255 to reflect the fact that beneficiaries pay on average 24.5% of actual plan premiums, with the remainder subsidized by taxpayers. The second term, z_k, represents the average cost of plan management and operations per enrollee (e.g. customer service) which we assume to be constant over any changes in enrollment induced by policy. The last term is the insurer's expected cost of drugs for the average enrollee; g_{ik} is the gross cost of the drugs used by consumer *i* so that $g_{ik} - oop_{ik}$ represents expenditures paid by the insurer.

Equation (13) shows the fixed point problem that we solve to obtain the new vector of premiums,

$$\pi_k^1[\psi_{ik}^1(p_k^1), p_k^1] - \pi_k^0 = 0. \tag{13}$$

Because z_k is assumed to be constant it cancels out of the difference in (12). We observe p_k^0, g_{ik}, and oop_{ik} for all person-plan combinations from our data and we use our parameter estimates for suspect and non-suspect choices to calculate ψ_{ik}^0. All that remains is to solve for p_k^1. The main challenge in doing so is to recognize that choice probabilities change with adjustments to the premium. All else constant, increasing the premium of plan *k* will reduce the probability that people select it. Therefore, we iterate between solving for a vector of premiums to satisfy (13), conditional on ψ_{ik}^1, and updating ψ_{ik}^1 to reflect changes in the vector of premiums.

ii. Calculating Changes in Insurer Revenue and Government Expenditures

After solving for new vectors of plan premiums and choice probabilities we use the results to calculate changes in insurer revenue and government expenditures. Equation (14) defines the predicted change in insurer revenue per enrollee.

$$\Delta\pi = \frac{1}{N}\sum_i \sum_{k\in K} \psi_{ik}^1 \pi_k^1 - \frac{1}{N}\sum_i \sum_{j\in J} \psi_{ij}^0 \pi_j^0. \tag{14}$$

While the revenue per enrollee for each plan is held fixed by (13), the overall market revenue per enrollee may change due to changes in the way enrollees sort themselves across the available plans.[27] This allows for the possibility that changes to choice architecture may mitigate or exacerbate adverse selection consistent with Handel (2013). Equation (15) defines the corresponding change in average government spending per enrollee.

$$\Delta\tau = \frac{1}{N}\sum_i \sum_{k\in K} \psi_{ik}^1 \left[\frac{p_k^1(1-.255)}{.255}\right] - \frac{1}{N}\sum_i \sum_{j\in J} \psi_{ij}^0 \left[\frac{p_j^0(1-.255)}{.255}\right]. \tag{15}$$

The term in brackets represents the component of the total plan premium paid by taxpayers.

iii. <u>Calculating Bounds on Consumer Welfare, Insurer Revenue, and Government Spending</u>

In Section V we discussed our approach to bounding the welfare effect of inertia and bounding the effect of the policy on consumer behavior. Considering every possible combination of these bounds would yield four sets of results for each policy. In the interest of brevity we focus on the two most extreme scenarios. At one extreme we consider the case where the policy is "most effective" as a nudge in the sense that it causes the suspect group to behave like the non-suspect group <u>*and*</u> the inertia parameters estimated for the non-suspect group are interpreted as psychological bias and hence have no direct effect on welfare, i.e. for non-suspect using equation 9' and for suspect using equation 11' with V_{ik}^{n1} and ψ_{ik}^{n1}). At the other extreme, we consider the case where the policy is "least effective" as a nudge in that it does not change the decision making process for the suspect group *and* the inertia parameters for the non-suspect group are interpreted as the hassle cost of switching and/or preferences for latent plan attributes and hence are welfare relevant (i.e. using equations 9 and 11). All else constant, the most effective nudge would be expected to predict larger increases in consumer welfare from policies that simplify choice architecture.

[27] This also means that average revenue per enrollee may change for any insurer offering multiple plans.

B. Distributional Effects of a Menu Restriction

In early 2014, CMS proposed a series of changes to Medicare Part D that included a provision to limit each parent organization to offering only one basic and one enhanced plan in each region (Department of Health and Human Services 2014).[28,29] This would have forced some current enrollees to switch plans. While the proposal was controversial and has yet to be implemented, it provides a prospective opportunity to investigate the effects of a realistic menu restriction.

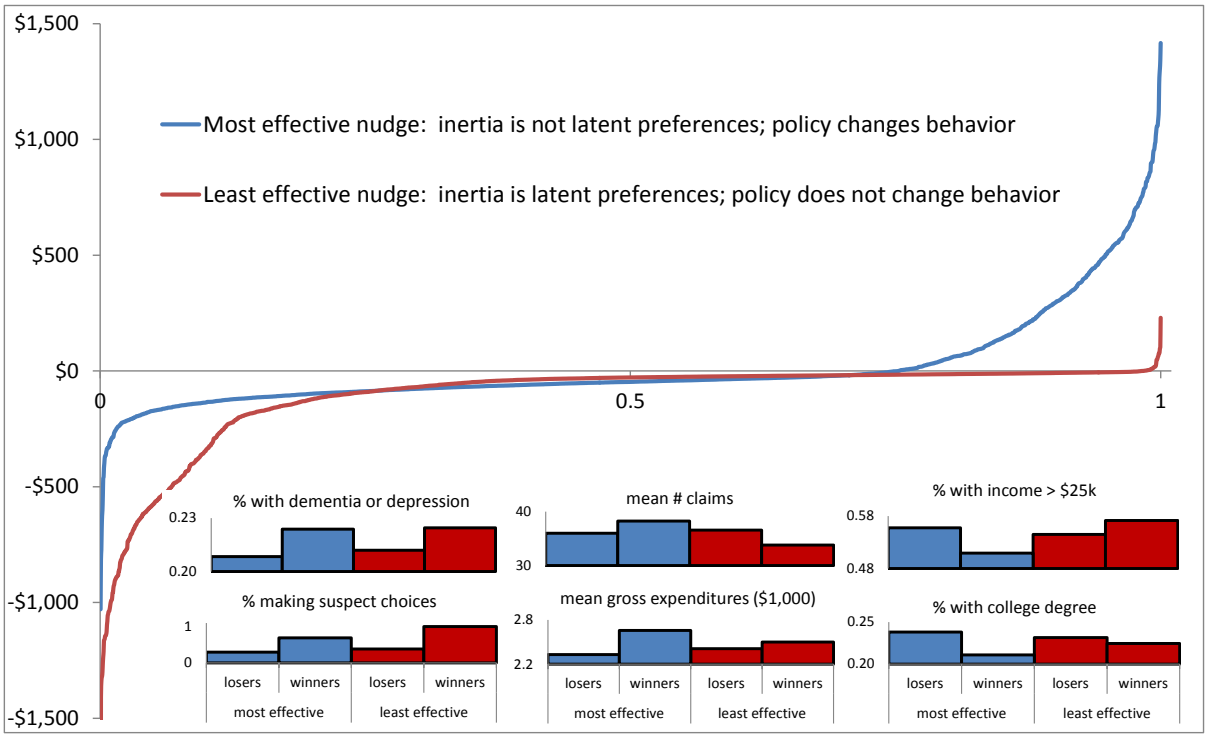
Our first policy experiment uses the set of enrollees and available plans in 2010—the last year of our enrollment sample—to simulate the welfare effects of the proposed menu restriction. Our data for that year describe 2,823 individuals, both new enrollees and those with experience. CMS must approve each PDP that an insurer offers, but the proposed regulation was unclear about how, exactly, CMS would determine which plans to retain. Therefore we start by assuming that CMS would require each sponsor to continue to offer their most popular plans; i.e. the single basic plan and the single enhanced plans with the highest enrollments. Then we consider alternative rules as robustness checks below. The menu restriction reduces the number of plans on the average enrollee's menu from 47 to 31.

The menu restriction affects consumer welfare in at least four ways. First, people may be made worse off when their utility maximizing plans are eliminated. Second, individuals who switch plans may incur hassle costs of switching. Third, individuals in the suspect group may be made better off if the policy forces them to switch out of a dominated plan or if the policy succeeds in reducing their inertia and nudging them to place greater emphasis on cost and risk reduction in ways that induce them to switch to plans that are cheaper, higher quality, and provide better insurance against health shocks. The magnitude of each of these gains (or losses) depends on which plans are eliminated and the relative benefits of switching. Finally, when enrollees switch plans their sorting behavior feeds into equilibrium premiums. As Handel (2013) points out, the direction of this effect is ambiguous. Increased sorting may increase or decrease adverse selection depending in part on whether the sorting is driven by suspect or non-suspect choosers.

[28] "Parent organizations" or "sponsors" are entities that contract with CMS to sell PDPs. They may include multiple brand names. Basic plans may differ in design but must be deemed actuarially equivalent to the standard benefits package for some representative enrollee(s). Enhanced plans offer supplemental benefits.

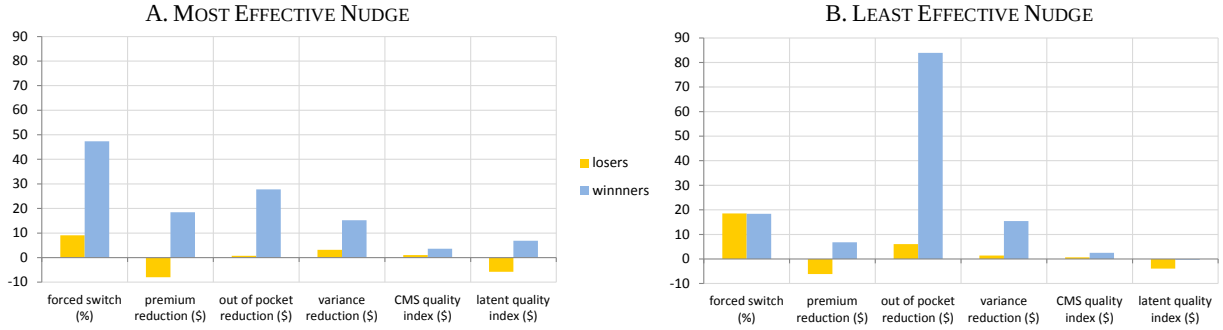
[29] The proposal included the rationale to "…ensure that beneficiaries can choose from a less confusing number of plans that represent the best value each sponsor can offer" (Federal Register 2014).

