Inattention and Switching Costs as Sources of Inertia in Medicare Part D

Florian Heiss (University of Düsseldorf)
Daniel McFadden (University of California, Berkeley)
Joachim Winter (LMU Munich)
Amelie Wuppermann (University of Halle-Wittenberg)
Bo Zhou (University of Southern California, Los Angeles)

Discussion Paper No. 221

December 18, 2019

Collaborative Research Center Transregio 190 | www.rationality-and-competition.de
Ludwig-Maximilians-Universität München | Humboldt-Universität zu Berlin
Spokesperson: Prof. Dr. Klaus M. Schmidt, University of Munich, 80539 Munich, Germany
+49 (89) 2180 3405 | info@rationality-and-competition.de

Inattention and switching costs as sources of inertia in Medicare Part D

Florian Heiss
University of Dusseldorf

Daniel McFadden
University of California, Berkeley
University of Southern California, Los Angeles

Joachim Winter
University of Munich

Amelie Wuppermann
University of Halle-Wittenberg

Bo Zhou
University of Southern California, Los Angeles

October 7, 2019

Abstract: Consumers' health plan choices are highly persistent even though optimal plans change over time. This paper separates two sources of inertia, inattention to plan choice and switching costs. We develop a panel data model with separate attention and choice stages, linked by heterogeneity in acuity, i.e., the ability and willingness to make diligent choices. Using data from Medicare Part D, we find that inattention is an important source of inertia but switching costs also play a role, particularly for low-acuity individuals. Separating the two stages and allowing for heterogeneity is crucial for counterfactual simulations of interventions that reduce inertia.

Keywords: Medicare; prescription drugs; health insurance demand; dynamic discrete choice.

JEL classification: I13; D12; J14; C25.

Acknowledgments: An earlier version of this paper was circulated under the title, "Plan switching and inertia in Medicare Part D: Evidence from administrative data." This research was funded by the Behavioral and Social Research program of the National Institute on Aging (grants P01AG05842-18, P01AG033559, R56AG026622-01A1, and RC4AG039036), with additional support from the E. Morris Cox Fund at the University of California, Berkeley, and from Deutsche Forschungsgemeinschaft through CRC TRR 190 (project 280092119). We gratefully acknowledge comments by, and discussions with, Abi Adams, Eric French, Dana Goldman, Ben Handel, Kathleen McGarry, Nicole Maestas, Filip Matějka, Takeshi Murooka, Maria Polyakova, and Justin Sydnor, participants at the 2014 AEA meetings in Philadelphia, the 2014 ASHEcon conference in Los Angeles, the 2014 CESifo Conference on Behavioral Economics in Munich, the 2014 CEAR/MRIC Behavioral Insurance Workshop in Munich, the 2014 German Health Econometrics Workshop in Wuppertal, the 2015 NBER Summer Institute, the 2015 Workshop on Policyholder Behavior at ETH Zurich, the 2016 EuHEA meetings in Hamburg, as well as seminar audiences at cemmap/UCL, DIW Berlin, George Mason University, Harvard University, Helmut Schmidt University Hamburg, Northwestern University, and the universities of Basel, Copenhagen, Essex, Leuven, Lugano, Lund, Oxford, Regensburg, St. Gallen, Tübingen, Wuppertal, Würzburg, and Zurich.

"A wealth of information creates a poverty of attention."
Herbert Simon

1 Introduction

The recent trend towards giving consumers more choice about their health plans has invited research on how good they are at making these decisions. Data come from private markets (such as employer-sponsored health coverage) as well as from health insurance programs that are offered or subsidized by governments. The Medicare Part D market for prescription drug insurance is an important example. A rapidly expanding literature analyzes Part D enrollment and plan choices, using both survey and administrative data. Generally, initial plan choices in Medicare Part D are difficult to reconcile with standard normative models of decision-making; see for instance Abaluck and Gruber (2011) and Heiss *et al.* (2013).[1] Evidence from other health insurance markets also finds that people make poor decision, such as choosing expensive, demonstratively dominated plans; see Bhargava *et al.* (2017).

Initial plan choice is only one aspect of consumer choice in Medicare Part D. Consumers stay in Medicare Part D for many years, and at the end of each year have an open enrollment period when they can switch plans.[2] Current enrollees are ordinarily reenrolled in their current plan by default if they take no action, so that attention is required only if the enrollee wants to consider a switch.[3] Enrollees experience changes over time in their health and prescriptions drug needs. Moreover, the menu of plans and features offered also change from year to year. In addition, recent reforms implement changes in the copayment and coverage structure of Medicare Part D plans, such as the gradual abolishment of the infamous coverage gap. As a result, enrollees who consider switching plans face a choice that is as complex as their initial plan choice.[4]

[1] We review the literature on plan choices in Medicare Part D in Section 2.2.

[2] Each year during the Medicare Part D annual enrollment period that runs between October 15 and December 7, individuals on Medicare have the opportunity to enroll in Part D, or if they are enrolled, to switch plans. Switching plans does not involve any fees, and as at initial enrollment, plans have to accept all individuals. Important institutional features of the Medicare Part D market are described in Bach and McClellan (2005), Duggan *et al.* (2008) and Heiss *et al.* (2010), among others. We provide an overview in Section 2.1.

[3] An infrequent exception that forces attention is the termination of an enrollee's current plan without a "cross-walk" to a new or consolidated plan that is the designated default for re-enrollment.

[4] Ketcham *et al.* (2012) conclude that Part D decisions improve over time with enrollee experience. Our results suggest that most of this learning comes from the first enrollment experience, and selection

Inertia may be attributed to the presence of defaults, lack of events that trigger consumers' attention, low expecations of improvement, or high costs of considering and carrying out switches.[5] If inertia leads to suboptimal switching choices in the Part D market, there are not only direct welfare losses to enrollees, but also overall effects on market efficiency and welfare due to strategic behavior of insurers.[6]

In this paper, our objectives are to characterize the sources of inertia, to study the events that trigger switching in the presence of inertia, and to simulate and compare the effects of interventions aimed at reducing inertia. We specify a model of consumer choice with two crucial features. First, the model comprises separate attention and plan-selection stages, to separate inattention to plan choice from switching costs as reasons for inertia. These stages are distinct and econometrically identified by fundamental exclusion restrictions: Attention is triggered by events that have occurred before the current enrollment period; e.g., discontinuation of previous plan, incidence of a new health condition, or a shock from an announced change in current plan premium or formulary. Plan choices are driven by future costs and expected benefits, information that is not acquired by the inattentive consumer. Then, past attention triggers should be irrelevant to and excluded from plan choice made by attentive consumers, while future features and costs of plans other than a previous plan, which the consumer can learn only by paying attention, are logically excluded from the attention stage. As Abaluck and Gruber (2016a) note, the distinction between attention and plan choice stages has important normative implications for the design of defaults and other policies that encourage thoughtful health decisions. Second, our choice model allows for unobserved heterogeneity in the ability and willingness to make diligent choices. In the following, we call this latent factor "acuity". It affects all

on decision-making skills is a better explanation than cumulative market experience for subsequent outcomes achieved by plan switchers.

[5] When attention is a scarce resource, bounded rationality or rational inattention may lead the consumer to focus on a limited number of decisions where there is potentially high payoff to diligence, leaving many choices to be made by "fast thinking" heuristics or reliance on defaults; see Simon (1955) and Kahneman (2011) for an early and a more recent perspectives on the broader issue, Gabaix (2017) for a comprehensive review of the economics literature on inattention, and Matějka and McKay (2015), Abaluck and Adams (2017), and Gibbard (2017) for the implications of inattention for discrete choice modelling. Attention involves direct costs of collecting and processing information on alternatives, and opportunity costs of inattention to other choice tasks.

[6] The mechanisms that suppliers commonly employ to exploit inertia, such as raising prices or reducing benefits to returning enrollees, are largely but not entirely eliminated by regulations on offered plans in the Part D market. More importantly, when inertia allows inefficient plans to survive in the market, they increase the average benefit paid, eventually increasing the burden on taxpayers from the market?s substantial government subsidy. For further discussion, see Handel (2013), Ericson (2014), Handel and Kolstad (2015), Polyakova (2016), Ho *et al.* (2017), Decarolis *et al.* (2019).

stages of the choice model: acuity increases the propensity to be attentive, influences apparent switching cost, and reduces noise in plan choice.

Estimation and simulation of our two-stage model of plan switching use administrative data on Medicare Part D plan choice and prescription drug use for 2007 through 2010.[7] A preliminary, descriptive analysis confirms that annual plan switching rates in the Medicare Part D market are low, a finding in line with earlier studies of plan switching in this and many other markets, see Section 2.2. We characterize switching behavior descriptively in terms of overspending, defined as the difference between the consumer's total costs[8] induced by the plan the consumer has chosen and that of the least-cost alternative (for given prescription drug use).[9] We find that overspending is much lower among individuals who switch plans than for stayers. Overspending is furthermore lower for those who switch when they have an available default than for those who are forced to make a choice (because they do not have a default plan as they enter Part D for the first time or their old plan is not available any more). This finding suggests that those individuals who make an active switching decision are selected, which is reflected in our structural model. Our descriptive analysis further suggests that overspending is not fully explained by unobserved plan characteristics that might lead consumers to optimally choose plans associated with higher overspending.[10]

Individuals making an active plan choice may stay with their default plan because it is optimal, or because switching costs are too high. In discrete choice models, such switching costs are typically captured by including a dummy variable for the default plan – which typically is just the old plan – with a coefficient interpreted as a monetary estimate of the disutility of switching; see Farrell and Klemperer (2007) for a review of the literature on switching costs in industrial organization and Shcherbakov (2016) for a more recent example. Studies of switching behavior in Medicare Part D and other health care markets that use such models include Miller and Yeo (2012), Nosal (2012), Handel (2013), Ericson

[7] We restrict our analysis to individuals enrolled in a stand-alone prescription drug plan (i. e., not in managed care) and who are not eligible for a Low-Income Subsidy (LIS). Further details are given in Section 3.

[8] The total costs that the consumer has to bear include premiums, deductibles, and copayments that she has to pay for prescription drugs.

[9] Overspending has been defined and studied in several earlier studies of Medicare Part D plan choice, including Abaluck and Gruber (2011), Heiss *et al.* (2013), and Abaluck and Gruber (2016a). In Section 3 we discuss how the overspending measure we use here is constructed.

[10] This observation is related to a more general discussion about the use of parameterized models of choice behavior to study the rationality of consumers in Medicare Part D and other markets, see Ketcham *et al.* (2016) and Abaluck and Gruber (2016b). We return to this issue in the concluding Section 7.

(2014), Handel and Kolstad (2015), Abaluck and Gruber (2016a), and Polyakova (2016). These studies also provide discussions of the sources of switching costs in health insurance markets, the most important one being transaction or "hassle" costs. We review these studies in more detail in Section 2.2.

The approach of including a dummy variable for the default plan in a discrete choice model implies that each individual compares the available plans in each year and deliberately makes a choice – which seems unrealistic when the availability of a default or high consideration costs invite inattention. Moreover, our simulations show that in such a model, switching costs have to be unreasonably large to produce typically low switching rates. Polyakova (2016) points out that including heterogeneity in these one-stage models tends to bring down estimated switching costs. The two-stage model we propose achieves this as well, with inattention to plan choice as a second source of inertia.

A few recent models of switching behavior in Medicare Part D and other markets allow for both inattention and switching costs. While they share certain features of our model, none comprises all of them.[11] Kiss (2014) studies liability insurance switching in Hungary. His specification of the attention stage is simpler than ours as he can rely on a natural experiment that creates variation in the attention probability. Hortaçsu *et al.* (2017) develop a model with attention and switching stages and use it to analyze inertia in the Texas residential electricity market. Ho *et al.* (2017) is substantively related to the present paper as they analyze inattention in Medicare Part D and discuss attention triggers. Their model focuses on the implications of inertia for firm behavior. None of the above three models allow unobserved heterogeneity that enters both stages. Finally, in recent work Abaluck and Adams (2017) provide a theoretical foundation for models of plan choice with inattention.[12] Their empirical analysis uses data from Medicare Part D, and while their empirical approach differs from ours, their key results on the importance of inattention are consistent with ours.

In order to characterize the individual and joint contributions of inattention, switching costs, and acuity to the observed persistence of plan choices, we estimate our model in five stages (I to V) by successively removing restrictions. The full model V provides

---

[11] Our study and the papers cited in this paragraph introduce inattention in two-stage models of consumer choice. Alternative but related motivations for introducing a first stage focus on decision costs. In a study of retirement investment, Luco (2018) distinguishes between decision costs and enrollment costs. The first-stage decision in Honka (2014) determines how many markets a consumers searches and thus endogenizes the size of her consideration set.

[12] There is a large and growing literature on rational inattention, see Gabaix (2017) for a review. Implications for empirical models of discrete choice are studied by Matějka and McKay (2015), Abaluck and Adams (2017), and Gibbard (2017), among others.

the best fit of the data, and the restricted models are statistically rejected. In the full model, we conclude that consumers are more likely to pay attention to plan choice if overspending in the last year is more salient and if their old plan gets worse, for instance due to premium increases. However, consumers are not more likely to pay attention when they experience the onset of a chronic health condition that involves costly prescription drugs, even though such an onset could make it worthwhile to search for a plan with a more suitable formulary.

Our model allows us to compare two classes of potential interventions aimed at reducing inertia. We simulate choices and overspending in counterfactual scenarios that force attention and remove switching costs, respectively. Our overall conclusion is that policies that increase attention, say by eliminating defaults or making them less automatic, or by providing individualized plan comparisons to reduce consideration cost, are at least as effective as policies designed to reduce switching costs.