FIGURE 1: DISTRIBUTION OF WELFARE EFFECTS FROM A MENU RESTRICTION



Note: The figure shows CDFs of the expected change in welfare from limiting each insurer to selling one basic plan and one enhanced plan, assuming that CMS requires insurers to keep the plans with the highest current enrollment. The bar charts report demographics for the average enrollee with welfare gains ("winners") and losses ("losers") under alternative assumptions about the efficacy of the nudge.

FIGURE 2: MECHANISMS UNDERLYING THE WELFARE EFFECTS OF A MENU RESTRICTION



Note: The first column reports the share of winners and losers who are forced to switch because their chosen plans are eliminated. The next two columns report average reductions in premiums and out of pocket expenditures. The last three columns use the marginal utility of income for the non-suspect group to report the reduction in variance and the increases in plan quality in monetary equivalents.

TABLE 6: SUMMARY OF OUTCOMES FOR ALTERNATIVE MENU RESTRICTION RULES

	Max enrollment		Max frontier		Min expenditures		Max profit	
	most effective	least effective	most effective	least effective	most effective	least effective	most effective	least effective
Δ insurer revenue / enrollee ($)	6	36	66	120	-24	3	261	316
Δ govt. spending / enrollee ($)	4	17	57	86	-19	-9	210	240
Δ expected welfare / enrollee ($)	12	-108	29	-165	39	-134	28	-226
% enrollees with expected welfare gain	25	2	30	1	30	2	34	1

Note: The table shows the sensitivity of outcomes to the menu restriction rule. Max enrollment is the baseline that corresponds to figures 1 and 2. Max frontier retains the basic and enhanced plans with the highest shares of enrollees on the efficiency frontier. Min expenditure retains plans with the lowest average expenditures. Max profit allows insurers to retain the plans with the highest average profit per enrollee.

To summarize results we start by focusing on the case in which CMS requires each insurer to retain their basic and enhanced plans with the highest numbers of enrollees. Figure 1 summarizes the distributional effects on the beneficiary population. It shows CDFs of the expected consumer surplus under the "most effective" and "least effective" scenarios for the efficacy of the policy in nudging consumers [henceforth ME and LE]. The bar charts in the bottom half of the figure summarize the demographic characteristics of the people who have expected welfare gains (i.e. winners) and losses (i.e. losers). In both scenarios fewer than 25% of consumers are made better off by the policy, yet the policy appears notably progressive. The winners are less likely to have a college degree; they are more likely to have been diagnosed with cognitive illnesses; they tend to have higher gross drug expenditures (consumer expenditures + insurer expenditures); and they are far more likely to belong to the suspect group.

Figure 2 summarizes the mechanisms that drive the welfare effects. It reports the shares of winners and losers who are forced to switch because the policy eliminates their current plans, followed by their average reductions in their premiums, their average reductions in out of pocket expenditures, their average reductions in variance, and the average increases that they experience in the CMS quality index as well as the index of latent quality defined by the insurer dummy variables. The last three effects are converted to dollar equivalents by dividing the changes in each variable by the marginal utility of income for the non-suspect group.

In the ME scenario just under 25% of consumers are made better off. Nearly half of them are forced to switch plans but, importantly, there is assumed to be no direct utility cost of switching. Many of the people who are forced to switch (especially those in the suspect group) are better off from switching because their new plans provide more generous coverage. Furthermore, people in the suspect group now place more emphasis on cost and risk reduction when selecting a plan. After the policy, the average winner pays $18 less in premiums and $28 less in out of pocket costs. Their exposure to risk declines by an amount equivalent to a certain payment of $15 and they experience an improvement in plan quality worth just over $10 (summing the effects of the CMS index and insurer fixed effects). Nevertheless, most people experience welfare losses. A small number of consumers (particularly those in the non-suspect group) experience relatively large losses because the policy eliminates their utility maximizing plans. However, most of the losers experience relatively small losses that accrue because the policy increases adverse selection and raises their plan premiums.

In the LE scenario, only 2% of consumers are made better off. For most people, the utility cost of being forced to switch plans (due to the welfare relevance of inertia) more than offsets the cost savings, risk reduction, and improvements in plan quality experienced by switchers. The small fraction of winners all belong to the suspect group. Hence, if we think that inertia primarily reflects hassle costs and consumer preferences, then the menu restriction makes the vast majority of consumers worse off in order to produce small benefits for a small share of people in the suspect group because they are less able to choose inferior plans.

The first two columns of Table 6 summarize how the policy affects insurer revenue per enrollee, government spending per enrollee via the premium subsidy, and the average change in expected welfare per enrollee. The ME scenario predicts a slight increase in average consumer surplus ($12), driven by large gains for a small fraction of consumers. The LE scenario predicts a larger reduction (-$108). Both scenarios predict increases in government spending ($4 to $17) and insurer revenue ($6 to $36). Hence, one of the main effects of the policy is to transfer revenue from consumers and taxpayers to insurers.

The last six columns show comparable results for three other hypothetical rules for how CMS could determine which plans to keep on the menu: the plans that are on the efficiency frontier for the greatest number of people; the plans with the minimum average cost to the enrollee; and the plans with the highest net revenue per enrollee.[30] Our results on consumer welfare are qualitatively robust across these scenarios. The most striking difference is the order of magnitude increases in insurer revenue and government spending that occur when insurers are allowed to retain their highest profit plans. The more profitable plans tend to be the higher-premium ones that provide more risk reduction and have higher quality ratings. Hence, insurers would have strong incentives to persuade regulators to allow them to retain their more comprehensive plans at the expense of consumers and taxpayers. With approximately 7.7 million people currently participating in the standalone Medicare prescription drug markets the change in annual government transfers would range from a reduction of $0.07 to $0.14 billion under the minimum expenditure rule to an increase of $1.6 to $1.8 billion under the maximum profit rule.

C. Distributional Effects of Personalized Decision Support

[30] For profitability, we assume that there is sufficiently little variation in z_k within the set of plans offered by each insurer that it does not affect the ranking of plans by revenue per enrollee. Under this assumption the ranking of plans within each brand is defined by $\frac{p_k^0}{.255} - \frac{1}{N}\sum_i \psi_{ik}^0 (g_{ik} - oop_{ik})$.

Our second policy experiment evaluates the welfare effects of a hypothetical information campaign modeled on a randomized field experiment conducted by Kling, Mullainathan, Shafir, Vermeulen, and Wrobel (2012) [henceforth KMSVW]. Their analysis is motivated by the observation that while Medicare enrollees can learn about their personal PDP options and potential savings by calling 1-800-Medicare or using various cost calculators available online, a minority of enrollees report doing so (as seen in Table 1). KMSVW attribute this to "comparison friction" which they define as the wedge between available information and consumers' use of it. KMSVW tested an intervention in which several hundred treatment group enrollees (who agreed to participate in the experiment) were sent a decision support letter containing personalized information about their potential personal cost savings from switching to their lowest cost available plan. The letter also identified the name of the low cost insurer and contact information to initiate a switch. KMSVW observed a 7 percentage point increase in the rate at which the treatment group switched to their lowest cost plan relative to a control group that received a general letter with no personalized decision support, and an 11.5 percentage point increase in the overall switching rate for the treatment group.

In this experiment we estimate the heterogeneous welfare implications of a national rollout of the decision support tool in which the government mails letters to all enrollees that would be similarly worded to the one sent to the treatment group in KMSVW's study. Such a policy may affect welfare via several pathways. First, as the authors suggest, providing enrollees with personalized information may mitigate psychological biases and/or reduce information costs, making them better off. In the context of our model, this would be realized as increases in the switch rate and cost savings, as well as potential reductions in risk and increases in quality. Second, an important feature of the information campaign—if it were implemented by the government—is that it would necessarily be based on incomplete information about enrollees' drug needs. While CMS has full information about an individual's claims over their prior years in the PDP market, the individual may have private information about their own drug needs over the upcoming year. If enrollees with private information about changes in their drug needs choose to switch plans based on outdated information provided by CMS then these misinformed individuals could experience welfare losses.[31] Finally, increased switching initiated by a national rollout could induce feedback effects on premiums that would further affect welfare.

[31] In principle such a phenomenon could exist if the free but imperfect information from CMS reduces efforts for people to acquire private infor-

We use KMSVW's estimated treatment effects as moments that we can use to calibrate V_{ij}^{n1} and V_{ij}^{s1}. Specifically, in the ME scenario we multiply the estimated inertia parameters by $\omega_1(1+\omega_2 1\{j=j^*\})$ as shown in (16.a) and (16.b), where $1\{j=j^*\}$ is an indicator for whether plan j is the individual's minimum cost plan that would be featured as part of the information treatment. We calibrate ω_1 to generate a 7 percentage point increase in the rate at which the treatment group switches to their lowest cost plan relative to the baseline that we observe in the data, and we calibrate ω_2 to simultaneously generate an 11.5 percentage point increase in the overall switch rate subject to the constraints that $0 \leq \omega_1, \omega_2, \omega_1+\omega_2 \leq 1$.

$$(16.a) \qquad V_{ijt}^{n1} = \hat{\alpha}_{it}^{n} c_{ijt} + \hat{\beta}_{it}^{n}\sigma_{ijt}^{2} + \hat{\gamma}_{it}^{n} q_{jt} + \omega_1(1+\omega_2 1\{j=j^*\})\left(\hat{\eta}_{it}^{n}\Delta B_{ijt} + \hat{\delta}_{it}^{n}\Delta P_{ijt}\right).$$

$$(16.b) \qquad V_{ijt}^{s1} = \hat{\alpha}_{it}^{s} c_{ijt} + \hat{\beta}_{it}^{s}\sigma_{ijt}^{2} + \hat{\gamma}_{it}^{s} q_{jt} + \omega_1(1+\omega_2 1\{j=j^*\})\left(\hat{\eta}_{it}^{s}\Delta B_{ijt} + \hat{\delta}_{it}^{s}\Delta P_{ijt}\right).$$

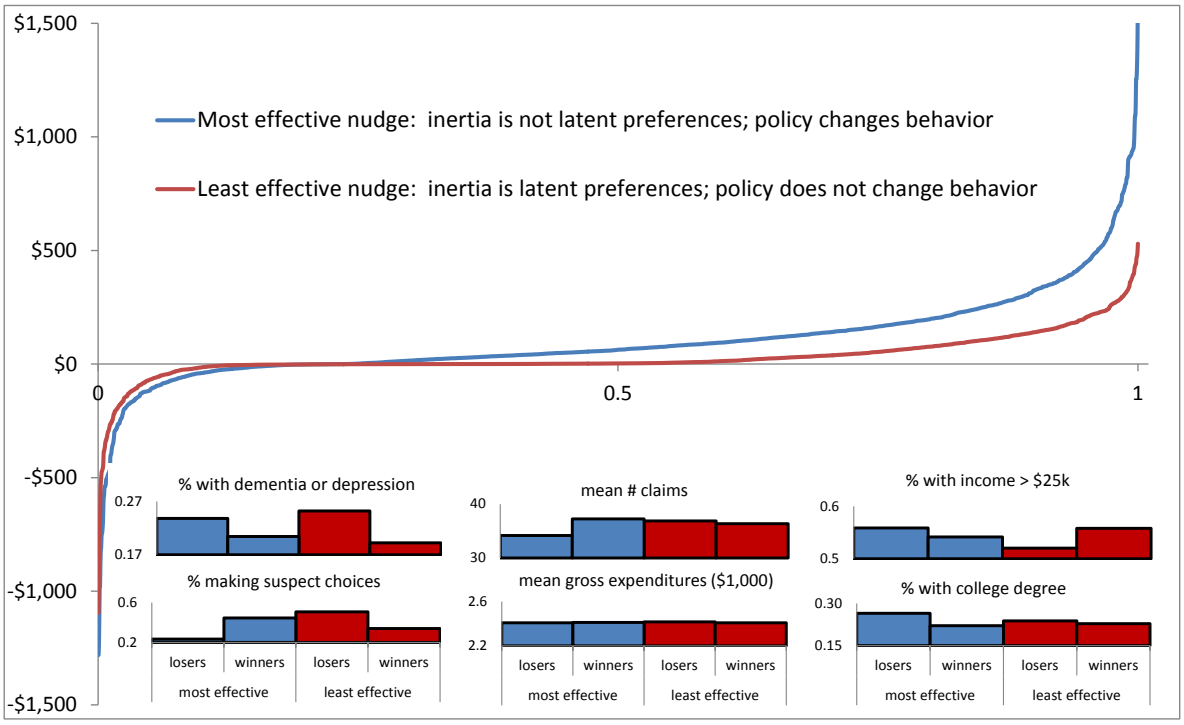
$$(16.c) \qquad V_{ijt}^{s1} = \hat{\alpha}_{it}^{s} c_{ijt} + \hat{\beta}_{it}^{s}\sigma_{ijt}^{2} + \hat{\gamma}_{it}^{s} q_{jt} + \hat{\eta}_{it}^{s}\Delta B_{ijt} + \hat{\delta}_{it}^{s}\Delta P_{ijt}.$$

In the LE scenario, there is assumed to be no change in the behavior of the suspect group so we use (16.a) and (16.c), in which case ω_1 and ω_2 will have to be larger than in the ME scenario in order to induce sufficient switching among the non-suspect group to replicate the treatment effects estimated by KSMVW.

Figure 3 summarizes the distributional effects of the personalized information treatment. In both the ME and LE scenarios the policy is welfare enhancing for more than two thirds of consumers. The winners are between 3 to 6 percentage points less likely to be diagnosed with dementia or depression and 1 to 4 percentage points less likely to have a college degree. There is virtually no difference between winners and losers in terms of gross drug expenditures.

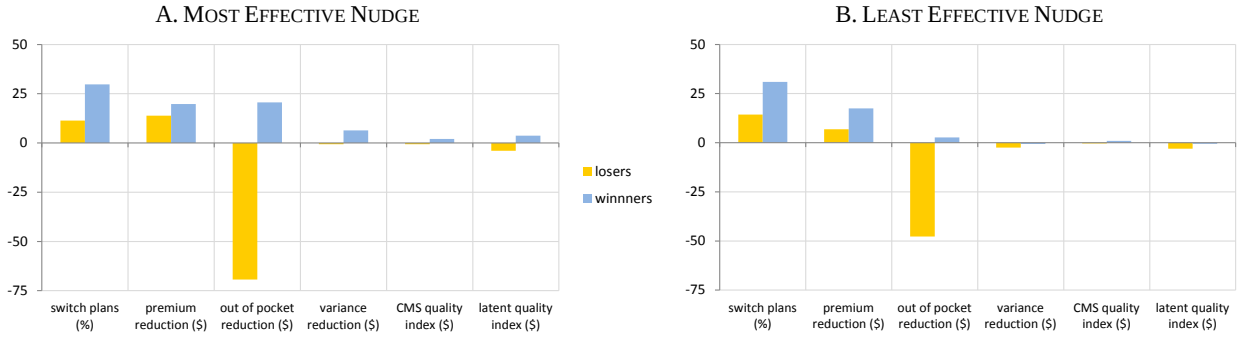
mation about their own future drug needs. Carlin, Gervais, and Manso (2013) provide and extended exploration of these ideas more generally.

FIGURE 3: DISTRIBUTION OF WELFARE EFFECTS FROM A PERSONALIZED DECISION SUPPORT TOOL



Note: The figure shows CDFs of the expected change in welfare from a personalized decision support tool that is based on the field experiments of Kling et al. (2012). The model is calibrated to reproduce their estimated treatment effects on the rates at which people switch plans. The bar charts report demographics for the average enrollee with welfare gains and losses under alternative assumptions about the efficacy of the nudge.

FIGURE 4: MECHANISMS UNDERLYING THE WELFARE EFFECTS OF DECISION SUPPORT



Note: The first column reports the share of winners and losers who choose to switch plans following the information treatment. The next two columns report average reductions in premiums and out of pocket expenditures. The last three columns use the marginal utility of income for the non-suspect group to report the reduction in variance and the increases in plan quality in monetary equivalents.

TABLE 7: SUMMARY OF OUTCOMES AND SENSITIVITY TO DECISION MAKERS' EXPECTATIONS

	Unbiased Expectations		Myopia	
	most effective	least effective	most effective	least effective
Δ insurer revenue / enrollee ($)	-69	-50	-73	-54
Δ govt. spending / enrollee ($)	-54	-41	-43	-35
Δ expected welfare / enrollee ($)	102	29	193	75
% enrollees with expected welfare gain	76	67	91	76

Note: The table shows the sensitivity of outcomes to the assumed form of decision makers' expectations for their own drug needs in the upcoming year. The baseline scenario that corresponds to figures 3 and 4 (perfect foresight) assumes that decision makers accurately forecast changes in their drug needs. The myopia scenario assumes that decision makers expect their future drug needs to be identical to the prior year.

In the ME scenario the winners are more likely to be in the suspect group. The information treatment induces them to place more emphasis on cost savings and many of them switch plans as a result. Figure 4 shows that the winners enjoy an average reduction in premiums and OOP expenditures of $41 along with risk reduction and quality improvements worth another $12. The losers have substantially higher OOP expenditures. This is because the low cost plan that is featured by the information treatment is the one that minimizes their expenditures based on their prior year of drug use. A small share of people who experience significant health shocks would spend substantially more in the recommended plan than in the plan that they actually chose for themselves. These individuals are concentrated in the non-suspect group. This illustrates the potential welfare losses that can arise from a nudge based on incomplete information. More broadly, this suggests a tradeoff between the potential benefits of simplifying the presentation of information and the potential costs of deemphasizing important details about the assumptions underlying that information.

In the LE scenario the winners are 4 percentage points more likely to have incomes over $25k and 17 percentage points less likely to be in the suspect group. This is because the information treatment is assumed to not change the behavior of individuals in the suspect group. The individuals with the largest welfare gains tend to be people in the non-suspect group who benefit because the information treatment lowers their cost of collecting information about alternative plans and induces them to actively switch plans. Overall, 67% of enrollees are better off as a result of the policy.