The remainder of this paper is structured as follows. We begin by describing the institutional arrangements of Medicare Part D that are relevant for our analysis and the related literature on plan choice and plan switching in this and other markets in Section 2. Section 3 describes the data sources, the construction of our analytic dataset, and the definitions of the model variables. It also reports some descriptive empirical facts about plan switching in Medicare Part D which motivate the set-up of our two-stage discrete choice model. This model is then introduced in Section 4. The estimation results are presented and discussed in Section 5. In Section 6, we report the results of our simulations of counterfactual scenarios aimed at reducing inertia in the Medicare Part D market. In Section 7, we summarize our findings and draw policy conclusions.

2 Background

2.1 Institutional framework of Medicare Part D

Medicare Part D, introduced in 2006, provides the Medicare-eligible population with universal access to standardized, heavily subsidized prescription drug coverage through government-approved plans sponsored by private insurance companies and health maintenance organizations (HMOs). In addition to providing access to affordable drug coverage to all Medicare beneficiaries (in particular to the chronically ill), a second policy goal was to create a "competitive, transparent marketplace offering a wide array of benefits" (Bach and McClellan (2005), p. 2733). The institutional design of Medicare Part D exemplifies the current trend toward "consumer-directed health care" (Goodman (2006)) as it relies on consumer behavior and competition among insurers to attain satisfactory market outcomes with limited government regulation. In the case of Medicare Part D, and arguably also in other similar programs, giving consumers more choice also means confronting them with difficult decisions. In the following, we describe the features of the program that are relevant for our present analysis.

Medicare beneficiaries can enroll in Part D when they become eligible for Medicare. If they enroll, they can choose between about 50 plans (the exact number varies across regions) that vary in monthly premium, the formulary benefit design, as well as the benefit structure. Once enrolled, beneficiaries can switch to a new plan annually during the open enrollment period at the end of each year. If individuals do not actively decide to switch to a different plan they are automatically re-enrolled in their old plan for the new year. The menu of available plans is the same for first-time enrollees and switchers. Importantly, plans might be taken off the market so that they are not available to enrollees for the next year. In most cases, the plan sponsor transfers enrollees in such plans to different ones; these transfers are documented in "cross-walk" tables. Otherwise, enrollees will not have a default plan.[13] As we discuss later, we exploit this fact for identification.

Each firm that wants to enter this market has to offer at least a plan that is actuarially equivalent to a standard plan whose features are tightly specified. The standard plan is defined in terms of the benefit structure, including specific values for the plan's deductible and two coverage thresholds that depend on the beneficiary's annual pharmacy bill or the beneficiary's out-of-pocket (OOP) cost for drugs on the plan's formulary. In 2006, the first year of the Part D market, the standard Medicare Part D plan had the following benefit structure:[14]

- The beneficiary has an annual deductible of $250.
- The beneficiary pays 25 percent of drug costs above $250 and up to $2,250 (the "initial coverage limit" or ICL). For a beneficiary whose annual pharmacy bill has reached $2,250, the OOP cost is then $750.
- The beneficiary pays 100 percent of drug costs above $2,250 and up to OOP cost of $3,600; this is referred to as the coverage gap or donut hole. The threshold of $3,600 is attained at a drug bill of $5,100.

[13] See Hoadley *et al.* (2013) for further details.

[14] In subsequent years, these numbers were adjusted for inflation, see for example Heiss *et al.* (2013).

- The beneficiary pays 5 percent of drug costs above the drug cost threshold of $5,100; this is referred to as catastrophic coverage.

The Medicare Part D plans sponsored by private insurance firms may differ from the standard plan, provided that their benefits for any drug costs are, on average, at least as high as those of the standard plan. Enhancements may in particular include coverage for the $250 deductible and for the gap in the standard plan.

This design implies that Medicare Part D stand-alone plans are characterized by the following variables: premium, deductible (if any), ICL for those plans that are not classified as "standard benefit", and the formulary benefit design which specifies drug tiers and copayments. Together with the specific prescription drugs an enrollee uses over the course of the year, these characteristics determine her OOP cost. The sum of these OOP cost and the premiums paid is the total cost to the consumer which for simplicity we refer to as "total costs" in the following.

There are two institutional features that provide beneficiaries with easy access to information on the plans that are available to them, and their cost implications: First, the Centers for Medicare & Medicaid Services (CMS) provide an online deicsion tool, Plan Finder, that gives the premium and out-of-pocket costs in all available plans, given the individual's current or anticipated list of prescriptions.[15] Second, before the beginning of the open enrollment period, each plan has to send a standardized "Plan Annual Notice of Change" (ANOC) to its enrollees, based on a template provided by CMS. This notice contains a table with cost-relevant plan features for the current and the following year (premium, deductible, copayments) as well as detailed information on all changes that will become effective in January of the following year.[16]

2.2 Existing literature on plan choice and switching in Medicare Part D

Since its introduction in 2006, a large number of studies have examined the quality of consumer decision-making in the Medicare Part D market. Early studies of enrollment decisions and initial plan choices, such as Heiss *et al.* (2010) and Kling *et al.* (2012),

[15] The CMS Plan Finder can be accessed at http://www.medicare.gov/find-a-plan/ (last visited: September 9, 2016). A limitation of Plan Finder is that while it suggests generic substitutes for an enrollee's branded drugs, it does not identify and provide convenient price information on branded therapeutic substitutes that are less expensive in alternative plans, or calculate risks associated with new conditions or complications, see Heiss *et al.* (2013).

[16] A recent example of an ANOC can be found at `https://www.yourmedicaresolutions.com/sites/default/files/2015%20%20Standard%20ANOC.pdf` (last visited: September 9, 2016).

analyzed survey data and documented choices that seemed unlikely to be optimal, even though initial enrollment rates were high and overall the introduction of Medicare Part D was deemed to be successful; see Heiss *et al.* (2006), Goldman and Joyce (2008), Duggan *et al.* (2008). More recent research uses much more detailed administrative data on plan choices and prescription drug claims provided by CMS. A finding of several early studies in this literature is that for substantial fractions of those enrolled in Part D, initial plan choices imply overspending: For a given use of prescription drugs, plans with lower total costs than that of the chosen plan exist, see Abaluck and Gruber (2011), Zhou and Zhang (2012), and Heiss *et al.* (2013). For example, Heiss *et al.* document that in 2006–2009, less than a quarter of individuals were enrolled in plans that were, from an *ex ante* viewpoint, as good as the least-cost plan covering the same drugs. Their estimates indicate that consumers overspent about $300 per year, on average. Unobserved plan and taste heterogeneity may explain some of the overspending documented in this and other studies, but most likely not all of it.

Given that beneficiaries' drug needs evolve over time and plan sponsors adjust key features such as premiums and the formulary benefit design, and switching involves no out-of-pocket monetary cost, the observed switching rates around 10 percent found by Heiss *et al.* (2010) and Hoadley *et al.* (2013) are surprisingly low. A natural approach to assess the effect of inertia at the individual level is to check how overspending evolves over time. Ketcham *et al.* (2012) find that enrollees were more likely to switch plans if their potential gain from doing so was larger and that overall, large reductions in overspending were realized from 2006 to 2007. In contrast, using the full universe of claims data that became available subsequently, Abaluck and Gruber (2016a) find that forgone savings from choosing suboptimal plans have increased during the first four years of Medicare Part D. They argue that there has been little consumer learning over time in the Part D market, and that increasing choice inconsistencies are driven by changes on the supply side that are not offset by consumers because of inertia. Relatedly, Abaluck *et al.* (2018) show in a reduced-form analysis of Medicare Part D plan choices that small price elasticities, substantial myopia, and excess sensitivity to salient plan characteristics impact choices beyond their effect on prices. As mentioned in the introduction, attributing the low switching rate solely to switching costs gives implausibly high estimates of these costs. For example, Miller and Yeo (2012) estimate switching costs around $1,700. Estimates based on individual level data that allow for some heterogeneity tend to be smaller, but still substantial; see Polyakova (2016) and Ketcham *et al.* (2015).

Another dimension of consumer behavior in health insurance markets that has been studied using administrative data from Medicare Part D concerns enrollees' reaction to nonlinearities in benefit schedule, particularly the out-of-pocket cost of drugs in the intial coverage region, the gap, and the catastrophic region. Einav *et al.* (2015) show that beneficiaries react to these nonlinearities by substituting drugs within and across years.

3 Data construction and descriptive evidence on plan switching

In our analysis of plan switching and inertia in Medicare Part D, we use the claims records of a 20% representative sample of Medicare enrollees in the years 2007–2010. These claims data have also been used by other researchers working on Medicare Part D plan choice and switching, including Abaluck and Gruber (2011), Abaluck and Gruber (2016a), Heiss *et al.* (2013), Hoadley *et al.* (2013), and Polyakova (2016). As in most of these earlier studies, we restrict our analysis to the claims for beneficiaries aged at least 65, in stand-alone prescription drug plans (PDPs) who do not receive low-income subsidies and are not dual-eligible for Medicare and Medicaid.

The claims data allow identification of major attributes of available and chosen plans, such as premium, deductibles, and gap coverage. However, plan identities were encrypted, so that it was not possible in our construction to name the insurer and retrieve the plan's exact formulary. Therefore, we estimated the formularies of the encrypted plans from the tiers and copayments of all drugs used in each plan by its enrollees from our sample. The algorithm is detailed in Heiss *et al.* (2013); a similar approach has been used by other researchers including Abaluck and Gruber (2016a).

Throughout the paper, we use the following convention for dating variables: Decisions with respect to plans covering prescription drug use in year t are taken during the open enrollment period at the end of year $t-1$. The plan in which a consumer is enrolled in year $t-1$ is referred to as the "default plan", the "old plan" or the "year $t-1$ plan", and the "new plan" is the one covering year t.

3.1 Sample

To study switching between Part D plans among seniors, we form working samples of Part D enrollees who have unrestricted choice among all the plans available in their Medicare region, and have sufficient data on their health, drug use, OOP costs, and premiums to estimate our plan-choice models.

The full 20% samples comprise between 9.3 and 10 million individuals per year, for the four years 2007–2010. We restrict our analytic sample to those individuals who are U.S. residents, aged 65 years or older, enrolled in stand-alone, non-employer (non-EGWP) Part D prescription drug plans; are neither entitled to the Low-Income Subsidy (LIS) nor dual eligible for Medicaid; are continuously enrolled in Part D in two adjacent years; and do not switch plans during the year (outside of the open enrollment period).[17]

Applying these criteria results in sample sizes of around 1.2 million individuals for the years 2008–2010 (Table 1). As we analyze plan choice for year t as made at the end of year $t-1$, conditional on year $t-1$ information, we estimate our choice models for these three years. The criteria for selecting individuals for our analytic sample are essentially the same as those used by Heiss *et al.* (2013) and other studies of plan choice in Medicare Part D. In the final step, we randomly select 100,000 individuals since the estimation of our model is computationally intensive. The number of observations available for each year is smaller (Table 1) as the resulting panel is unbalanced. The total number of choices across all individuals and years is 234,455. All descriptive statistics and model estimates reported in the remainder of the paper refer to this estimation sample, or to sub-samples thereof.

3.2 Plan choice, prescription drug use, and individual characteristics

For our subsequent analysis, we construct various variables that measure individual characteristics, health status, prescription drug use, and plan choice. These data are augmented with data on the features of the available plans (which may vary every year) such as premium and whether they provide gap coverage.

In addition, we construct cost variables based on the individual's prescription drug use during year $t-1$, i. e., the year at the end of which decisions for year t are made. The list of all prescription drugs used combined with the formulary and benefit design of a Part D plan and the plan's premium gives the total costs of that plan, i. e., the sum of all

[17] For those who are not in a stand-alone plan, we do not have data on individual prescription drug claims. Eligibility for the LIS, in turn, changes the nature of the choice problem an individual faces considerably, as discussed for instance by Decarolis (2015). These groups are commonly excluded in recent studies of plan choice and switching behavior using claims data, as for instance in Ho *et al.* (2017), Abaluck and Gruber (2016a), and Polyakova (2016). Similar to these studies, we do not adjust for or estimate selection into (or out of) stand-alone PDP plans. There is an extensive and growing literature focusing on selection in the Medicare context, mainly focusing on advantageous selection into Medicare Advantage (for an overview see Newhouse and McGuire (2014)) and how insurers encourage this type of selection, e.g. through advertising (Aizawa and Kim (2018)) or formulary design (Lavetti and Simon (2018)).

copayments, the premiums, and any deductibles. We construct total costs for year t for the plan chosen in year $t-1$ and for any plan available in the next year.

We also construct variables that indicate whether according to her OOP costs, an individual (in her current year plan) ended up in the Part D catastrophic region, in the coverage gap, or below the coverage gap.

Finally, we construct indicators for the incidence of new health conditions, classified by whether they are costly in terms of prescription drugs or not. We classify conditions as costly if having the specific condition predicts high prescription drug costs in supplementary regressions.[18]

The claims data contain no demographic information on the individual other than age, sex, race, and geographic location. To close this gap we augment our data with regional information from the American Community Survey (ACS): the ZIP-code share of individuals aged 65 and above without a high school degree ("low education") and the share of the population aged 65 or above with income below the federal poverty level (FPL) ("low income") and above 5× the FPL ("high income").

Descriptive statistics for these variables are reported below in Section 5. In the remainder of the present section, we discuss the choice and cost variables in more detail.

3.3 Individuals with and without a default plan

In our subsequent analysis, we distinguish individuals with and those without a default plan: individuals whose choice set in t includes the plan that they were enrolled in in t–1– as well as those whose $t-1$ plan left the market but was assigned a new plan for t by the plan sponsor – have a default plan. If they do not take action, they will – by default – stay in their old (or the designated default) plan. On the contrary, individuals without a default are forced to pay attention to plan choice if they do not want to end up without Part D coverage.

Individuals with a default plan either switch to a new plan or stay in the default – typically the old one. Thus, every year, each individual falls into one of three groups: without default plan, switcher, and stayer. Table 2 displays the shares of these groups.

[18] Costly conditions include acute myocardial infarction, ischemic heart disease, stroke, breast and lung cancer, Alzheimer's disease and dementia, diabetes, depression and asthma. In comparison to these conditions, hip fracture, prostate cancer, endometrial cancer, colorectal cancer, chronic obstructive pulmonary disease, chronic kidney disease, osteoporosis, heart failure, rheumatoid arthritis/osteoarthritis, glaucoma, anemia, benign prostatic hyperplasia, hyperlipidemia, acquired hypothyroidism, hypertension, atrial fibrillation are relatively cheap and thus included as "cheap" conditions.

The majority of individuals with default ends up staying in their default plan. On average across all three years, only a little more than 10% of the individuals with default plan switch plans. On average, and across all years, about 3.4% of individuals do not have a default plan and are thus forced to make a plan choice. The majority of them do not have a default plan because they are new to Medicare Part D (97%). Only 3% have been enrolled in a $t-1$ plan that is no longer available in year t and was not mapped to a new plan by its sponsor.[19] Despite some differences in sample definitions, these numbers are in line with those reported in an extensive descriptive analysis of administrative data on Medicare Part D plan switching between 2006 and 2010 by Hoadley *et al.* (2013).

3.4 Total costs and overspending

To motivate our analysis of inertia in Medicare Part D plan choices and the effect of interventions to reduce it, we now present descriptive evidence on the quality of plan choices of individuals without default plan, switchers and stayers. In the literature on plan-choice quality in Medicare Part D, and in health insurance more generally, a measure of overspending is frequently used for this purpose, see Abaluck and Gruber (2011), Heiss *et al.* (2013), and Abaluck and Gruber (2016a). Conceptually, overspending is defined as the difference between the total costs (to the consumer) of the plan a consumer has chosen and the total costs of the least-cost alternative. This definition reflects the fact that relevant features of the plans such as premiums, deductibles, the plan formulary and copayments are summarized in the total costs.

In practice, measures of total costs and overspending can be constructed in different ways. A key distinction is that between an *ex ante* and an *ex post* perspective. Tables 3 and 4 display *ex post* and *ex ante* measures of total costs and of overspending for individuals without a default plan and those with a default overall and separately for stayers and switchers.

Ex post overspending for year t is calculated based on the prescription drugs that an individual has taken in that year. It thus reflects the costs that beneficiaries actually realized during year t. As Table 3 shows, individuals who stay in their default plan have the highest overspending, on average, compared to both switchers and individuals without a default plan. This is true in absolute values as well as when looking at overspending relative to total costs. These numbers might be taken to imply that actively making

[19] This low share is inline with Hoadley *et al.* (2013) who note that these situations have occurred relatively rarely. In most cases when plans left the market, they were mapped to a different plan as designated default by the plan sponsor.

decisions results in better outcomes. This is not what one would expect if individuals stayed in their default plans because of a preference for unobserved plan characteristics.

Individuals without a default plan do not choose quite as well as those who actively decide to switch to a different plan, which indicates substantial heterogeneity in the ability to choose plans. This implies that switchers are a selected group of individuals who appear to make better choices than individuals without a default plan who, in turn, are mostly newly enrolled.

In Table 4, we turn to *ex ante* overspending for stayers and switchers, which is based on prescription drugs taken in year $t-1$ (this measure cannot be computed for the newly enrolled individuals, as their prescriptions for $t-1$ are not observed). This measure is meant to capture beneficiaries' information on their prescription drug needs at the time when they make their plan choice. This is also the information beneficiaries are asked to provide to the Medicare Plan Finder. The table shows *ex ante* overspending in year t for the plan that individuals chose for year t as well as for the default plan. For the group of stayers, the two plans are obviously the same. For the switchers, overspending based on the default plan is the amount they would have overspent had they stayed in their default. The difference in overspending between the default and new plans thus captures how much switchers have saved by switching plans – $180 per year, on average. Compared to stayers, switchers' overspending in the default plan is slightly higher in all years except for 2010. This suggests that high *ex ante* overspending in the default plan may encourage individuals to switch to a different plan.

A particularly interesting group of individuals are those who have chosen the *ex ante* cost-minimizing plan. As discussed in Heiss *et al.* (2013), this is the choice implied by static optimization. It can easily be implemented, for instance by using a decision aid such as the Medicare Plan Finder discussed above. The lower panel of Table 4 displays the share of individuals with zero *ex ante* overspending, separately for the stayers and switchers. A much larger share of switchers than stayers have zero *ex ante* overspending. These numbers also support the notion that switchers make better choices than stayers.

In the structural model of plan choice presented in the remainder of this paper, we take an *ex post* perspective. We cannot use the *ex ante* measures because they are not available for the individuals without a default plan. *Ex post* measures of overspending might seem problematic as they condition on information on outcomes, realized during year t, that are not known to the individual at the end of year $t-1$. Using these measures is akin to assuming perfect foresight (Heiss *et al.* (2013)). *Ex ante* measures, in contrast, do not condition on year t outcomes but require assumption on how individuals form expectations at the end of year $t-1$ with respect to these variables. These assumption can also be problematic. We discussed the trade-off between using *ex ante* and *ex post* measures in detail in earlier work, see Heiss *et al.* (2013). To the extent that individuals have more information on the future development of their drug use than just the drugs used in $t-1$, *ex post* drug cost might even be a better measure of their expectations than *ex ante* measures. In our analysis of static Medicare Part D choices, we found that the key substantive findings do not depend on whether *ex ante* or *ex post* measures are used.

4 A two-stage model of plan-switching decisions

In this section, we introduce a comprehensive model of plan selection that is behaviorally more plausible and can explain more of the empirical findings presented in Section 3 than typical models used in the literature; Figure 1 provides a schematic overview. The model includes two features that have not been combined in previous papers.

First, our model comprises separate attention and plan-choice stages. Attention is triggered by events known to the consumer before the open enrollment period, such as onset of new health conditions, announcement of premium or formulary changes of the old plan. Plan choice requires attention, and is influenced by the future expected benefits and costs of alternative plans, as well as the cost of switching from the default plan. The attention stage excludes information on new plan features that are unavailable to the inattentive consumer, while the plan-choice stage excludes past triggering information that is irrelevant to future plan choice. These are the behavioral foundations of the exclusion restriction that identifies the two stages.

Second, our model allows for unobserved heterogeneity in the ability and willingness to incur the consideration costs of diligent decisions; we call this latent factor acuity. Persons may have low acuity because they have difficulty processing information and making decisions or because the opportunity costs of careful decision making in this situation are high.[20] Acuity affects both stages, allowing for the possibility that individuals who are attentive make more careful choices than the average. To identify switching costs in the second stage, we assume that enrollees with and without a default plan have the same distribution of preferences, conditioned on acuity.

Below, we first describe the structure of the choice problem in Section 4.1. The concept of acuity is introduced in Section 4.2. Then we discuss the two stages of the choice model in

[20] Distinguishing between these various factors would require a structural model for acuity and data on factors that influence opportunity cost; this is beyond the scope of this paper.

Sections 4.3 and 4.4. In Section 4.5, we describe how various sets of parameter restrictions give rise to simpler versions of this quite comprehensive choice model, some of which have been studied in the earlier literature on inertia. We refer to the full model as Model V; the restricted versions are labeled Model I–IV. The identification of the full Model V is discussed in Section 4.6. Since inattention is unobserved and acuity enters both stages simultaneously as a latent variable, all model parameters have to be estimated jointly. We use a maximum likelihood estimator with a numerical approximation of the analytically unavailable likelihood function; details are discussed in Section 4.7.

4.1 Structure of the choice problem

We model the plan choice for a set of consumers $i = 1, \ldots, N$ for years $t = 1, \ldots, T$.[21] Since the individuals start enrolling in Part D in different years, can die during the time our panel data covers, or can change the low-income subsidy status or other criteria for inclusion in our estimation sample discussed in Section 3.1, our panel data set is unbalanced and T differs by individual. For notational convenience, we do not explicitly index T by i in the remainder of this section. Individual i chooses among the plans available to her for year t. The number of plans and their features differ by year and Medicare region, so the choice set varies over time and across individuals. We index the plans in an individual's choice set by $j \in \mathcal{Y}_{it}$.

4.2 Acuity

Individuals differ in both their ability and their willingness to put effort into and incurring the consideration cost of making good plan-choice decisions. We introduce a latent variable, acuity, that captures these aspects. Acuity may comprise such diverse factors as intelligence, financial literacy, mental health, opportunity costs of time, or the availability of support from children, doctors, or other sources. It can affect choices in both stages: Individuals with high levels of acuity might be more likely to pay attention in the first stage of the model, i. e., they collect and compare information on the available plans. If they do pay attention, acuity might also affect the weight they put on plan features, including the objective cost consequences such as their premiums and implied OOP costs, and the error variance in the second stage (the multinomial choice model).

[21] The time index t refers to the year for which the plan is chosen. The actual choice is made in the open enrollment period at the end of the previous year.

Acuity q_{it} of individual i at time t is a latent variable that depends on observed individual characteristics $\boldsymbol{w}_{it}$. We include variables such as age, gender, depression, and ZIP-code level socio-economic characteristics. Acuity also includes a time-constant unobserved heterogeneity term c_i:

$$q_{it} = \boldsymbol{w}_{it}\boldsymbol{\alpha} + c_i \tag{1}$$

We normalize the elements of $\boldsymbol{w}_{it}$ such that q_{it} has a mean of zero. We further assume that c_i is normally distributed with mean zero; its variance, σ_c^2, will be estimated. Furthermore, we denote with q_{it}^* acuity, standardized to have a mean of zero and a variance of one.

4.3 Attention stage

Let a_{it} denote an unobserved indicator for whether individual i pays attention at time t ($a_{it} = 1$) or not ($a_{it} = 0$). Individuals who are inattentive always stay in their default plan. Some individuals are forced to pay attention and choose a plan, for example in the first year of enrollment or if a plan is discontinued without a cross-walk to a new default. While attention is unobserved, the dummy variable f_{it} which is equal to one if individual i is forced to make a choice as she has no default plan at time t is observed in the data. For individuals with a default plan ($f_{it} = 0$), we model the attention probability as a function of observable attention triggers, collected in the vector $\boldsymbol{x}_{it}$, and unobserved individual acuity q_{it} from equation (1).

Whether individual i pays attention in choice situation t is determined by a latent propensity, a_{it}^*. We use a linear specification

$$a_{it}^* = \boldsymbol{x}_{it}\boldsymbol{\zeta} + q_{it} + e_{it}. \tag{2}$$

The scale of latent acuity, q_{it}, is identified by including it additively in (2). Individuals with default plan pay attention if a_{it}^* is positive while individuals without a default always pay attention. Assuming i.i.d. logistic errors e_{it}, this implies a conditional attention probability of

$$p_{it}^a(c_i) \equiv \Pr(a_{it} = 1 | f_{it}, \boldsymbol{x}_{it}, \boldsymbol{w}_{it}; c_i) = \begin{cases} \frac{1}{1+\exp(-\boldsymbol{x}_{it}\boldsymbol{\zeta} - q_{it})} & \text{if } f_{it} = 0 \\ 1 & \text{if } f_{it} = 1 \end{cases} \tag{3}$$

4.4 Plan-choice stage

The plan-choice stage is specified as a generalized multinomial logit model with mixed (i. e., random) parameters and heteroskedasticity, building on McFadden (1974) and McFadden and Train (2000). The parameters of all plan characteristics may vary with the individual's acuity (which also appears in the attention equation, as described above).[22] Given an individual pays attention, she compares the plans j in her choice set $\mathcal{Y}_{it}$ according to their perceived utilities

$$u_{itj} = \mathbf{z}_{itj}\boldsymbol{\beta}_{it} + \gamma_{it} y_{it-1j} + v_{itj}. \tag{4}$$

We first discuss how the effects of the plan characteristics, collected in the vector $\mathbf{z}_{itj}$, are modeled. We allow all parameters $\boldsymbol{\beta}_{it}$ associated with the plan characteristics to depend on acuity q_{it}, i. e., they vary across individuals i and choice situations t. We divide the plan characteristics, $\mathbf{z}_{itj}$, into two groups that differ in their parameterization. For the coefficients associated with price characteristics (premium and OOP costs), we impose a negativity constraint for all acuity levels.[23] The signs of the coefficients of the other characteristics remain unconstrained. Indexing the plan characteristics by k, their coefficients can be written as

$$\beta_{kit} = \begin{cases} -\exp\left(\beta_k^0 + \beta_k^q q_{it}^*\right) & \text{if} \quad k \in \{\text{premium}, \text{OOP costs}\} \\ \beta_k^0 + \beta_k^q\, \Phi(q_{it}^*) & \text{if} \quad k \notin \{\text{premium}, \text{OOP costs}\} \end{cases} \tag{5}$$

where $\Phi(\cdot)$ denotes the c.d.f. of the standard normal distribution. In order to facilitate the interpretation of the size of the non-cost coefficients, we use standardized acuity, $\Phi(q_{it}^*)$. This specification implies that these parameter values are bounded between β_k^0 (for very low acuity) and $\beta_k^0 + \beta_k^q$ (for very high acuity).