By inducing consumers to switch to lower cost plans, the information treatment reduces insurer revenue and government expenditures per enrollee. These figures are shown in the first two columns of Table 7. Recall that our model of decision making assumes that consumers have unbiased expectations of their actual drug use in the upcoming year. This could cause us to understate the policy's benefits. If consumers are myopic in the sense that they expect their drug use to be the same as the prior year then there is less scope for the information treatment to be welfare reducing. The last two columns of Table 7 demonstrate this and show that when we repeat the estimation and simulation based on the assumption that consumers are myopic, then between 76% and 91% of consumers benefit from the policy and the average change in welfare is an increase of between $75 and $193. Adverse selection still makes some people worse off by raising plan

premiums. Across the scenarios in Table 7 the implied annual reduction in government expenditures ranges from $0.27 to $0.41 billion.

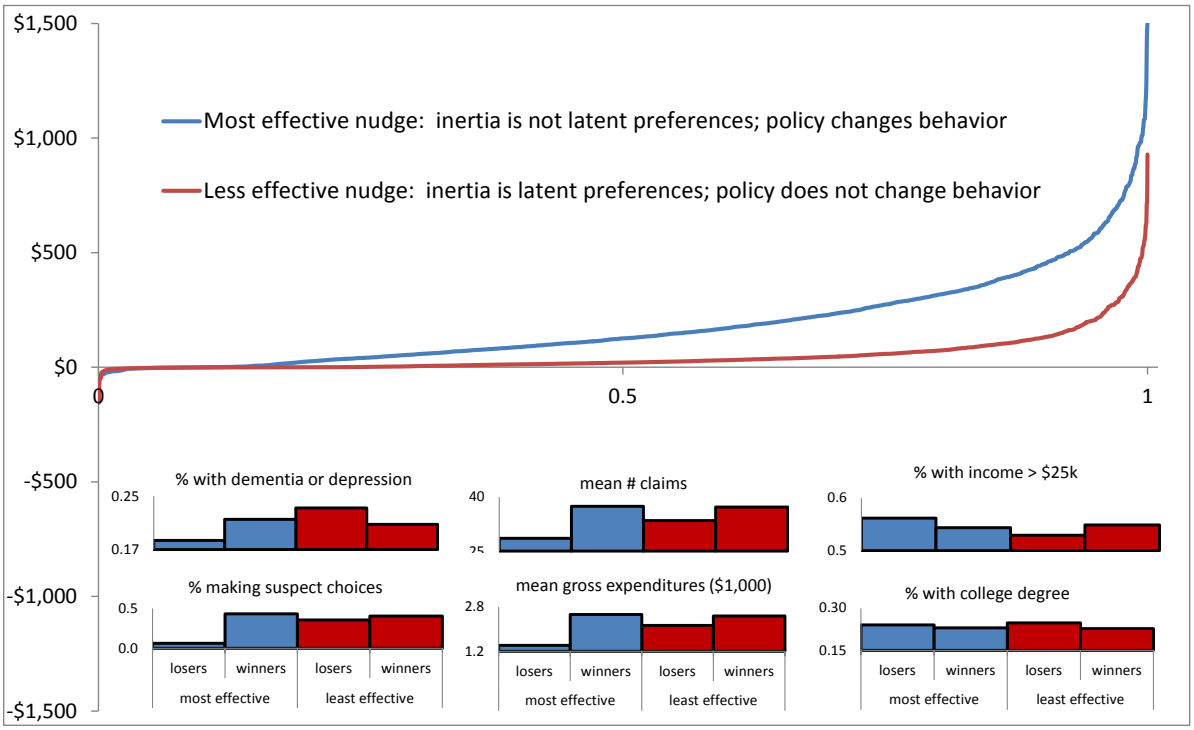
D. Distributional Effects of Default Assignment to a Low Cost Plan

Our final policy experiment evaluates the welfare effects of replacing CMS's current revealed preference approach to defining each person's reenrollment default plan with an alternative policy that would set the default to be the plan that would minimize each enrollee's costs. We envision the policy being implemented as a stronger version of the decision support nudge. Not only would enrollees be informed of their minimum cost options, they would be automatically assigned to those options unless they chose to opt out by actively switching to a different plan. As before, we assume CMS would predict each enrollee's minimum cost plan using their drug claims from the prior year. Consistent with CMS's current approach, first-time enrollees would still be required to make active decisions.

In the ME scenario, the policy completely erases inertia for enrollment in the new low-cost default. Nevertheless, some consumers may still prefer their original plans if those plans provide greater quality or variance reduction. Assuming it is costless for enrollees to opt out and continue in their old plans, the policy could reduce consumer welfare for only two reasons: (1) endogenous increases in premiums, or (2) (mis)assignment to plans requiring higher expenditures due to changes in drug needs. Figure 5 shows that across the distribution of enrollees the aggregate effect of these two mechanisms is almost always dominated by the aggregate effect of lower expenditures and the elimination of inertia. Overall, 90% of consumers have gains in expected welfare. Figure 6 shows that approximately half of them choose to stay in the low cost default.[32] The others choose to opt out and remain in their current plans. The average winner experiences a substantial reduction in their premium ($56) and OOP expenditures ($85) and a smaller reduction in the monetary value of plan quality ($38). The policy tends to benefit people who have more drug claims, more drug expenditures, higher rates of cognitive illness, and who are more likely to be among the suspect group.

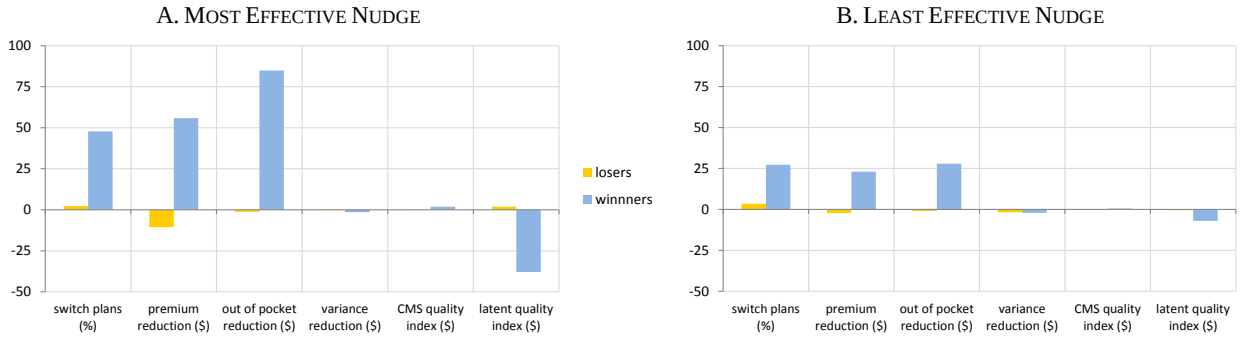
[32] While the aggregate probability of remaining in the low cost default is 48% the probabilistic nature of the logit model requires that there is a positive probability that every enrollee chooses to remain in their low cost default. The probability weighted consumer surplus from this potential outcome feeds into each individual's expected welfare gain.

FIGURE 5: DISTRIBUTION OF WELFARE EFFECTS FROM ASSIGNMENT TO A DEFAULT PLAN



Note: The figure shows CDFs of the expected change in welfare from automatically assigning people to default plans, assuming it is costless to opt out. People are automatically assigned to the plan that would minimize their expenditures based on their prior year of drug use. The bar charts report demographics for the average enrollee with welfare gains and losses under alternative assumptions about the efficacy of the nudge.

FIGURE 6: MECHANISMS UNDERLYING THE WELFARE EFFECTS OF DEFAULT ASSIGNMENT



Note: The first column reports the share of winners and losers who end up in plans that differ from their true default plans following the counterfactual assignment. The next two columns report average reductions in premiums and out of pocket expenditures. The last three columns use the marginal utility of income for the non-suspect group to report the reduction in variance and the increases in plan quality in monetary equivalents.

TABLE 8: SUMMARY OF OUTCOMES AND SENSITIVITY TO DECISION MAKERS' EXPECTATIONS

	Unbiased Expectations		Myopia	
	more effective	less effective	more effective	less effective
Δ insurer revenue / enrollee ($)	-254	-92	-261	-93
Δ govt. spending / enrollee ($)	-144	-54	-154	-57
Δ expected welfare / enrollee ($)	186	49	275	67
% enrollees with expected welfare gain	90	82	92	84
opt out cost needed to set average Δ in expected welfare to zero ($)	470	73	816	100

Note: The table shows the sensitivity of outcomes to the assumed form of decision makers' expectations for their own drug needs in the upcoming year. The first four rows assume no opt out cost. See the note to Table 7 and the text for additional details and definitions.

In the LE scenario being assigned to a default plan does not eliminate the hassle cost of learning to navigate a plan offered by a different insurer (e.g. prior authorization paperwork, new pharmacy networks, new customer service protocols). To account for this we recalibrate the model so that the policy reduces the cost of switching to the low-cost default from $\hat{\eta}_{it}\Delta B_{ijt} + \hat{\delta}_{it}\Delta P_{ijt}$ to $(\hat{\eta}_{it} - \hat{\delta}_{it})\Delta B_{ijt}$. Under this interpretation, the welfare-relevant hassle costs are the difference in the estimated cost of switching between brands relative to switching within brands. The continued presence of navigation costs reduces the share of enrollees choosing their assigned default to 23% which, in turn, reduces the average changes in premiums, oop expenditures, and plan quality experienced by winners and losers albeit with no qualitative changes in the pattern of results.[33]

Even with less than one quarter of consumers switching, the reduction in premiums translates to savings to taxpayers of $54 per enrollee or about $0.42 billion per year. Table 8 shows that the share of consumers who benefit, their average welfare gain, and the implications for government spending and insurer revenue are virtually unchanged if we repeat the estimation and simulation under the assumption that consumers have myopic expectations of their own drug needs for the upcoming year.