The lagged dependent variable y_{it-1j} indicates the default plan, i.e., the plan in which individual i is enrolled in year $t-1$ or the plan that the individual's $t-1$ plan was cross-walked to. Thus, $y_{it-1j} = 0$ is associated with a switch; a value of 1 indicates that individual i's chosen plan j is the same as the default plan. As in prior literature, the coefficient of the lagged dependent variable, γ_{it}, is interpreted as reflecting switching costs. We allow for the possibility that switching costs vary across individuals and may also depend on acuity. Thus, the most general specification is

$$\gamma_{it} = \gamma^0 + \gamma^q\, \Phi(q_{it}^*) + \gamma_i \tag{6}$$

[22] In an earlier version if this paper, Heiss *et al.* (2016), the most general model we described and estimated only allowed for the variance of the error term to vary with acuity whereas the parameters of the plan characteristics were fixed.

[23] Specifications without the explicit restriction of negative cost coefficients had a worse model fit. Moreover, cost coefficients with a positive density in a neighborhood of zero create problems of interpretation and for the calculation of willingness to pay; see McFadden and Robles (2019).

with a random component that is assumed to by normally distributed,

$$\gamma_i \sim \mathcal{N}\left(0, \sigma_\gamma^2\right). \tag{7}$$

We allow individual acuity to affect choices in the sense that individuals with higher values of q_{it}^* give relatively more weight to utility contributions of observed plan characteristics and the switching cost. Those with lower values of q_{it}^* are relatively more affected by the noise involved in the error terms. Thus, $v_{itj} = \lambda_{it}\tilde{v}_{itj}$, where $\tilde{v}_{itj}$ is an Extreme Value Type I random variable. The error term of the choice model is heteroscedastic, and we allow its variance to depend on acuity by specifying

$$\lambda_{it} = \exp(\eta q_{it}^*). \tag{8}$$

With a plan-choice parameterization resembling a heteroscedastic mixed multinomial logit model, the conditional probability that individual i chooses plan j from choice set $\mathcal{Y}_{it}$ at time t given she pays attention is

$$p_{itj}^y(c_i, \gamma_i) \equiv \Pr(y_{it} = j | a_{it} = 1; c_i, \gamma_i, \cdot) = \frac{\exp\left(\frac{\boldsymbol{z}_{itj}\boldsymbol{\beta}_{it} + \gamma_{it} y_{it-1j}}{\lambda_{it}}\right)}{\sum_{l \in \mathcal{Y}_{it}} \exp\left(\frac{\boldsymbol{z}_{itl}\boldsymbol{\beta}_{it} + \gamma_{it} y_{it-1l}}{\lambda_{it}}\right)}. \tag{9}$$

This choice model incorporates the possibility that individuals with default plan who end up paying attention have a relatively high acuity and therefore tend to pick "good" plans with low total costs. We have seen a first indication of this effect in Section 3.4 where we reported that switchers have a lower realized overspending than individuals without a default plan. Ignoring this effect would lead to biases in the policy conclusions. Specifically, if active choices are observed only for a selected group of high-acuity individuals, simpler models would overestimate the gain from reductions in inattention and switching costs as they cannot predict the behavior of the full population using the estimates of a selected population without adjusting for the acuity differences.

We end this section by discussing how our model captures inertia. At the end of a year, an individual may stay in her default plan for three reasons. First, she is inattentive and so she stays in her default plan. Second, she is attentive, realizes that there is a better plan but does not switch because of switching costs. Third, she is attentive and decides that her default plan is the best option. Importantly, not all individuals that stay in the default plan do so because of inertia. Inertia reflects only the first two reasons. Our model allows us to quantify these channels.

As a thought experiment, we can compare the probability of staying in the default plan to that of switching to an otherwise identical plan. We use the difference between these two probabilities to quantify inertia in the empirical analysis below.

4.5 Alternative model specifications

The general model presented in the previous sections nests several special cases that have been studied in prior literature. These special cases are obtained by restricting parameters of the general model to zero. We specify and estimate five increasingly complex versions. The behavioral features of these models are described, together with the associated parameter restrictions, in Table 5.

Model I, the standard model in the literature, consists only of a choice stage; it is a standard mixed (random parameters) multinomial logit model with heterogeneous switching costs. The only free parameters are $\boldsymbol{\beta}^0, \mu_\gamma, \sigma^2_\gamma$; all others are restricted to zero. As discussed in Section 2, the coefficient of a dummy variable for the default plan has been interpreted as a measure of switching costs in earlier literature. Importantly, this simple model attributes all inertia to switching costs.

In Model II, we add an attention stage and thus allow for inattention as an additional source of inertia. The attention stage includes acuity, a latent measure of decision-making competence modeled as a function of observable characteristics, associated with parameter vector $\boldsymbol{\alpha}$ and attention triggers with the parameter vector $\boldsymbol{\zeta}$. Model III adds unobserved heterogeneity in the equation for the latent acuity variable, with estimated standard deviation σ_c.

Model IV further allows for an effect of acuity on plan choice, conditional on attention, which introduces heteroskedasticity at the plan-choice stage. In other words, individuals' plan choices vary not only with observed plan characteristics and switching costs, but also with their acuity. The parameter η captures this effect. We expect that individuals with higher levels of acuity tend to make better plan-choice decisions ($\eta < 0$).

Finally, Model V corresponds to the full two-stage model described above; in addition to Model IV it includes interactions of the coefficients of the plan characteristics with acuity, captured by the parameter vector $\boldsymbol{\beta}^q_{it}$.

When estimating the restricted models I through IV, the parameters are optimized to best explain the observed patterns such as inertia in choices. For example, in Model IV heterogeneity is suppressed in switching costs but not in the attention stage. Therefore, some of the non-switching observed in the data that would be attributed to switching costs in the full model is captured by the attention stage in the restricted model.

Given the nested parametric structure of the full model, we can test the behavioral restrictions implied by Models I–IV. The results are reported together with measures of model fit in Section 5.2 below. To preview our findings, the more flexible models tend to fit the data better and they produce more realistic estimates of switching costs.

4.6 Identification

Data on repeated plan choices reveal whether a beneficiary switches plans for the next period. The identification problem arises because if she does not switch, we do not observe the reason why.

The simplest explanation for why a beneficiary decides to stay in the default plan is that she carefully compared plans and came to the conclusion that her default plan is the most attractive one available for the next year, t. Given that on average around 50 plans are available and that both the list of available plans and their features (like the premium, deductible, list of covered drugs, copayments and others) change on a yearly basis, this alone is unlikely to explain the low switching rates observed in the data. We are able to account for a large part of the plan characteristics that should be relevant to an individual who chooses an insurance plan to minimize total costs. We observe exactly which drugs an individual takes and can reconstruct the financial consequences of buying these drugs under each available plan, as reflected in our total cost measure. We also account for specific plan features such as the premium and deductible as well as the risk associated with each plan. On top of that, we include alternative-specific constants for the largest plans. Allowing for preference heterogeneity provides a comprehensive model of deliberate plan choices. The remaining persistence that cannot be explained by plan features and preference heterogeneity is what we call inertia and attribute to inattention and switching costs.

For identification, we can exploit the fact that some individuals are forced to make an active choice because they enter the market for the first time or their $t-1$ plan is discontinued without a cross-walk to a new default. For similar arguments, see Handel (2013), Handel and Kolstad (2015) and others. These individuals by definition do not have the opportunity to be inattentive. If they are otherwise comparable to the other consumers and have the same distribution of preferences – conditional on acuity – their behavior alone identifies the plan choice parameters and the difference to the individuals with default plans is attributed to inertia. As individuals with and without default plan have similar means of the observed variables (in particular age, education, and poverty rate; see Table 6 below), the assumption of similar preferences across the two groups, particularly when conditioned on acuity, seems reasonable. Also, they are very similar in terms of the choice sets they face; see Table 7 below.

The explanatory variables that drive the attention probability and the plan choices are different in a quite natural fashion, creating exclusion restrictions that add to the identification of the model. Take the premium as an example. If the premium of the $t-1$ plan *increases* in t – a change that is made salient in the ANOC letter – our model implies that attention is triggered and the individual starts to compare alternatives. If on the other hand the *level* of the premium of the $t-1$ plan is high for year t relative to the alternatives – a fact that is only apparent after comparing plans – this might contribute to overcoming switching costs and changing the plan. The same holds for the other predictors. Attention is triggered by different shocks or changes in the past – we use experience with Part D in $t-1$, changes of the $t-1$ plan – including whether it was cross-walked (consolidated) to a different plan, and health shocks and health care use in $t-1$, see Table 6 – , whereas for the plan choice, comparison to the other plans provide the relevant predictors. Importantly, in this market attention triggers such as changes in the old ($t-1$) plan's premium and other features are highly salient as the ANOC letters prominently contain pairwise comparisons of the old and new values of plan features, see also Ho *et al.* (2017). The assumption that changes in state variable trigger attention is also in line with theoretical models of inattention such as Bordalo *et al.* (2013). In a study of Medicare Part D choices that is comparable to ours, Abaluck and Adams (2017) test the exclusion restriction that characteristic levels impact utility while characteristic changes conditional on levels impact only attention, and they find it "roughly statisfied" (p. 36).[24]

A final identification issue arises in dynamic panel data models with lagged dependent variables if unobserved initial conditions and unobserved heterogeneity are correlated, see Heckman (1981). In principle, we could include initial choices for everyone in our sample as we observe Part D plan choice since the program's introduction in 2006. However, as discussed in Heiss *et al.* (2010) and Heiss *et al.* (2013), the choice situation faced by consumers in the first year of Medicare Part D was special due to an extended enrollment period which ended in May as well as not fully functioning decision aids such as Plan Finder. We therefore opted against including the data from the first year.

[24] We have also estimated the model relaxing the exclusion restrictions by including experiences with Part D in $t-1$ and health care use in $t-1$ in the acuity equation (1). The results (available from the authors upon request), are very similar.

4.7 Maximum likelihood estimation

Given our parametric specification, we can estimate all model parameters simultaneously using maximum likelihood. For notational convenience, we collect all parameters in the vector $\boldsymbol{\theta}$; for an overview, see Table 5. Since we assume independence across individuals, the likelihood function is the product of the individual likelihood contributions $\mathcal{L}_i(\boldsymbol{\theta})$. The observed outcome for individual i is the sequence of plan choices made in years $t = 1, \ldots, T$ (while attention is unobserved). Let j_{it} denote the observed plan choice in year t. Thus, the likelihood contribution is given by

$$\mathcal{L}_i(\boldsymbol{\theta}) = \Pr(y_{i1} = j_{i1}, \ldots, y_{iT} = j_{iT} | \boldsymbol{\theta}, \boldsymbol{d}_i)\,, \tag{10}$$

where for notational convenience, $\boldsymbol{d}_i$ collects the histories of the observed covariates $\boldsymbol{w}_{it}$, $\boldsymbol{x}_{it}$, and $\boldsymbol{z}_{it}$. Further, conditional on these covariates as well as individual i's unobserved acuity c_i and her time-constant switching cost component γ_i, her choices are independent over time. Put differently, any observed dependence is due to these two latent characteristics, conditional on observables. We can therefore write

$$\Pr(y_{i1} = j_{i1}, \ldots, y_{iT} = j_{iT} | \boldsymbol{\theta}, \boldsymbol{d}_i, c_i, \gamma_i) = \prod_{t=1}^{T} \Pr(y_{it} = j_{it} | \boldsymbol{\theta}, \boldsymbol{d}_i, y_{it-1}, c_i, \gamma_i)\,. \tag{11}$$

These probabilities are readily available given our assumptions on the attention and plan-choice stages. We first obtain the choice probabilities for period t:

$$\begin{aligned} \Pr(y_{it} = j_{it} | \boldsymbol{\theta}, \boldsymbol{d}_i, y_{it-1}, c_i, \gamma_i) &= p^a_{it}(c_i) \cdot \Pr(y_{it} = j | a_{it} = 1, \boldsymbol{d}_i, y_{it-1}, c_i, \gamma_i) \\ &+ (1 - p^a_{it}(c_i)) \cdot \Pr(y_{it} = j | a_{it} = 0, \boldsymbol{d}_i, y_{it-1}, c_i, \gamma_i) \end{aligned} \tag{12}$$

This expression involves probabilities that condition on attention, which is not observable. By plugging in the quantities derived in (3) and (9) above, we see that the choice probabilities have different forms depending on whether the individual has a default plan or not and whether she stays in the default plan or switches to a different plan, which are both observable events. For individuals without a default plan ($f_{it} = 1$), the attention probability is $p^a_{it}(c_i) = 1$. If individuals with default plan ($f_{it} = 0$) do not pay attention, the probability of choosing a different than the default plan (i. e., $y_{it-1 j_{it}} \neq j_{it}$), is zero, the probability of choosing the default plan is one.

Putting everything together, for any plan $j \in \mathcal{Y}_{it}$, we get

$$\Pr(y_{it} = j | \cdot, c_i, \gamma_i) = \begin{cases} p^y_{itj}(c_i, \gamma_i) & \text{if } f_{it} = 1 \\ p^a_{it}(c_i) \cdot p^y_{itj}(c_i, \gamma_i) & \text{if } f_{it} = 0 \wedge y_{it-1} \neq j \\ p^a_{it}(c_i) \cdot p^y_{itj}(c_i, \gamma_i) + (1 - p^a_{it}(c_i)) & \text{if } f_{it} = 0 \wedge y_{it-1} = j \end{cases} \tag{13}$$

To obtain the probability of individual i's entire choice sequence, we plug this expression into equation (11). To calculate the likelihood contribution, the final step is to integrate out the remaining two latent quantities, c_i and γ_i:

$$\Pr(y_{i1} = j_{i1}, \ldots, y_{iT} = j_{iT} | \cdot\,) = \iint \prod_{t=1}^{T} \Pr(y_{it} = j_{it} | \cdot\,, c, \gamma)\, g(c) h(\gamma)\, \mathrm{d}c\, \mathrm{d}\gamma\,, \tag{14}$$

where $g(\cdot)$ and $h(\cdot)$ are the densities of c and γ, respectively. Both latent variables are assumed to follow a normal distribution. Recall that we impose the normalization that c is standard normally distributed, whereas mean and variance of the normal variate γ are left unrestricted. The double integral does not have an analytic solution. We use a Gaussian quadrature product rule to accurately approximate it.[25]

5 Results

In this section, we present the main empirical results. We begin by providing descriptive statistics of the observed covariates and choice outcomes. We then present the parameter estimates of our choice model, focusing on the most comprehensive specification (Model V), and interpret these results in terms of the implied marginal effects. We end the section by comparing these results with those obtained from estimating the simpler nested Models I–IV.

5.1 Descriptive statistics

Table 6 contains descriptive statistics of the covariates that enter the acuity equation and the attention stage, separately for individuals with and without a default plan and stratified by switching status. The top panel shows variables that are potential triggers of attention. As most of the individuals without a default newly enter Medicare Part D, the attention trigger variables that refer to information in $t-1$ are not available for them. The attention trigger variables are thus not shown for individuals without a default. The bottom panel of this table shows demographic controls that may affect individuals' acuity when making Part D choices.

There are three different sets of attention triggers: (1) experience in Part D in year $t-1$, (2) changes in features of the year $t-1$ plan, and (3) health shocks and prescription drug

[25] With an even more flexible model including several random parameters, the dimension of this integral would increase and product rule Gaussian quadrature would become infeasible. However, in that case it could be approximated using either Monte-Carlo simulation or powerful multidimensional integration rules such as sparse grids, see Heiss and Winschel (2008).

use in year $t-1$. The descriptive statistics suggest that Part D experience and changes in the old plan may be particularly relevant triggers: while among those who decide to switch plans 23% were in the coverage gap in year $t-1$, only 18% among stayers hit the gap. Similarly, switchers face an average premium increase of $130 per year while stayers only face an average increase of $70. 15% of switchers see changes in cost sharing, either from coinsurance to copayment or vice versa, compared to only 8% among stayers.

With respect to the variables that enter the acuity equation, switchers and stayers appear to be rather similar. A small difference is observed in the fraction of non-whites, which is 5% among switchers and 7% among stayers. As most individuals without a default plan are newly enrolled in Medicare Part D, it is not surprising that these individuals are younger, more likely to be male (which is correlated with age), and have fewer years of experience with Part D than individuals with a default. With respect to all other demographic variables, the two groups are very similar.

Table 7 compares plan characteristics in the choice sets faced by the individuals in the estimation sample, again for individuals with and without a default plan and separately by switching status. Overall, the choice sets seem to be very similar across the three groups of individuals. On average, they can choose among roughly 52 plans, with very similar average premiums, deductibles, ICLs, and gap coverage. The last plan characteristic is the variance of the OOP costs on the plan level. This measure is meant to capture the variability of costs that consumers can expect with each plan. Including the plan-level variance of OOP as a regressor allows risk preferences to play a role.

Coming back to our discussion of identification in Section 4.6 above, these descriptive findings are reassuring for two reasons. First, if individuals without a default were very different from those with default plan, it would be questionable whether individuals without default can help to identify preferences for plan characteristics for those with default plan. Second, the choice sets that the two groups face are similar on average, i. e., conditional on attention the groups face similar choices, except that the default plan is included in the choice set for those with default plan.

5.2 Main estimation results

As all models are estimated by Maximum Likelihood, their predictive power can be readily assessed and compared. Table 8 reports the values of the Akaïke and Bayesian information criteria (AIC and BIC) as well as the likelihood-ratio test statistics for pairwise comparisons to Model V. According to both the AIC and the BIC, Model V fits the data best, and the LR tests clearly reject the simpler models as well. Thus, the most comprehensive Model V is our preferred model. In this subsection, we present the estimation results obtained using Model V and discuss their substantive interpretation.[26]

Table 9 displays the estimates of the parameters of the attention stage, equation (2), and of the acuity equation (1). Table 10 shows those for the plan-choice stage, equation (5). These tables also show the implied marginal effects of the covariates. For the attention stage and the acuity equation, we calculate the marginal effect on the probability of paying attention as well as on the probability of switching plans. For the plan-choice stage, we calculate the marginal effect on the probability of choosing a plan conditional on attention as well as overall (i. e., not conditioning on attention).

Attention stage

Recall that attention is a latent binary outcome. Our model estimates imply a mean attention probability of 32 percent. The attention probability varies with observed covariates and it is also heterogeneous conditional on these covariates. We discuss the effects of observed covariates first and return to conditional heterogeneity below.

The attention trigger variables fall into three groups: experience in Part D in t–1, changes in the features of the $t-1$ plan announced for year t, and health shocks as well as health-care use in $t-1$. As Table 9 shows, the coefficients of all experience variables and several of those in the other two groups are statistically significant at the 0.01 level. Despite the large sample size, some of the health variables are not significant; they do not seem to raise enrollees' awareness of the opportunity to switch plans. In the following, we discuss the effect sizes of these covariates in terms of their marginal effects on the probability of attention. Table 9 also reports marginal effects on switching. As switching requires attention, the marginal effects of the covariates on switching are always smaller than those on attention.

Perhaps the most salient aspect of an individual's Part D experience is whether she hit the coverage gap and if so, whether she was enrolled in a plan that provides gap coverage. We thus include two dummy variables and their interaction. The marginal effect of having gap coverage in year t–1 for individuals who did not hit the gap in t–1 is -13.8 percentage points. For those who did hit the gap in $t-1$, it is 6.7 percentage points smaller, for a total effect of -20.5 percentage points. This is a large effect: Individuals who hit the gap and do have gap coverage are much less likely to even consider switching for year t. Having hit the gap without gap coverage in year $t-1$ increases the attention probability

[26] The estimation results for the other four models can be found in Tables A1 to A7 in the Appendix.

by 14.8 percentage points, which also is a sizeable effect. Similarly, individuals who have hit the catastrophic coverage region in $t-1$, i. e., those with the largest drug bills, are more likely to be attentive as well (by 3.9 percentage points).

As plans are required to send their enrollees ANOC letters before the open enrollment period each year, all changes of the features of the year $t-1$ plan should be salient. We find that individuals indeed are more likely to pay attention if the premium of their old plan increases, the deductible increases, the ICL increases, their old plan introduces changes to the formulary that would lead to an increase in OOP cost, or their old plan switches from copayments to coinsurance or *vice versa*. For instance, an increase of the year t–1 premium by 100 dollars results in a 10.4 percentage point increase of the attention probability. Interestingly, we find that premium changes are particularly salient, as the effect of a premium increase is much higher than the effects of an increase in the deductible or the annual OOP spending implied by a formulary change. The finding that individuals put more weight on premiums than on the cost implications of other plan features is in line with earlier research, including Abaluck and Gruber (2011) and Heiss *et al*. (2013). Furthermore, we find that the effect of the implied change in OOP cost due to formulary changes is much smaller (0.9 percentage point increase with a 100 dollar increase in OOP), than the effect of a change in the premium or the deductible. This likely reflects that cost consequences of formulary changes are less salient than changes in premium and deductible.[27]

Attention probabilities are higher if a $t-1$ plan is consolidated and if there is a change in cost-sharing (from copay to coinsurance and *vice versa*). Such changes might make a plan more or less attractive, so we do not have a strong prior on the sign of these effects. An increase of the ICL (as a higher limit indicates that the plan gets better) and a lower share of tiers with increases in cost-sharing have insignificant or very small effects. However, as the regression already controls for the change in total OOP spending, these effects are not straightforward to interpret.

One would expect that having been diagnosed with a new health condition that is costly in terms of prescription drugs or having had emergency room visits or hospital stays in the year $t-1$ raises attention to the Medicare Part D decision. However, this is not borne out by the data. The variables that measure health shocks and health-care use are mostly statistically insignificant or have small effects. However, we find a statistically significant

[27] While the ANOC letters contain summary tables on changes in premiums, deductibles and copayments, the exact formulary changes, i. e., changes in the list of drugs and their tiers, are only provided in an appendix to the letter.

but small effect of a dummy variable for whether an individual had five doctor visits or more in the previous year. As the model controls for the other health variables, this finding suggests that repeated interaction with a health professional raises attention. The data do not allow us to pursue such potential information channels further.

Acuity equation

An important determinant of attention is acuity – a latent variable that captures how careful individuals are in making their choices. It is modeled as a function of observable individual characteristics and an unobserved heterogeneity term. As the log likelihood values and the information criteria in Table 8 show, including acuity in the model improves predictive power considerably.

The structure of Model V allows us to characterize the association between acuity and attention. Figure 2 shows the distribution of the estimated individual-level attention probability (given the values of all other variables), stratified by decile of acuity. The median attention probability for individuals in the lowest acuity decile is about 12 percent whereas for those in the top acuity decile it is above 50 percent.

The bottom panel of Table 9 shows the coefficients of the observable characteristics in the acuity equation as well as their marginal effects on attention and switching. The coefficient estimates imply higher levels of acuity for females, whites, younger individuals, individuals who have less experience with Part D, individuals who live in ZIP code areas with higher average education among seniors, those living in middle-income ZIP codes, and those who have no depression diagnosis in their claims history.

We comment on the implied marginal effects of these covariates on attention in turn.[28] Beginning with the socio-demographic characteristics, the difference in acuity between males and females is small. The effect of being non-white is larger, with an implied marginal effect on attention of about 3 percentage points. (This estimate obviously also captures the effects of any unobserved individual characteristics that are correlated with race.) Acuity declines with age. Relative to the base category (69 or younger) those in the oldest age category (80 years or older) have an implied attention probability that is lower by 2.9 percentage points.

An important issue in the literature on Medicare Part D is whether choices improve over time as consumers learn to navigate this rather complicated market. The prior literature on this question delivered mixed results, see Ketcham *et al.* (2012) and Abaluck and

[28] As mentioned before, the marginal effects of switching are smaller and we do not not discuss them separately.

Gruber (2016a). We find that for each additional year an individual has been enrolled in Part D, the attention probability is reduced by 0.8 percentage points, *ceteris paribus*. The fact that individuals become less attentive to plan switching over time, conditional on all other variables including attention triggers, suggests that individuals' decision-making does not improve over time – rather, acuity declines.

It would be interesting to test whether acuity is predicted by measures of cognitive ability, decision making competence, or financial literacy, see for example Barcellos *et al.* (2014). Measures of these constructs are not available in the Medicare claims data. However, the attention probability of individuals who have been diagnosed with depression is lower by 1 percentage point. This partial measure of mental health, which might also be related to decision making competence, predicts at least part of the heterogeneity in Medicare Part D decision-making (controlling for several other predictors).

The proxies for socio-economic status, constructed at the ZIP code level, also predict acuity and therefore attention. An increase of the share of seniors without a high-school degree by 1 percentage point is associated with a decrease in the attention probability of 0.065 percentage points. As ZIP code level income measures, we include the shares of poor (income below the FPL) and rich (income above 500% of the FPL) seniors. Attention probabilities are largest for the reference group of individuals with incomes between 100% and 500% of the FPL, as the poor and rich shares increase by 1 percentage point, attention probabilities decline by 0.025 and 0.09 percentage points, respectively. As these ZIP-code level measures are noisy, the relatively small magnitudes of the marginal effects are not surprising. Nevertheless, they suggest that socio-economic status might play a role in determining the quality of Medicare Part D plan choices. We speculate that the relatively poor show lower levels of acuity because of lower decision-making competence. Among the relatively rich, acuity might be lower because they feel less inclined to worry about plan choice in Part D, which in turn might reflect their opportunity cost of time. This interpretation is consistent with the finding that richer individuals tend to select high-premium plans in Medicare Part D, independent of their risk; see Heiss *et al.* (2013).

Plan-choice stage

Turning to the results for the plan-choice stage, shown in Table 10, note first that in our two-stage model, plan choices are made only by attentive individuals. While individuals without a default plan always pay attention, our model estimates imply an average attention probability of 32% for individuals who have a default. Consequently, there is a 68% probability of being inattentive. Individuals with default plan who are inattentive stay in their default plan and do not enter the choice stage. For the others, plan choices are modeled by a mixed (random parameters) multinomial logit specification with heteroskedasticity. In our preferred Model V, the variance of the errors is a function of acuity and the coefficients of plan characteristics vary with acuity as well. For each plan characteristic k, we therefore report and discuss the values of β_k^0 and β_k^q as defined in equation (5).

The results for the plan-choice stage show sensible signs for all coefficients, and to the extent that they are comparable, they are also in line with earlier results from static plan-choice models such as Abaluck and Gruber (2011) and Heiss *et al.* (2013). In particular, individuals are more likely to choose a plan if it has lower OOP, lower premium, lower deductible, and higher ICL. The marginal effect of having gap coverage is about 0 on average. Also, the effect of a plan's premium is three times as large as that of its implied OOP cost, which confirms earlier findings that individuals are too sensitive to the premium as the most salient feature, see e.g. Abaluck and Gruber (2011) and Abaluck and Gruber (2016a).