Finally, the last column of Table 7 illustrates the importance of what may seem like a small detail—the design of the opt out feature. We repeat the simulation except that now we make it costly for enrollees to switch back to their previously-chosen plans. Intuitively, people may incur a cost from paying attention to the new policy, learning how the opt out feature works, determining whether they expect to prefer their newly assigned default to their old plan and, if not, exercising their opt out option. Under the assumption that everyone faces the same opt out cost we solve for the cost needed to set the average change in expected welfare to zero. It ranges from a low of $73 in the LE scenario with unbiased expectations to a high of $816 in the ME scenario under myopia. When people pay a significant utility cost of opting out, some of them choose the newly assigned default even though it is welfare reducing.

VIII. Robustness Checks

Table 9 reports the sensitivity of our main estimates for consumer welfare, taxpayer spend-

[33] This approach may still overstate benefits to the extent that $\hat{\eta}$ and $\hat{\delta}$ represent latent preferences. As we increase the post-policy cost of switching to the new default option to $\hat{\eta}\Delta B_{ijt} + \hat{\delta}\Delta P_{ijt}$ the benefits to consumers approach zero. The extreme case in which $\hat{\eta}$ and $\hat{\delta}$ are entirely latent preferences is equivalent to reverting to the pre-policy equilibrium in which case the policy has no effect on consumer welfare.

ing, and insurer revenue to several alternative specifications of our model. The columns match the main policy scenarios summarized in the tables and figures of the previous section, and the first three rows of Panel A repeat the results from those scenarios for convenience.

Panel B summarizes results from the alternative approach in which we replace the assumption that consumers have perfect foreknowledge about their future drug costs and distribution with an assumption that their expectations are fully myopic, e.g. based on their prior year's drug consumption (the "ex ante" approach reported in Ketcham et al. 2012 and Ketcham, Lucarelli and Powers 2015). This involves using the ex ante measures of expected cost to reclassify choices as suspect or non-suspect, estimating the logit models based on that information, and repeating the policy analysis. The results are qualitatively identical to those from the ex post assumption reported in the article. The main quantitative difference is that the welfare gains from personalized decision support and default assignment are larger because the ex ante approach assumes there is no information asymmetry between consumers and the government.

Panels C and D of Table 9 report the sensitivity of our main results to two alternative approaches to defining suspect choices under the baseline approach using ex post drug claims to determine plan costs and choice of dominated plan. Panel C ignores the MCBS knowledge question and defines choices as suspect based solely on dominated plans. Panel D uses a more inclusive definition based on the union of dominated plan choices, the knowledge question, and being able to reduce expenditures by more than 50%. Moving from C to D increases the set of choices labeled as suspect from 18% to 47%, with the base results in Panel A fitting logically between these figures. Altering how suspect choices are defined has little effect on our main results. This remains true if we further expand the set of suspect choices to include people who could have reduced expenditures by more than 33% (Table A8). The reason is that of the three suspect choice indicators considered here, the choice of a dominated plan has the largest effects on our estimates for θ^s. This means that when we classify a greater share of choices as suspect, the difference between θ^s and θ^n declines. More people benefit from certain simplifications to choice architecture, but the average gain among those who benefit is smaller. These effects offset each other in a way that leads to small increases in expected welfare in some scenarios and small decreases in expected welfare in others.

TABLE 9—ROBUSTNESS CHECKS ON OUR MAIN RESULTS

	Menu Restriction		Decision Support		Default Assignment	
	most effective	least effective	most effective	least effective	most effective	least effective
			A. Baseline results			
Δ insurer revenue / enrollee ($)	6	36	-69	-50	-254	-92
Δ govt. spending / enrollee ($)	4	17	-54	-41	-144	-54
Δ expected welfare / enrollee ($)	12	-108	102	29	186	49
% enrollees with expected welfare gain	25	2	76	67	90	82
			B. Enrollees expect their drug needs to be the same as last year			
Δ insurer revenue / enrollee ($)	7	34	-73	-54	-261	-93
Δ govt. spending / enrollee ($)	3	17	-43	-35	-154	-57
Δ expected welfare / enrollee ($)	22	-141	193	75	275	67
% enrollees with expected welfare gain	24	1	91	76	92	84
			C. Suspect choices based on dominated plans only			
Δ insurer revenue / enrollee ($)	15	40	-62	-52	-247	-93
Δ govt. spending / enrollee ($)	10	20	-48	-44	-139	-55
Δ expected welfare / enrollee ($)	4	-116	111	21	198	48
% enrollees with expected welfare gain	21	1	77	70	88	84
			D. Suspect choices expanded to include potential savings > 50%			
Δ insurer revenue / enrollee ($)	-27	34	-101	-42	-285	-86
Δ govt. spending / enrollee ($)	-16	16	-77	-36	-162	-51
Δ expected welfare / enrollee ($)	31	-90	88	22	178	44
% enrollees with expected welfare gain	30	3	72	67	90	82
			E. Exclude mid-year enrollment decisions			
Δ insurer revenue / enrollee ($)	25	49	-53	-43	-220	-81
Δ govt. spending / enrollee ($)	17	26	-51	-40	-122	-47
Δ expected welfare / enrollee ($)	-23	-108	74	28	175	47
% enrollees with expected welfare gain	22	2	72	67	90	83
			F. Exclude beneficiaries who get help choosing plans			
Δ insurer revenue / enrollee ($)	17	50	-77	-48	-268	-93
Δ govt. spending / enrollee ($)	14	29	-62	-41	-156	-55
Δ expected welfare / enrollee ($)	10	-88	91	25	176	45
% enrollees with expected welfare gain	25	3	74	70	90	82

Note: The table summarizes the sensitivity of our main results to alternative specification of the model. Panel A repeats our baseline results for convenience. The numbers match those in tables 6, 7, and 8. Panels B through F show comparable results after altering the baseline specification in various ways. Panel B assumes that decision makers expect the upcoming year's drug needs to be the same as last year. Panel C uses a more exclusive definition for suspect choices based entirely on the selection of a dominated plan. Panel D uses a more inclusive definition for suspect choices that includes anyone who could reduce their drug spending by 50% or more by choosing a different plan. Panel E drops mid-year enrollment decisions that occur, for example, because the enrollee turns 65 in the middle of the year. Panel F drops beneficiaries who had help choosing a plan or who relied on a proxy to choose a plan for them. See the text for additional details.

As a next set of robustness checks, we refine the sample in multiple ways. In Panel E we exclude 640 enrollees who entered the market mid-year. A potential concern in that they may have been forward looking with respect to the following year's drug needs at the time they made their enrollment decisions, especially as they neared or entered the open enrollment period for the following year. Given that they represent a small share of our overall sample (7%) is it not surprising that dropping them has little effect on our results. In Panel F we drop the 3,740 enrollees (38% of our sample) who had help choosing a plan or relied on a proxy to choose a plan for them. While the smaller sample reduces the statistical significance in the relationships between observed demographics and the probability of making an uninformed decision, the logit estimates and subsequent policy implications are virtually identical to the full sample. This suggests that while the research value of having access to better information on how family, friends, and advisors influence decision making is self-evident, in our context of Medicare Part D it does not alter the predicted effects of policy reforms.

IX. Summary

We have developed a structural model capable of evaluating who would win and who would lose from a wide range of paternalistic reforms in a differentiated product market. We used the model to evaluate three prospective policies that have been proposed to simplify markets for prescription drug insurance created under Medicare Part D. Our analysis was enabled by a novel combination of administrative records and survey data on consumers' knowledge of the market, their enrollment decisions, and the financial consequences of those decisions for a nationally representative sample of the Medicare population. We used the data to first identify which consumers appear to make informed decisions that reveal their preferences to us then to estimate separate models of decision making for the informed and misinformed groups.

The results from our policy experiments suggest that CMS's recent proposal to simplify the choice process by reducing the number of drug plans would reduce welfare for the median consumer and increase transfers from taxpayers to insurers. In contrast, our results suggest that providing personalized information about the potential savings from switching plans and assigning people to low-cost default plans would benefit the median enrollee. However, these gains are always less than 19% of consumer expenditures, typically under 10%, and are often overshadowed by transfers from insurers to taxpayers. We note three limitations with our study. First, our

analysis largely excludes supply side responses to the prospective policies apart from the adjustments in premiums and the insurers' decisions about which plans to provide under menu restrictions. Second, our analysis does not embed any responses by consumers in their decisions about whether to participate in the PDP market or not. Given the large taxpayer subsidies to all PDP enrollees, such enrollment decisions likely have large effects on expected consumer surplus and taxpayer costs to the extent that such decisions change under the prospective policies. Third, our study holds constant the drugs consumed across plans and under alternative policies, again excluding some potentially welfare-relevant changes from the policies. We consider each of these limitations as important avenues for further research.

REFERENCES

Abaluck, Jason, and Jonathan Gruber. 2011. "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program." *American Economic Review*. 101(4): 1180-1210.

Ackerberg, Daniel, Lanier Benkard, Steven Berry, and Ariel Pakes. Econometric Tools for Analyzing Market Outcomes. In: Heckman J, Leamer E The Handbook of Econometrics. Amsterdam: North-Holland; 2007. pp. Chapeter 63.

Agarwal, Sumit, John Driscoll, Xavier Gabaix, and David Laibson. 2009a. "The Age of Reason: Financial Decisions over the Life-Cycle and Implications for Regulation." *Brookings Papers on Economic Activity*, Fall: 51:117.

Agarwal, Sumit and Bhashkar Mazumder. 2013. "Cognitive Abilities and Household Financial Decision Making". *American Economic Journal: Applied Economics*. 5(1): 193-207.