In addition to these plan features, the plan-choice model also contains a dummy variable for the default plan. It has a statistically significant and quantitatively important effect. We derive an estimate of switching costs from the associated coefficient below. Finally, we find that individuals are more likely to choose a plan whose implied OOP cost variance is lower, suggesting that individuals are risk averse and value the insurance component of Medicare Part D.

As before, the quantitative interpretation of the results focuses on marginal effects, which we compute both unconditional and conditional on attention. The unconditional marginal effect measures the overall change in the probability of choosing a plan associated with a change in the respective plan feature, averaged across plans and individuals. For example, the probability that a plan is chosen decreases by 0.11 percentage points with an increase in the annual premium by \$100. Given that individuals choose, on average, between roughly 50 plans and thus *a priori* the probability of choosing a plan is roughly 2 percent, this corresponds to a sizable 5.5 percent increase.

When conditioning on attention, marginal effects are larger in absolute value, except for the marginal effect of the default plan dummy, which captures switching costs. This increase reflects the fact that the effects of the plan characteristics are watered down by inattention when not conditioning on attention. For example, a plan's annual premium has a lower impact on its choice probability among all individuals than among the subset of individuals who pay attention. That is, those who are inattentive do not even consider the plan's premium when making their choice.

The unconditional marginal effect of the default plan dummy measures the overall difference in choice probabilities between the default plan and an otherwise identical alternative. We interpret this quantity as a measure of average inertia. Part of this inertia is driven by inattention as discussed above. The marginal effect conditional on attention reflects the part of inertia that is due to switching costs.

Figure 3 illustrates how the unconditional marginal effect of the default plan dummy varies with acuity. Almost all individuals in the lowest acuity decile stick with their default plan. As acuity increases this share declines. In the top decile the median marginal effect of the default plan dummy is only about 50 percentage points.

As discussed before, the coefficient of the default plan dummy variable can be interpreted as a measure of switching costs. The raw model parameters are measured in 'utility units'. We can translate them into dollar units because a price variable (plan premium) is included among the regressors.[29] Switching costs, c_{it}, are calculated as ratio of the coefficient of the old plan dummy and the coefficient of the premium (multiplied by 1000 because of the scaling of the premium variable),

$$c_{it} = -1000 \frac{\gamma_{it}}{\beta_{premium,it}}. \tag{15}$$

In our preferred Model V both parameters are allowed to vary with acuity. Figure 4 shows the resulting relationship between acuity and average switching costs.[30] We condition acuity on attention, as switching costs are only relevant for attentive individuals. For individuals with the lowest acuity, switching costs estimates become very large whereas for high acuity people they approach 0. There is also an association between attention and switching costs as Figure 5 illustrates. Those with low switching costs are more attentive and *vice versa*.

Returning to Table 10, switching costs imply an increase of the probability of choosing the default plan by 30.4 percentage points conditional on attention. Not conditioning on attention, the overall probability of choosing the default plan is 76.1 percentage points higher than the probability of choosing an otherwise identical plan. This large unconditional effect arises because all inattentive individuals stay in the default plan by definition.

[29] We choose premium to translate 'utility units' into dollar units because as discussed above it is the most salient among the price variables included in the model.

[30] We set the random coefficient γ_i to zero so that switching costs we report are the conditional average w.r.t. acuity.

Finally, we discuss the estimates of the acuity interaction terms, $\boldsymbol{\beta}^q$ and γ^q. The estimated coefficients are highly statistically significant, which shows that individuals' valuations of plan characteristics varies with acuity. The overall picture is that individuals with high acuity value the financial characteristics of plans more strongly.[31] However, the specific estimates are not straightforward to interpret because of the underlying heterogeneity together with the acuity interactions. One way to illustrate the heterogeneity is to look at the distribution of the marginal effects, which is partly due to acuity. Table 11 shows the 5$^{\text{th}}$ percentile, the median, and the 95$^{\text{th}}$ percentile of the conditional and unconditional marginal effects (whereas Table 10 shows just the means). For example, the marginal effect of an increase in the annual premium by \$100 varies between minus 2.7 percentage points at the 5$^{\text{th}}$ percentile and about 0 at the 95$^{\text{th}}$ percentile. The variation in marginal effects is considerable for all plan characteristics, including the default plan dummy. This is consistent with Figure 3, which however shows heterogeneity with respect to acuity alone.

5.3 Comparison of Models I–V

In this section, we compare the five models in terms of their behavioral implications. In particular, it turns out that relative to our preferred Model V, the simpler Models I–IV lead to different estimates of switching costs, inattention, and inertia. Thus, the five models would also have different implications for policy interventions aimed at reducing inertia to which we return in Section 6 below.

The five models lead to quite different estimates of switching costs, as shown in Figure 6. Recall that this measure of switching costs reflects how much individuals are, on average, willing to pay in order to stay in their default plan. Model I attributes all inertia to switching costs. It yields large average switching costs of more than \$1000. When we allow for inattention as a source of inertia, estimated switching costs are reduced substantially to roughly \$660. Given that average premiums are only about \$550 per year, this is still a rather high number. Further reductions in estimated switching costs arise when we allow for unobserved heterogeneity in the attention stage (Model III) and in the plan-choice stage (Model IV). The reduction in switching costs as we move from Model I to Model IV is somewhat mechanical: In Model I, all inertia realized in the data

[31] The fact that in our preferred Model V acuity has a rich interaction structure explains that by itself is not statistically significant (see Table 10). In the simpler Model IV, where acuity is not interacted with plan characteristics it is significant (see Table A7).

can only result from switching costs. As additional channels that might explain inertia are opened up, switching costs tend to go down.

The situation is more complicated once we, in addition, allow for heterogeneity in switching costs. Figure 4 already showed that this heterogeneity is substantial, with extremely high switching costs for those with low acuity. Figure 6 shows the mean within each quartile of switching costs. For individuals in the bottom quartile (i. e., likely those with high acuity) switching costs are even lower than in Model IV. In contrast, in the upper quartile switching costs are, on average, about three times as large as in Model I. This result might be surprising but it reflects the fact that low acuity individuals very rarely switch plans. We draw two important implications from these results: First, estimates of switching costs depend heavily on the structure of the model and tend to decline as the attention stage is introduced and made more flexible. Second, switching costs are related to individuals' acuity (and indirectly to those characteristics that enter the acuity model); allowing for preference heterogeneity in the choice stage is particularly important.

Similarly to the estimates of switching costs, the other parameter estimates and marginal effects also change across model specifications. In the Appendix we report these estimates, see Tables A1–A7 which correspond to Tables 9 and 10 above. Furthermore, our models differ in how much of inertia is attributed to inattention and switching costs, respectively.

As explained above, inertia is captured by the unconditional marginal effect of the default plan dummy. Model I – which does not allow for inattention – results in a large estimate of switching costs (see Figure 6) and underestimates average inertia (41 percentage points, see Table A1). Along with declining switching costs, the estimate of average inertia increases to 81.4 percentage points in Model IV (see Table A7). In Model V the estimate of average inertia is similar (76.1 percentage points with a median 0.78) but there is considerable heterogeneity, see Figure 3.

6 Simulations: Forcing attention and removing switching costs

Several policy interventions aimed at increasing switching rates have been suggested in the literature, both in Medicare Part D and in other markets where consumer inertia is an issue. Such proposals typically aim to reduce the transaction costs associated with switching plans (e. g., Hoadley *et al.* (2013)); increase attention using reminders, nudges, and other information shocks (e. g., Sinaiko *et al.* (2013)); or reduce the complexity of the choice environment, for instance by reducing the number of available options (e. g., Frank and Lamiraud (2009), Besedeš *et al.* (2015)). Our simulations are also related to a series of studies investigating "active choice" policies, including Carroll *et al.* (2009), Keller *et al.* (2011), and Chetty *et al.* (2014).

We use the estimates of our five models to simulate the effects of two counterfactual scenarios on the probability that individuals switch plans and on *ex post* overspending. In the first scenario, we force all individuals to pay attention. In the second scenario, we set switching costs to zero. These two scenarios are extreme and therefore allow us to bound the effects of more realistic interventions that increase attention or reduce switching costs. We focus on the most comprehensive and statistically superior Model V. As discussed in section 5, Models I–IV which are nested in Model V are rejected both by likelihood ratio tests and perform much worse in explaining observed choices according to the AIC and BIC information criteria. However, it is not clear *a priori* whether this translates into different conclusions when they are used for counterfactual simulations. It is therefore instructive to simulate the counterfactual scenarios for these four models as well.

Figures 7 and 8 display the simulation results. For Models I–V, we report the average probability to switch and average overspending for three scenarios: the baseline case (i. e., using the estimates based on the observed choice data as reported in Section 5) and the two counterfactual interventions.

As panel (a) of Figure 7 shows the effects of the counterfactual interventions vary substantially across the five models. In Model I, which does not have an attention stage, removing switching costs increases the mean switching probability dramatically from 10.4 percent to 85.5 percent. Once the model includes an attention stage, the effect of removing switching costs becomes smaller, while forcing attention increases the switching probability. In Model IV the effect of forcing attention is large with an implied switching probability of almost 68 percent. However, in Model IV switching costs do not vary with acuity. To the extent that attention and switching costs both vary with acuity, the estimates of the effects of forcing attention or removing switching costs are distorted. In the full Model V, these two counterfactual interventions lead to smaller increases in the switching probability of 25.5 percent and 28.8 percent, respectively, which are arguably more realistic.

Panel (b) of Figure 7 illustrates how much acuity matters for the effects of these counterfactual interventions. Already in the baseline situation, switching probabilities for low acuity individuals are negligible at 1.3 percent, whereas in the high acuity group they average to 25 percent. The effects of the counterfactual interventions are also larger in absolute terms for the high acuity individuals.

Figure 8 shows the effects of the two counterfactual interventions on average overspending. The picture is similar as for switching probabilities. The reduction in overspending induced by forcing attention is more modest in Model V than in Model IV, but still relevant with a decrease by almost 10 percent. Again, heterogeneity with respect to acuity is important as shown in panel (b). Overspending decreases with acuity in baseline and the relative reduction associated with the two counterfactual interventions is larger for the high acuity group. The strongest effect arises for the high acuity group in the forced attention scenario (a reduction of about 22 percent). One implication of these results is that both interventions do not reduce overspending for low acuity individuals and they help high acuity individuals the most.

The simulation results obtained from Model V provide additional insights on the heterogeneity of the effects of counterfactual interventions. Figure 9 shows the empirical cumulative distribution functions (CDFs) of overspending in the baseline scenario and in the two counterfactual scenarios: Both interventions lead to a reduction over much of the distribution but removing switching costs seems to lead to an increase in overspending for some individuals. This can be seen more clearly in Figure 10 which displays the empirical CDFs of the reduction in overspending that is induced by the two counterfactual simulations, relative to the baseline case. Forcing attention is associated with higher reductions in overspending. Interestingly, in both counterfactual scenarios a sizable fraction of individuals choose plans that imply higher overspending. The increase in overspending is larger in the no switching costs scenario. Again, underlying these results is substantial heterogeneity with respect to acuity.

The two graphs in Figure 11 show the distributions of the reduction in overspending induced by policy interventions stratified by acuity. Graph (a) shows that the savings induced by forcing attention are much higher for high acuity individuals as they make better choices when forced to do so even if they were inattentive in baseline. In contrast, the reduction in overspending does not vary as much with acuity when switching costs are removed, as shown in Graph (b). Our interpretation is that with the attention stage in place, low acuity people are much less likely to enter the choice stage, so removing switching costs will not affect them much. Moreover, low acuity people tend to choose worse plans in the choice stage, as shown in Figure 8 (b) above. High acuity individuals have lower switching costs to begin with, so the effect of this intervention is relatively small for them, too.

An important implication of our simulation results is that while some individuals profit when they are forced to pay attention, others are harmed – at least with respect to *ex post* overspending. For policymakers, the balance between winners and losers of such reforms is of particular interest. The magnitudes of these effects depend on whether heterogeneity in acuity is allowed for, as we do in our preferred Model V.

These are interesting interaction effects that should be accounted for when policy interventions are simulated using behavioral models. Beyond the specific structure and the application to Medicare Part D, we believe that highlighting this channel is an important contribution of our model.

7 Summary and conclusions

In Medicare Part D and other health insurance markets, only few consumers switch plans across years; many fail to switch to cheaper plans with comparable coverage and thus forego significant savings. In this paper, we separated channels that contribute to this high level of inertia: inattention to plan choice and perceived costs of switching plans. To characterize these two sources of inertia, we developed a two-stage panel data model of plan choices. The first stage models whether individuals pay attention to their plan choice or not. Given that individuals pay attention, the second stage models the plan choice in a discrete choice framework. The model includes switching costs and allows for unobserved individual-specific heterogeneity in acuity that leads to correlation of the two stages. We estimate the model based on a sample of individuals in stand-alone PDP plans.

Our analysis highlights the fact that in the market for Medicare Part D stand-alone plans, attention to the possibility of switching plans is triggered by events and experiences that are salient during the open-enrollment period. In particular, for those individuals who do not have gap coverage in their current plan, having reached the gap is associated with a higher propensity to switch. Premium and formulary changes also trigger attention, but the effect of premium changes is much larger than that of changes in OOP costs induced by changes to the formulary. This finding is interesting since formulary changes are clearly relevant for switching decisions but not as salient as premium changes. An implication for firm behavior is that consumers' reactions to increases in their plans' costs will be larger when they are implemented via the premium rather than other characteristics, as for example copayments or obscure details of the formulary benefit design including tier assignment of specific drugs. Furthermore, our results suggest that new plans in the market may not be considered by inattentive consumers, to the extent that market entry is not salient to them.