Allcott, Hunt and Dmitry Taubinsky. 2015. "Evaluating Behaviorally-Motivated Policy: Experimental Evidence from the Lightbulb Market." *Forthcoming, American Economic Review.*

Allcott, Hunt and Judd B. Kessler. 2015. "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons." *NBER Working Paper #21671*.

Ambuehl, Sandro, B. Douglas Bernheim and Annamaria Lusardi. 2014. "The Effect of Financial Education on the Quality of Decision Making". *NBER Working Paper #20618*.

Bayer, Patrick, Fernando Ferreira, and Robert McMillan. 2007. "A Unified Framework for Measuring Preferences for Schools and Neighborhoods." *Journal of Political Economy*, 115(4): 588-638.

Bayer, Patrick and Christopher Timmins. 2005. "On the Equilibrium Properties of Locational Sorting Models." *Journal of Urban Economics*, 57(3): 462-77.

Bernheim, B. Douglas and Antonio Rangel. 2009. "Beyond Revealed Preference: Choice-Theoretic Foundations for Behavioral Welfare Economics-super." *The Quarterly Journal of Economics*, 124(1): 51-104.

Bernheim, B. Douglas, Andrey Fradkin, and Igor Popov. 2015. "The Welfare Economics of Default Options in 401(k) Plans." *Forthcoming in American Economic Review.*

Busse, Meghan, Devin Pope, Jaren Pope, and Jorge Silva-Risso. 2015 "The Psychological Effect of Weather on Car Purchases". *Forthcoming in Quarterly Journal of Economics.*

Camerer, Colin, Samuel Issacharoff, George Lowenstein, Ted O'Donoghue, and Matthew Rabin. 2003. "Regulation for Conservatives: Behavioral Economics and the Case for "Asymmetric Paternalism"". *University of Pennsylvania Law Review*. 151(3): 1211-1254.

Carlin, Bruce, Simon Gervais, and Gustavo Manso. 2013. "Libertarian Paternalism, Information Production, and Financial Decision Making." *Review of Financial Studies*, 26(9): 2204-28.

Chetty, Raj, John N. Friedman, Soren Leth-Petersen, Torben Heien Nielsen, and Tore Olsen. 2015. "Active vs. Passive Decisions and Crowd-Out in Retirement Savings Accounts: Evidence from Denmark". *Forthcoming in the Quarterly Journal of Economics.*

Choi, Syngjoo, Shachar Kariv, Wieland Müller and Dan Silverman. 2014. "Who is (More) Rational?" *American Economic Review*, 104(6): 1518-50.

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

Fang, Hanming, Michael P. Keane, and Dan Silverman. 2008. "Sources of Advantageous Selection: Evidence from the Medigap Insurance Market." *Journal of Political Economy*, 116(2): 303-350.

Federal Register. 2014. "Policy and Technical Changes to the Medicare Advantage and the Medicare Prescription Drug Benefit Programs for Contract Year 2015 (CMS-4159-P)." January 10: 1917-2073. Accessed online September 1, 2015, http://www.gpo.gov/fdsys/pkg/FR-2014-01-10/html/2013-31497.htm.

Fleurbaey, Marc and Erik Schokkaert. 2013. "Behavioral Welfare Economics and Redistribution." *American Economic Journal: Microeconomics*, 5(3): 180-205.

Galiani Sebastian, Alvin Murphy, and Juan Pantano. 2015. "Estimating Neighborhood Choice Models: Lessons from a Housing Assistance Experiment". *Forthcoming in the American Economic Review*.

Handel, Benjamin R. 2013. "Adverse selection and inertia in health insurance markets: When nudging hurts." *American Economic Review*, 103(7): 2643-2682.

Handel, Benjamin R. and Jonathan T. Kolstad. 2015. "Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare." *Forthcoming, American Economic Review*.

Handel, Benjamin R., Jonathan T. Kolstad, and Johannes Spinnewijn. 2015. "Information Frictions and Adverse Selection: Policy Interventions in Health Insurance Markets." *NBER Working Paper #21759.*

Health and Human Services. 2014. "Proposed HHS Notice of Benefit and Payment Parameters for 2016 Fact Sheet." November 20, 2014. Accessed online November 27, 2015, https://www.cms.gov/cciio/resources/fact-sheets-and-faqs/downloads/fact-sheet-11-20-14.pdf.

Heiss, Florian, Daniel McFadden, and Joachim Winter. 2010. "Mind the Gap! Consumer Perceptions and Choices of Medicare Part D Prescription Drug Plans." In *Research Findings in the Economics of Aging*. ed. David A. Wise. Chicago, IL: University of Chicago Press.

Howell, Benjamin L., Jennifer Wolff, and Bradley Herring. 2012. "Medicare Beneficiary Knowledge of the Part D Program and Its Relationship with Voluntary Enrollment." *Medicare Research & Review*. 2(4): E1-E21.

Hsu, John, Vicki Fung, Mary Price, Jie Huang, Richard Brand, Rita Hui, Bruce Fireman, and Joseph P. Newhouse. 2008. "Medicare beneficiaries' knowledge of Part D prescription drug program benefits and responses to drug costs." *Journal of the American Medical Association*, 299(16): 1929{1936.

Kahneman, Daniel, Peter P. Wakker, and Rakesh Sarin. 1997. "Back to Bentham? Explorations of Experienced Utility." *Quarterly Journal of Economics* 112(2): 375-405.

Kariv, Shachar and Dan Silverman. 2013. "An Old Measure of Decision-making Quality Shed New Light on Paternalism." *Journal of Institutional and Theoretical Economics* 169(1): 29-44.

Keane, Michael and Nada Wasi. 2013. "Comparing alternative models of heterogeneity in consumer choice behavior." *Journal of Applied Econometrics* 28: 1018-1045.

Keane, Michael P. and Kenneth I. Wolpin. 2007. “Exploring the usefulness of a non-random holdout sample for model validation: Welfare effects on fe-male behavior.” International Economic Review 48(4): 1351-1378.

Ketcham, Jonathan D., Claudio Lucarelli, Eugenio Miravete and Christopher Roebuck. 2012. “Sinking, Swimming or Learning to Swim in Medicare Part D.” *American Economic Review*. 102(6): 2639-2673.

Ketcham, Jonathan D., Claudio Lucarelli, and Christopher Powers. 2015. “Paying Attention or Paying Too Much in Medicare Part D.” *American Economic Review*. 105(1): 204-233.

Ketcham, Jonathan D., Nicolai Kuminoff, and Christopher Powers. 2015. “Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program: Comment.” *Forthcoming, American Economic Review.*

Kling, Jeffrey, Sendhil Mullainathan, Eldar Shafir, Lee Vermeulen, and Marian V. Wrobel. 2012. “Comparison Friction: Experimental Evidence from Medicare Drug Plans.” *Quarterly Journal of Economics*, 127(1): 199-235.

Lancaster, Kelvin J. 1966. “A New Approach to Consumer Theory.” *Journal of Political Economy*, 74(2): 132-57.

Leggett, Christopher G. 2002. “Environmental Valuation with Imperfect Information”. *Environmental and Resource Economics*. 23: 343-355.

Liebman, Jeffrey and Richard Zeckhauser. 2008. “Simple Humans, Complex Insurance, Subtle Subsidies. *NBER Working Paper 14330*.

Lucarelli, Claudio, Jeffrey Prince, and Kosali Simon. 2012. “The Welfare Impact of Reducing Choice in Medicare Part D: A Comparison of Two Regulation Strategies.” *International Economic Review*, 53(4): 1155-1177.

MedPAC, 2006. “Chapter 8. How beneficiaries learned about the drug benefit and made plan choices.” In *Report to the Congress, Increasing the Value of Medicare.*

McFadden, Daniel. 2006. “Free Markets and Fettered Consumers.” *American Economic Review*, 96 (1): 5–29.

Madrian, Brigitte, and Dennis F Shea. 2001. “The Power of Suggestion: Inertia in 401(k) Participation and Savings Behavior.” *The Quarterly Journal of Economics*, 116(4): 1149-87.

Pope, Jaren C. 2008. “Buyer Information and the Hedonic: the Impact of a Seller Disclosure on the Implicit Price of Airport Noise.” *Journal of Urban Economics.* 63: 498-516.

Small, Kenneth A. and Harvey S. Rosen. 1981. "Applied Welfare Economics with Discrete Choice Models." *Econometrica*. 49(1): 105-130.

Stigler, George J. and Gary S. Becker. 1977. "De Gustibus Non Est Disputandum". *American Economic Review*. 67(2): 76-90.

Taubinsky, Dmitry and Alexander Robert Rees-Jones. 2015. "Attention Variation and Welfare: Theory and Evidence from a Tax Salience Experiment." *Working Paper*.

Thaler, Richard H. and Cass R. Sunstein. 2008. *Nudge*. New York: Penguin Books.

Train, Kenneth. 2009. *Discrete Choice Methods with Simulation.* Cambridge University Press.

Tymula, Agnieszka, Lior A. Rosenberg Belmaker, Lital Ruderman, Pau. W. Glimcher, and Ifat Levy. 2013. "Like Cognitive Function, Decision Making Across the Life Span Shows Profound Age-Related Changes. *Proceedings of the National Academy of Sciences*, 110(42): 17143-17148.

US Department of Health and Human Services. 2014. "2014 CMS Statistics." Centers for Medicare & Medicaid Services Office of Information Products and Data Analytics CMS Pub. No. 03510, August. Accessed online on September 3, 2015 at https://www.cms.gov/Research-Statistics-Data-and-Systems/Statistics-Trends-and-Reports/CMS-Statistics-Reference-Booklet/Downloads/CMS_Stats_2014_final.pdf.

Wiswall, Matthew and Basit Zafar. 2015. "Determinants of College Major Choice: Identification using an Information Experiment". Forthcoming in *Review of Economic Studies.*

Woodward, Susan E and Robert E Hall. 2012. "Diagnosing Consumer Confusion and Sub-Optimal Shopping Effort: Theory and Mortgage-Market Evidence." *American Economic Review*, 102(7): 3249-3276.

SUPPLEMENTAL APPENDIX: FOR ONLINE PUBLICATION

TABLE A1—COMPARING MCBS SAMPLE MEANS WITH ADMINISTRATIVE DATA

	2006	2007	2008	2009	2010
	Medicare beneficiary survey sample				
age	77	77	78	78	78
% female	62	62	62	62	62
white (%)	94	93	93	94	94
Alzheimer's or dementia (%)	7	8	9	10	11
Depression (%)	9	8	10	11	11
number of available brands	20	24	23	23	20
number of available plans	43	56	55	50	47
premium ($)	363	362	406	476	513
out-of-pocket costs ($)	1,010	842	873	920	903
mean potential savings, ex post ($)	546	347	295	332	337
	Random 20% Sample of all Part D Enrollees				
age	76	76	76	76	76
% female	63	64	63	63	62
white (%)	93	92	92	92	93
Alzheimer's or dementia (%)	7	9	9	10	10
Depression (%)	9	9	10	10	11
number of available brands	19	24	22	23	20
number of available plans	43	56	55	50	47
premium ($)	362	369	415	487	516
out-of-pocket costs ($)	994	890	857	892	886
mean potential savings, ex post ($)	521	355	298	337	333

Note: The top half of the table reports means based on enrollees in the merged administrative-MCBS sample that we use for estimation. The bottom half of the table reports means based on a random 20% sample of all individuals who enrolled in Medicare Part D for the entire year. The two data sets differ in that 6.5% of the observations in our merged sample are individuals who enrolled during the middle of the year. We drop these 640 individuals before calculating sample means in order to ensure comparability between the two data sets.

TABLE A2—ASSOCIATION BETWEEN MCBS KNOWLEDGE QUESTION AND MARKET OUTCOMES

	Unconditional	active choice, no help	active choice, help	active choice, no help, dominated plan choices	active choice, help, dominated plan choices
Chose a dominated plan	-0.002	0.021	-0.007		
	[0.008]	[0.015]	[0.021]		
Potential savings within chosen brand	4.173	23.617	2.949	61.123	-15.415
	[4.810]	[9.438]**	[12.977]	[34.181]*	[44.732]
Number of plan choices	16,112	3,299	1,755	577	327
Number of enrollees					
Share choosing dominated plans	0.18	0.17	0.18	1.00	1.00
Average potential savings within brand ($)	98	89	101	329	336

Note: The table reports coefficients on an indicator for whether the individual answered the MCBS knowledge question incorrectly. Robust standard errors are clustered by enrollee. *,**, and *** indicate that the p-value is less than 0.1, 0.05, and 0.01 respectively.

The first row reports results from regressing an indicator for whether the individual was enrolled in a dominated plan on an indicator for whether the decision maker gave the wrong answer to the MCBS knowledge question. Control variables include the explanatory variables in Table 4 and indicators for year and CMS region. Hence, the coefficient measures the partial effect of knowledge on decision making outcomes. The first column shows that there is no relationship between knowledge and the probability of choosing a dominated plan in the full sample. Column 2 conditions on active choices made without help. The point estimate is insignificant (p=.172) but implies a 2.1% increase in the probability of choosing a dominated plan. In contrast, the point estimate is approximately zero for active choices made with help in column 3

The second row reports results after replacing the dependent variable with the amount of money (in dollars) that the individual could have saved by purchasing a cheaper plan offered by the same brand. In column 1 the point estimate is $4 but insignificant. In column 2 we see that answering the knowledge question incorrectly was associated with a statistically significant $24 increase in potential savings for people making active choices without help. This effect is equivalent to 27% of the potential savings within this group. In contrast, when we focus on active choices made with help in column 3 the point estimate is a statistically insignificant $2. The last two columns condition on dominated plan choices. For this group we see that those answering the knowledge question incorrectly who did not have help could have saved approximately $61 than those who answered the question correctly, whereas the effect for those getting help is negative and imprecisely estimated.

TABLE A3—LOGIT MODELS WITH ADDITIONAL DEMOGRAPHIC INTERACTIONS

	All Choices		Non-Suspect choices		Suspect choices	
expected cost	-0.253	[0.019]***	-0.427	[0.034]***	-0.142	[0.021]***
variance	-0.555	[0.258]**	-2.090	[0.724]***	0.793	[0.514]
quality (CMS index)	0.256	[0.080]***	0.151	[0.112]	0.272	[0.120]**
within-brand switch	-3.342	[0.105]***	-3.242	[0.149]***	-3.459	[0.151]***
between-brand switch	-5.320	[0.086]***	-5.232	[0.120]***	-5.572	[0.123]***
cost x 1{ income > $25k }	-0.014	[0.019]	0.022	[0.033]	-0.030	[0.021]
cost x 1{ bottom tercile of claims }	-0.134	[0.028]***	-0.167	[0.041]***	-0.061	[0.036]*
cost x 1{ top tercile of claims }	0.083	[0.019]***	0.135	[0.035]***	0.039	[0.021]*
cost x 1{ help }	0.010	[0.019]	0.042	[0.034]	-0.023	[0.021]
cost x 1{ sought CMS info }	-0.054	[0.020]***	-0.063	[0.034]*	0.009	[0.022]
variance x 1{ college graduate }	-0.581	[0.318]*	-0.946	[0.558]*	0.297	[0.476]
variance x standardized age	-0.020	[0.116]	-0.280	[0.254]	-0.003	[0.215]
variance x 1{ female }	0.424	[0.284]	0.625	[0.559]	0.100	[0.506]
variance x 1{ help }	-0.009	[0.280]	-0.745	[0.502]	0.322	[0.398]
variance x 1{ sought CMS info }	0.190	[0.277]	0.777	[0.464]*	0.577	[0.473]
quality x 1{ income > $25k }	0.236	[0.084]***	0.276	[0.114]**	0.089	[0.140]
quality x 1{ help }	-0.043	[0.086]	-0.046	[0.116]	0.010	[0.138]
quality x 1{ sought CMS info }	0.244	[0.089]***	0.202	[0.118]*	0.358	[0.153]**
switch within brand x standardized age	-0.180	[0.069]***	-0.202	[0.095]**	-0.101	[0.102]
switch within brand x 1{ income > $25k }	-0.347	[0.124]***	-0.278	[0.169]	-0.429	[0.186]**
switch within brand x 1{ help }	0.280	[0.122]**	0.137	[0.173]	0.437	[0.182]**
switch within brand x 1{ sought CMS info }	0.086	[0.131]	0.254	[0.167]	-0.235	[0.219]
switch within brand x 1{ nonwhite }	-0.714	[0.283]**	-0.480	[0.342]	-0.851	[0.445]*
switch brand x standardized age	-0.095	[0.049]*	-0.093	[0.071]	-0.039	[0.070]
switch brand x 1{ income > $25k }	-0.303	[0.095]***	-0.298	[0.134]**	-0.341	[0.141]**
switch brand x 1{ help }	0.264	[0.094]***	0.293	[0.128]**	0.266	[0.140]*
switch brand x 1{ sought CMS info }	0.310	[0.094]***	0.283	[0.129]**	0.276	[0.147]*
switch brand x 1{ nonwhite }	-0.647	[0.211]***	-0.676	[0.339]**	-0.446	[0.285]
pseudo R^2	0.67		0.66		0.71	
number of enrollment decisions	9,831		5,465		4,366	
number of enrollees	3,511		2,166		1,675	

Note: The table reports parameter estimates from logit models estimated from data on all choices; from non-suspect choices only; and from suspect choices only. All models include indicators for insurers. Robust standard errors are clustered by enrollee. *,**, and *** indicate that the p-value is less than 0.1, 0.05, and 0.01 respectively.

TABLE A4—RISK PREMIUMS FOR 50-50 BETS FOR NON-SUSPECT CHOICES

Risk	Size of Bet
0.04	100
0.35	1,000
0.58	2,000
0.71	3,000
0.78	4,000
0.82	5,000
0.85	6,000
0.87	7,000
0.89	8,000
0.90	9,000
0.91	10,000

To assess the estimates from the logit model for non-suspect choices, we compare its implied risk premiums in a manner comparable with prior literature. Specifically, deriving the risk premium from the logit model as a 1st order approximation to a CARA model yields the following expression for the risk aversion coefficient:

$$\rho_{it} = \frac{-2\beta_{it}/1{,}000{,}000}{\alpha_{it}/100}, \text{ where } U_{ijt} = \alpha_{it}\acute{c}_{ijt} + \beta_{it}\acute{\sigma}^2_{ijt} + \gamma_{it}\acute{q}_{ijt} + \eta_{it}\Delta\acute{B}_{ijt} + \delta_{it}\Delta\acute{P}_{ijt} + \epsilon_{ij1}.$$

The estimates in Table 5 for the reference individual in the non-suspect group yields $\rho = .000773$. Table A5 translates this into a risk premium for various 50-50 bets. These results are broadly consistent with the range of prior results, e.g. as reported in Table 5 of Cohen and Einav (2007). Cohen and Einav find the mean consumer would be indifferent between a 50-50 bet of winning $100 and losing $76.5, whereas the median consumer is virtually risk neutral. In contrast, our results imply the mean non-suspect consumer is indifferent between a 50-50 bet of winning $100 and losing $96.3 although Cohen and Einav argue that preferences likely differ between their automobile insurance context other contexts like drug insurance. In the health insurance context, Handel (2013) finds that the median individual is indifferent between a bet of winning $100 and losing $94.6. In the model preferred by Handel and Kolstad (2015), the mean consumer is indifferent between a bet of winning $1,000 and losing $913. This controls for friction and inertia. In comparison, our results imply indifference between winning $1,000 and losing $739.