Another interesting finding is that acuity declines with experience, which seems inconsistent with the notion that people learn. This result corroborates recent findings of Abaluck and Gruber (2016a) who also argue that choice inconsistencies in Medicare Part D increased over time. Also, individuals who pay attention in the baseline scenario have higher acuity and therefore make better decisions, at least in terms of total costs associated with their plan choice and current prescription drug use, than those who do not pay attention. Finally, our estimates indicate that even conditional on attention, there are still considerable implicit switching costs. Understanding the sources of these switching costs is an important topic for future research. Another interesting extension would be to modify the model so that it applies to the Medicare population eligible for the low-income subsidy whose acuity may be lower than that of the stand-alone enrollees. These open issues not withstanding, our two-stage model opens up the possibility to simulate the effects of policies that aim at increasing plan switching by making the economic determinants of optimal health insurance plan choices more salient.

The simulations we performed using the estimates of our full two-stage choice model provide several interesting insights that differ from those one would obtain from simpler ones. For example, as a thought exercise consider eliminating default re-enrollment and forcing consumers to pay attention. This would almost certainly increase switching rates, and probably improve overall welfare, but with exceptions that would tend to be concentrated among low-acuity consumers. Alternatively, consider assigning each returning enrollee a default plan that minimizes their *ex ante* overspending, rather than automatically taking their previous plan as a default. This would not only increase attention, but would forcefully nudge inattentive enrollees in the direction of reduced overspending. Even low-acuity individuals should see welfare improvements, with their high switching costs now operating to keep them from switching away from the favorable default.

More generally, our findings suggest that removing inattention, in our case by forcing every beneficiary to make an active choice, has effects on switching rates that are at least as large as those associated with removing switching costs. This finding echoes Kiss (2014) who found that in the Hungarian car insurance market, a public campaign aimed at increasing switching rates mainly acted through manipulating attention, rather than switching costs. Moreover, we demonstrate that the evaluation of such interventions requires structural models that allow for heterogeneity of consumers with respect to their acuity – that is, their ability and diligence in making diligent decisions.

Recent literature has stressed that reducing the number of available plans is another intervention that would conceivably improve plan choices and increase switching rates.

We leave to future research the issue of whether a structural choice model like ours could reliably capture the effects of changing the number of available options, as discussed by Ericson and Sydnor (2017) and Gruber (2017).[32] Our model could potentially also be used to study the effects of simplifying the Plan Finder or of providing more elaborate, personalized decision support tools. Recent studies by McGarry *et al.* (2018) and Bundorf *et al.* (2019), respectively, suggest that such interventions can improve choices in health insurance markets like Medicare Part D considerably.

The welfare consequences of consumer inertia in health insurance markets are complex and depend on the specific setting. Our findings add to the existing evidence on the excess costs associated with suboptimal plan choices and on the importance of inertia observed in health insurance markets. It seems obvious that consumers should be encouraged to review their health plans more frequently. However, Handel (2013) argues that policies that nudge consumers to better decisions by reducing inertia also exacerbate adverse selection, which potentially leads to a reduction in overall welfare. A further issue is that removing frictions might lead to market unravelling, but as Handel and Kolstad (2015) argue, risk-adjustment could mitigate the negative effects of reducing information frictions. Handel *et al.* (2015) explore this issue further. We leave an extension of our present analysis that would allow for the computation of welfare effects to future research. We note, however, that in addition to these issues, such an extension would necessarily involve additional functional from assumptions, as pointed out in a related context by Ketcham *et al.* (2015) and Abaluck and Gruber (2016b).

Another area of research where our two-stage model might be fruitfully employed is the industrial organization of health insurance markets, along the lines of Ho *et al.* (2017) and Polyakova (2016). Our two-stage model could be used to study how different interventions that reduce inertia affect market equilibria and prices over time. However, as Abaluck and Gruber (2016a), Decarolis (2015), and Decarolis *et al.* (2019) argue, analyzing the choice set evolution is especially complicated due to the rules governing assignment of low-income subsidy enrollees to plans.

More generally, our empirical findings are relevant for the emerging literature on the effects of inattention on consumer choices and their implications for firm behavior, reviewed for

[32] The key issue here is that in data on Medicare Part D, the number of available plans has varied rather little over time. In experimental studies such as Besedeš *et al.* (2015), the number of options can be manipulated over a much larger range.

instance by Grubb (2012) and Grubb (2015).[33] Our results add to the still relatively scarce evidence showing that inattention and attention shocks matter in high-stakes financial decisions.[34] Further, we show that interventions such as active choice policies might alleviate market distortions that arise from inattention. Our findings also stress that allowing for unobserved heterogeneity in such models is crucial.

[33] These papers also highlight connections between inattention and biased expectations or overconfidence. Exploring these biases in the context of Medicare Part D would be interesting but seems infeasible with administrative data on plan choices alone.

[34] This evidence includes Stango and Zinman (2014) on bank overdraft fees, Helmers *et al.* (2016) on online shopping, and Crossley *et al.* (2017) on retirement savings.

References

Abaluck, J. and A. Adams (2017): What do consumers consider before they choose? identification from asymmetric demand responses. NBER Working Paper No. 23566.

Abaluck, J. and J. Gruber (2011): Choice inconsistencies among the elderly: Evidence from plan choice in the Medicare Part D program. *American Economic Review*, 101, 1180–1210.

Abaluck, J. and J. Gruber (2016a): Evolving choice inconsistencies in choice of prescription drug insurance. *American Economic Review*, 106(8), 2145–2184.

Abaluck, J. and J. Gruber (2016b): The robustness of tests for consumer choice inconsistencies: Reply to Ketcham, Kuminoff and Powers. *American Economic Review*, forthcoming.

Abaluck, J., J. Gruber, and A. Swanson (2018): Prescription drug use under medicare part d: A linear model of nonlinear budget sets. *Journal of Public Economics*, 164, 106 – 138.

Aizawa, N. and Y. S. Kim (2018): Advertising and risk selection in health insurance markets. *American Economic Review*, 108(3), 828–867.

Bach, P. B. and M. B. McClellan (2005): A prescription of a modern Medicare program. *New England Journal of Medicine*, 353, 2733–2735.

Barcellos, S. H., A. Wuppermann, K. G. Carman, S. Bauhoff, D. McFadden, A. Kapteyn, J. Winter, and D. Goldman (2014): Preparedness of Americans for the Affordable Care Act. *Proceedings of the National Academy of Sciences of the United States of America*, 111(15), 5497–5502.

Besedeš, T., C. Deck, S. Sarangi, and M. Shor (2015): Reducing choice overload without reducing choices. *Review of Economics and Statistics*, 97(4), 793–802.

Bhargava, S., G. Loewenstein, and J. Sydnor (2017): Choose to lose: Health plan choices from a menu with dominated option*. *The Quarterly Journal of Economics*, 132(3), 1319–1372.

Bordalo, P., N. Gennaioli, and A. Shleifer (2013): Salience and consumer choice. *Journal of Political Economy*, 121(5), 803–843.

Bundorf, M. K., M. Polyakova, C. Stults, A. Meehan, R. Klimke, T. Pun, A. S. Chan, and M. Tai-Seale (2019): Machine-based expert recommendations and insurance choices among Medicare Part D enrollees. *Health Affairs*, 38(3), 482–490.

Carroll, G., J. Choi, D. Laibson, B. Madrian, and A. Metrick (2009): Optimal defaults and active decisions. *Quarterly Journal of Economics*, 124, 1639–1676.

Chetty, R., J. N. Friedman, S. Leth-Petersen, T. Nielsen, and T. Olsen (2014): Active vs. passive decisions and crowdout in retirement savings accounts: Evidence from Denmark. *Quarterly Journal of Economics*, 129, 1141–1219.

Crossley, T. F., J. de Bresser, L. Delaney, and J. Winter (2017): Can survey participation alter household saving behavior? *Economic Journal*, 127, 2332–2357.

Decarolis, F. (2015): Medicare Part D: Are insurers gaming the low income subsidy design? *American Economic Review*, 105(4), 1547–1580.

Decarolis, F., M. Polyakova, and S. P. Ryan (2019): Subsidy design in privately-provided social insurance: Lessons from Medicare Part D. *Journal of Political Economy*, forthcoming.

Duggan, M., P. Healy, and F. Scott Morton (2008): Providing prescription drug coverage to the elderly: America's experiment with Medicare Part D. *Journal of Economic Perspectives*, 22(4), 69–92.

Einav, L., A. Finkelstein, and P. Schrimpf (2015): The response of drug expenditure to nonlinear contract design: Evidence from Medicare Part D. *Quarterly Journal of Economics*, 130(2), 841–899.

Ericson, K. M. and J. Sydnor (2017): The questionable value of having a choice of levels of health insurance coverage. *Journal of Economic Perspectives*, 31(4), 51–72.

Ericson, K. M. M. (2014): Consumer inertia and firm pricing in the Medicare Part D prescription drug insurance exchange. *American Economic Journal: Economic Policy*, 6(1), 38–64.

Farrell, J. and P. Klemperer (2007): Coordination and lock-in: Competition with switching costs and network effects. In M. Armstrong and R. Porter (Eds.), *Handbook of Industrial Organization*, Volume 3, chapter 31, 1967–2002. Amsterdam: North-Holland.

Frank, R. and K. Lamiraud (2009): Choice, price competition and complexity in markets for health insurance. *Journal of Economic Behavior and Organization*, 71(2), 550–562.

Gabaix, X. (2017): Behavioral inattention. NBER Working Paper No. 24096.

Gibbard, P. (2017): Disentangling preferences and limited attention: Random utility models with consideration sets. Unpublished manuscript, Australian National University.

Goldman, D. P. and G. F. Joyce (2008): Medicare Part D: A successful start with room for improvement. *Journal of the American Medical Association*, 299(16), 1954–1955.

Goodman, J. (2006): Consumer directed health care. Policy Brief 2006-PB-20, Networks Financial Institute, Indiana State University.

Grubb, M. D. (2012): Dynamic nonlinear pricing: Biased expectations, inattention, and bill shock. *International Journal of Industrial Economics*, 30, 287–290.

Grubb, M. D. (2015): Overconfident consumers in the marketplace. *Journal of Economic Perspectives*, 29(4), 9–36.

Gruber, J. (2017): Delivering public health insurance through private plan choice in the United States. *Journal of Economic Perspectives*, 31(4), 3–22.

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

Handel, B. and J. T. Kolstad (2015): Health insurance for "humans": Information frictions, plan choice, and consumer welfare. *American Economic Review*, 105(8), 2449–2500.

Handel, B. R., J. T. Kolstad, and J. Spinnewijn (2015): Information frictions and adverse selection: Policy interventions in health insurance markets. NBER Working Paper No. 21759.

Heckman, J. J. (1981): Heterogeneity and state dependence. In C. F. Manski and D. McFadden (Eds.), *Structural Analysis of Discrete Data and Econometric Applications*, 91–139. Cambridge, MA: MIT Press.

Heiss, F., A. Leive, D. McFadden, and J. Winter (2013): Plan selection in Medicare Part D: Evidence from administrative data. *Journal of Health Economics*, 32, 1325–1344.

Heiss, F., D. McFadden, and J. Winter (2006): Who failed to enroll in Medicare Part D, and why? Early results. *Health Affairs*, 25, w344–w354.

Heiss, F., D. McFadden, and J. Winter (2010): Mind the gap! Consumer perceptions and choices of Medicare Part D prescription drug plans. In D. A. Wise (Ed.), *Research Findings in the Economics of Aging*, 413–481. Chicago and London: University of Chicago Press.

Heiss, F., D. McFadden, J. Winter, A. Wuppermann, and B. Zhou (2016): Inattention and switching costs as sources of inertia in Medicare Part D. NBER Working Paper No. 22765.

Heiss, F. and V. Winschel (2008): Likelihood approximation by numerical integration on sparse grids. *Journal of Econometrics*, 144, 62–80.

Helmers, C., P. Krishnan, and M. Patnam (2016): Attention and saliency on the internet: Evidence from an online recommendation system. CEPR Discussion Paper No. 10939.

Ho, K., J. Hogan, and F. Scott Morton (2017): The impact of consumer inattention on insurer pricing in the Medicare Part D program. *RAND Journal of Economics*, 48(4), 877–905.

Hoadley, J., E. Hargrave, L. Summer, J. Cubanski, and T. Neuman (2013): To switch or not to switch: Are Medicare beneficiaries switching drug plans to save money? Issue Brief, The Henry J. Kaiser Family Foundation, Menlo Park, CA.

Honka, E. (2014): Quantifying search and switching costs in the US auto insurance industry. *RAND Journal of Economics*, 45(4), 847–884.

Hortaçsu, A., S. A. Madanizadeh, and S. L. Puller (2017): Power to choose? An analysis of consumer inertia in the residential electricity market. *American Economic Journal: Economic Policy*, 9(4), 192–226.

Kahneman, D. (2011): *Thinking, Fast and Slow*. New York: Farrar, Straus and Giroux.

Keller, P. A., B. Harlam, G. Loewenstein, and K. G. Volpp (2011): Enhanced active choice: A new method to motivate behavior change. *Journal of Consumer Psychology*, 21, 376–383.

Ketcham, J. D., N. V. Kuminoff, and C. A. Powers (2016): Choice inconsistencies among the elderly: Evidence from plan choice in the Medicare Part D program: Comment. *American Economic Review*, forthcoming.