TABLE A5

CHARACTERISTICS OF PEOPLE WHO ALWAYS, SOMETIMES, OR NEVER MAKE SUSPECT CHOICES

	Always suspect	Sometimes suspect	Never suspect
number of enrollees	3,749	1,376	4,706
Medicare Current Beneficiary Survey			
high school graduate (%)	74	79	81
college graduate (%)	17	22	26
income>$25k (%)	48	50	59
currently working (%)	11	9	14
married (%)	49	54	60
has living children (%)	93	91	93
uses the internet (%)	27	34	41
searched for CMS info: internet (%)	21	30	32
searched for CMS info: 1-800-Medicare (%)	11	16	15
makes own health insurance decisions (%)	59	60	65
gets help making insurance decisions (%)	28	29	25
insurance decisions made by proxy (%)	13	11	10
CMS Administrative Data			
mean age	79	78	77
female (%)	66	69	59
white (%)	91	95	94
dementia including Alzheimer's (%)	14	10	8
depression (%)	12	12	9
mean number of drug claims	39	37	32
mean number of available plans	52	52	52
mean number of available brands	23	23	22
has a default plan (%)	85	80	79
switches out of the default plan (%)	9	35	11
active enrollment decisions (%)	22	53	31
mean premium ($)	427	398	447
mean out-of-pocket costs ($)	1,025	989	800
mean potential savings, ex post ($)	362	332	284

TABLE A6

LOGIT ESTIMATES FOR PEOPLE WHO ALWAYS, SOMETIMES, AND NEVER MAKE SUSPECT CHOICES

			Sometimes suspect					
	Always suspect		suspect choice		non-suspect choice		Never suspect	
expected cost	-0.176	[0.021]***	-0.123	[0.035]***	-0.411	[0.063]***	-0.405	[0.032]***
variance	0.807	[0.268]***	2.057	[0.423]***	-1.424	[0.492]***	-1.488	[0.354]***
quality (CMS index)	0.229	[0.127]*	0.567	[0.229]**	0.063	[0.208]	0.196	[0.124]
within-brand switch	-3.762	[0.191]***	-2.376	[0.275]***	-1.458	[0.242]***	-3.829	[0.188]***
between-brand switch	-6.054	[0.149]***	-4.105	[0.258]***	-3.738	[0.208]***	-5.597	[0.150]***
cost x 1{ bottom tercile of claims }	-0.074	[0.043]*	-0.029	[0.058]	-0.176	[0.118]	-0.172	[0.042]***
cost x 1{ top tercile of claims }	0.042	[0.024]*	0.040	[0.044]	-0.031	[0.079]	0.166	[0.039]***
cost x 1{ sought CMS info }	-0.011	[0.025]	0.020	[0.040]	0.008	[0.079]	-0.067	[0.036]*
variance x 1{ college graduate }	-0.148	[0.549]	1.535	[0.933]*	-1.758	[1.952]	-0.873	[0.714]
quality x 1{ income > $25k }	0.055	[0.161]	-0.058	[0.312]	0.053	[0.245]	0.269	[0.132]**
quality x 1{ sought CMS info }	0.418	[0.182]**	0.259	[0.289]	0.477	[0.261]*	0.071	[0.136]
switch within brand x standardized age	0.026	[0.117]	-0.409	[0.217]*	-0.228	[0.137]*	-0.287	[0.134]**
switch within brand x 1{ income > $25k }	-0.363	[0.221]	-0.463	[0.346]	-0.171	[0.290]	-0.292	[0.223]
switch within brand x 1{ help }	0.644	[0.213]***	-0.089	[0.382]	-0.271	[0.304]	0.235	[0.228]
switch within brand x 1{ sought CMS info }	-0.623	[0.288]**	0.148	[0.397]	-0.092	[0.287]	0.422	[0.220]*
switch within brand x 1{ nonwhite }	-0.420	[0.452]	-18.521	[0.468]***	0.091	[0.477]	-0.602	[0.467]
switch brand x standardized age	0.024	[0.085]	-0.064	[0.152]	-0.027	[0.121]	-0.227	[0.088]***
switch brand x 1{ income > $25k }	-0.235	[0.177]	-0.455	[0.280]	-0.539	[0.265]**	-0.281	[0.165]*
switch brand x 1{ help }	0.536	[0.165]***	-0.168	[0.313]	0.241	[0.257]	0.213	[0.164]
switch brand x 1{ sought CMS info }	0.117	[0.191]	0.013	[0.270]	-0.135	[0.254]	0.415	[0.158]***
switch brand x 1{ nonwhite }	-0.430	[0.327]	-0.437	[0.720]	-1.152	[0.905]	-0.510	[0.385]
pseudo R^2	0.75		0.52		0.47		0.71	
number of enrollment decisions	3,749		617		759		4,706	

Note: The table reports parameter estimates from logit models estimated from data on all choices; from non-suspect choices only; and from suspect choices only. All models include indicators for insurers. Robust standard errors are clustered by enrollee. *,**, and *** indicate that the p-value is less than 0.1, 0.05, and 0.01 respectively.

TABLE A7—VALIDATION OF LOGIT MODELS STRATIFIED BY SUSPECT VS NON-SUSPECT AGAINST ANALOG POOLED MODEL

	In-sample fit (2008)						Out-of-sample fit (2009)								Weighted absolute errors			
	suspect			non-suspect			suspect				non-suspect				in-sample		out-of-sample	
	data	|model error|		data	|model error|		data	|model error|			data	|model error|			|model error|		|model error|	
		s=ns	s		s=ns	ns		s=ns	s	ns		s=ns	s	ns	s=ns	s≠ns	s=ns	s≠ns
Percent of consumers choosing:																		
gap coverage	12	1	1	12	2	2	10	0	0	1	11	1	0	1	2	2	0	1
dominated plan	29	7	5	8	8	6	26	7	5	9	7	6	7	5	8	6	7	5
min cost plan within brand	51	7	4	70	9	11	50	7	2	5	67	5	10	7	8	8	6	5
Mean consumer expenditures ($)																		
premium + OOP	1,430	16	0	1,279	15	0	1,668	20	4	37	1,369	19	34	6	17	0	21	6
overspending on dominated plans	62	24	18	17	8	9	49	20	16	25	13	4	2	5	16	14	12	11
Percent of consumer switching plans	16	3	0	21	3	0	13	5	2	7	23	6	9	4	3	0	6	3

Table A7 reports results from a logit model validation exercise. The purpose is to determine whether the models estimated separately by suspect and non-suspect choices outperform the pooled model, and whether the suspect model better predicts suspect choices than the non-suspect model does and vice versa. For this exercise the estimation sample is 2008 while the prediction sample is 2009. We chose these two years because they incorporate the largest year-to-year change in the choice set in our data—a central aspect to out-of-sample validation methods (Keane and Wolpin 2007). In particular, the number of plans available fell by 10%, although three new brands entered the market, precluding our use of brand indicators in the models. The results show that both in-sample and out-of-sample predictions are closer to the data along a number of policy-relevant outcomes when we base the predictions on separate models for the given type of choice. Blue shading is used to indicate the moments where our preferred model that distinguishes between suspect and non-suspect choices outperforms the pooled model. Red shading indicates moments where the pooled model performs better.

TABLE A8—ADDITIONAL ROBUSTNESS CHECKS ON OUR MAIN RESULTS

	Menu Restriction		Decision Support		Default Assignment	
	most effective	least effective	most effective	least effective	most effective	least effective
	A. Baseline results					
Δ insurer revenue / enrollee ($)	6	36	-69	-50	-254	-92
Δ govt. spending / enrollee ($)	4	17	-54	-41	-144	-54
Δ expected welfare / enrollee ($)	12	-108	102	29	186	49
% enrollees with expected welfare gain	25	2	76	67	90	82
	B. Suspect choices expanded to include potential savings > 33%					
Δ insurer revenue / enrollee ($)	-52	30	-134	-43	-320	-87
Δ govt. spending / enrollee ($)	-37	14	-88	-28	-192	-53
Δ expected welfare / enrollee ($)	69	-102	169	70	267	60
% enrollees with expected welfare gain	36	3	91	74	94	85
	C. Including choices from 2006					
Δ insurer revenue / enrollee ($)	4	43	-81	-49	-259	-83
Δ govt. spending / enrollee ($)	4	22	-60	-39	-147	-49
Δ expected welfare / enrollee ($)	7	-119	115	36	188	48
% enrollees with expected welfare gain	25	2	77	66	90	83

Note: The table summarizes the sensitivity of our main results to alternative specification of the model. Panel A repeats our baseline results for convenience. The numbers match those in tables 6, 7, and 8. Panels B through F show comparable results after altering the baseline specification in various ways. See the text for details.

Panels A repeat our baseline results. Panel B uses a more inclusive definition based on the union of dominated plan choices, the knowledge question, and being able to reduce expenditures by more than 33%. Panel C includes enrollment decisions from the inaugural year of Medicare Part D.