Ketcham, J. D., C. Lucarelli, E. J. Miravete, and M. C. Roebuck (2012): Sinking, swimming, or learning to swim in Medicare Part D. *American Economic Review*, 102, 2639–2673.

Ketcham, J. D., C. Lucarelli, and C. A. Powers (2015): Paying attention or paying too much in Medicare Part D. *American Economic Review*, 105, 204–233.

Kiss, A. (2014): Salience and switching. Unpublished manuscript.

Kling, J. R., S. Mullainathan, E. Shafir, L. Vermeulen, and M. V. Wrobel (2012): Comparison friction: Experimental evidence from Medicare drug plans. *Quarterly Journal of Economics*, 127(1), 199−235.

Lavetti, K. and K. Simon (2018): Strategic formulary design in medicare part d plans. *American Economic Journal: Economic Policy*, 10(3), 154−192.

Luco, F. (2018): Switching costs and competition in retirement investment. *American Economic Journal: Microeconomics*, forthcoming.

Matějka, F. and A. McKay (2015): Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, 105, 272−298.

McFadden, D. (1974): Conditional logit analysis of qualitative choice behavior. In P. Zarembka (Ed.), *Frontiers in Econometrics*, 105−142. New York: Academic Press.

McFadden, D. and P. Robles (2019): Instability in mixed logit demand models for analysis of concentrated markets. Schaeffer Center Working Paper, USC.

McFadden, D. and K. Train (2000): Mixed MNL models for discrete response. *Journal of Applied Econometrics*, 15(5), 447−470.

McGarry, B. E., N. Maestas, and D. C. Grabowski (2018): Simplifying the Medicare Plan Finder tool could help older adults choose lower cost Part D plans. *Health Affairs*, 37(8), 1290−1297.

Miller, D. P. and J. Yeo (2012): Estimating dynamic discrete choice models of product differentiation: An application to Medicare Part D with switching costs. Unpublished manuscript.

Newhouse, J. P. and T. G. McGuire (2014): How successful is medicare advantage? *The Milbank Quarterly*, 92(2), 351−394.

Nosal, K. (2012): Estimating switching costs for Medicare Advantage plans. Unpublished manuscript, University of Mannheim.

Polyakova, M. (2016): Regulation of insurance with adverse selection and switching costs: Evidence from Medicare Part D. *American Economic Journal: Applied Economics*, 8(3), 165 −195.

Shcherbakov, O. (2016): Measuring consumer switching costs in the television industry. *RAND Journal of Economics*, 47(2), 366−393.

Simon, H. A. (1955): A behavioral model of rational choice. *Quarterly Journal of Economics*, 69(1), 99−118.

Sinaiko, A. D., C. C. Afendulis, and R. G. Frank (2013): Enrollment in Medicare Advantage plans in Miami-Dade County: Evidence of status quo bias? *Inquiry*, 50(3), 202–215.

Stango, V. and J. Zinman (2014): Limited and varying consumer attention: Evidence from shocks to the salience of bank overdraft fees. *Review of Finance*, 27(4), 990−1030.

Zhou, C. and Y. Zhang (2012): The vast majority of Medicare Part D beneficiaries still don't choose the cheapest plans that meet their medication needs. *Health Affairs*, 31(10), 2259−2265.

Tables and figures

Table 1: Sample selection

	Entire 20%	**Selection criteria**	**Estimation sample**
2007	9,300,230		
2008	9,530,809	1,082,665	74,848
2009	9,781,213	1,144,094	79,181
2010	10,016,372	1,158,315	80,426

Notes: Selection criteria: US residents, aged 65+, enrolled in stand-alone non-employer PDP plan, do not receive low-income subsidies (non-LIS), not dual-eligible, enrolled in Part D continuously in 12 months in the reference year (t) and prior year ($t-1$) – unless the individual has no default plan – and no missing data. Estimation sample randomly selects 100,000 individuals who meet all selection criteria in at least one year. Because of later enrollment, death, or not meeting all selection criteria for the whole time period, the panel data set is unbalanced and the number of observations is smaller than 100,000 in each year.

Table 2: Individuals with and without default plan

	With default plan			Without default plan	
	Stayers	**Switchers**	**All**		
	(%)	(%)	(%)	(%)	N
2008	86.59	10.16	96.75	3.25	74,848
2009	86.42	10.27	96.69	3.31	79,181
2010	88.05	9.44	97.49	2.51	80,426
All years	87.03	9.95	96.98	3.02	234,455

Notes: Individuals who do not have a default plan are new to Part D (roughly 97% of those without a default plan). The others do not have a default plan because their $t-1$ plan is terminated and there is no cross-walked other plan that they are defaulted to (roughly 3%).

Table 3: *Ex post* total costs and overspending

	With default plan			Without default plan	Total
	Stayers	**Switchers**	**All**		
Mean *ex post* total costs					
2008	1271.8	1328.1	1277.7	1241.2	1276.5
2009	1440.2	1273.5	1422.5	1257.8	1417
2010	1459.2	1296	1443.3	1294	1439.6
All years	1393.3	1298.6	1383.6	1262.4	1379.9
Mean *ex post* overspending					
2008	329.9	218.7	318.2	283.6	317.1
2009	397.4	203.9	376.8	260.4	373
2010	398.3	206.7	379.8	255.2	376.6
All years	376.3	209.6	359.2	266.9	356.4

Notes: Based on estimation sample. *Ex post* total costs are costs that consumers have realized in their chosen plan, i. e., costs based on prescription drugs taken in year t. *Ex post* overspending is the difference between *ex post* total costs in the plan that a consumer has chosen and the *ex post* total costs that a consumer would have realized in the least-cost alternative plan.

Table 4: *Ex ante* total costs and overspending in default and chosen plans

	Stayers	**Switchers**	
	Default plan = chosen plan	Default plan	Chosen plan
Mean *ex ante* total costs			
2008	1315.7	1591.1	1417
2009	1428.7	1502.9	1303
2010	1470	1509.8	1331.4
All years	1407.1	1533.9	1349.4
Mean *ex ante* overspending			
2008	354.8	424.2	250.1
2009	409.7	423.9	223.9
2010	425.4	409.3	230.9
All years	397.7	419.2	234.7
Median *ex ante* overspending			
2008	274.3	325.2	179.2
2009	341.5	357.6	158.8
2010	347.6	316.4	171.8
All years	323.6	335.3	170.1
Share with zero *ex ante* overspending			
2008	3.39	1.05	13.21
2009	1.95	0.89	17.31
2010	2.65	1.17	17.04
All years	2.65	1.03	15.88

Notes: Based on estimation sample. *Ex ante* total costs are costs in (default or chosen) plan in t, calculated based on each consumer's prescription drug use in $t-1$. *Ex ante* overspending is the difference between *ex ante* total costs in (default or chosen) plan and *ex ante* total costs in the least-cost alternative. As *ex ante* overspending is not available for most beneficiaries without a default, they are not included in this table.

Table 5: Model specifications

		Model				
		I	**II**	**III**	**IV**	**V**
Behavioral component	**Parameters**					
Plan choice with switching costs	$\boldsymbol{\beta}^0, \gamma^0, \sigma_\gamma^2$	✓	✓	✓	✓	✓
Attention stage	$\boldsymbol{\zeta}$		✓	✓	✓	✓
Acuity (latent) in attention stage	$\boldsymbol{\alpha}$		✓	✓	✓	✓
Unobserved heterogeneity in acuity	σ_c			✓	✓	✓
Acuity in choice stage via error variance	η				✓	✓
Acuity in choice stage via interactions	$\boldsymbol{\beta}^q, \gamma^q$					✓

Notes: Parameters not included in a model are restricted to zero. The vectors $\boldsymbol{\beta}^0$ and $\boldsymbol{\beta}^q$ collect the coefficients of the plan characteristics as defined in equation (5).

Table 6: Descriptive statistics – estimation sample

	With default plan						**Without default plan**	
	Switchers		**Stayers**		**All**			
Attention triggers								
Part D experience in $t-1$								
Gap coverage [D]	0.15	(0.35)	0.13	(0.34)	0.13	(0.34)		
Hit the gap [D]	0.23	(0.42)	0.18	(0.39)	0.19	(0.39)		
Gap coverage & hit the gap [D]	0.06	(0.23)	0.05	(0.21)	0.05	(0.21)		
Hit catastrophic region [D]	0.03	(0.17)	0.03	(0.16)	0.03	(0.16)		
Changes in features of the $t-1$ plan								
Plan consolidated [D]	0.1	(0.31)	0.19	(0.39)	0.18	(0.38)		
Change of premium [$1000]	0.13	(0.11)	0.07	(0.08)	0.08	(0.09)		
Change of deductible [$1000]	0.01	(0.05)	00	(0.05)	00	(0.05)		
Change of ICL [$1000]	0.15	(0.05)	0.15	(0.05)	0.15	(0.05)		
OOP cost effect of formulary change [$1000]	0.05	(0.17)	0.04	(0.16)	0.05	(0.16)		
Tiers with increases in copayments [share]	0.38	(0.35)	0.42	(0.3)	0.41	(0.31)		
Switch between copay and coinsurance [D]	0.15	(0.36)	0.08	(0.26)	0.08	(0.28)		
Health shocks and health care use in $t-1$								
Onset of costly condition [D]	0.11	(0.31)	0.11	(0.32)	0.11	(0.32)		
Onset of cheap condition [D]	0.29	(0.46)	0.3	(0.46)	0.3	(0.46)		
Five or more doctor visits [D]	0.79	(0.41)	0.78	(0.41)	0.78	(0.41)		
At least one ER visit [D]	0.25	(0.43)	0.26	(0.44)	0.25	(0.44)		
At least one hospital stay [D]	0.17	(0.38)	0.18	(0.38)	0.17	(0.38)		
Determinants of acuity								
Male [D]	0.35	(0.48)	0.37	(0.48)	0.37	(0.48)	0.46	(0.5)
Non-white [D]	0.05	(0.22)	0.07	(0.26)	0.07	(0.25)	0.09	(0.28)
Age 70–79 years [D]	0.46	(0.5)	0.46	(0.5)	0.46	(0.5)	0.43	(0.49)
Age 80 years or older [D]	0.28	(0.45)	0.31	(0.46)	0.30	(0.46)	0.27	(0.44)
PDP experience [years]	2.75	(0.87)	2.8	(0.88)	2.8	(0.88)	0.24	(0.78)
Low education [ZIP-code share]	0.21	(0.11)	0.21	(0.12)	0.21	(0.12)	0.21	(0.12)
Low income [ZIP-code share]	0.09	(0.07)	0.09	(0.07)	0.09	(0.07)	0.08	(0.07)
High income [ZIP-code share]	0.24	(0.16)	0.25	(0.16)	0.25	(0.16)	0.25	(0.16)
Ever had depression [D]	0.19	(0.39)	0.19	(0.39)	0.19	(0.39)	0.15	(0.36)
N	23,331		204,053		227,384		7,071	

Notes: Means with standard deviations in parentheses. Dummy variables are marked with [D]. Pooled across years 2007/8-2009/10. Information on experience in $t-1$ plan for beneficiaries without a default plan omitted as only available for small fraction among them (those whose $t-1$ plan was terminated). Change in cost sharing is defined as the share of tiers on the formulary for which an increase in cost-sharing occurs. Dollar amounts are measured in 2010 dollars using the CPI.

Table 7: Descriptive statistics – choice sets in estimation sample

	With default plan						Without default plan	
	Switchers		Stayers		All			
Number of plans	51.75	(4.33)	51.12	(3.72)	51.18	(3.80)	51.59	(3.79)
OOP cost [$1000]	1.04	(1.17)	0.96	(1.08)	0.97	(1.09)	0.94	(1.14)
Plan characteristics								
Annual premium [$1000]	0.55	(0.04)	0.55	(0.04)	0.55	(0.04)	0.54	(0.04)
Deductible amount [$1000]	0.13	(0.02)	0.13	(0.02)	0.13	(0.02)	0.12	(0.02)
ICL amount [$1000]	2.75	(0.10)	2.75	(0.10)	2.75	(0.10)	2.74	(0.10)
No gap coverage [D]	0.76	(0.04)	0.76	(0.04)	0.76	(0.04)	0.76	(0.04)
Default plan [D]	0.02	(0.00)	0.02	(0.00)	0.02	(0.00)	0.00	(0.00)
Variance of OOP cost [($1000)^2]	10.31	(5.87)	10.03	(5.60)	10.06	(5.63)	9.74	(5.86)
N	23,331		204,053		227,384		7,071	

Notes: Means across averages in individuals' choice sets with standard deviations in parentheses. Dummy variables are marked with [D]. Pooled across years 2007/8-2009/10. Default plan is a dummy variable for the $t-1$ plan or the cross-walked new plan if the $t-1$ plan was discontinued. Dollar amounts are measured in 2010 dollars using the CPI.

Table 8: Estimated models

	Model				
	I	**II**	**III**	**IV**	**V**
Number of parameters	13	41	42	43	55
Log likelihood	-162,598.6	-158,564.9	-157,617.0	-157,110.7	-154,617.8
LR test against model V (df)	15961.6 (42)	7894.2 (14)	5998.4 (13)	4985.8 (12)	—
AIC	325,223.2	317,211.8	315,318.0	314,307.4	309,345.6
BIC	325,357.9	317,636.8	315,753.3	314,753.1	309,915.7

Notes: All models are estimated on the unbalanced estimation sample with $234,455$ observations on $100,000$ individuals. On average, they faced around 51.2 alternatives in each choice situation which translates into 12,003,258 alternatives in total.

Table 9: Attention stage and determinants of acuity (Model V)

	Coefficient	SE	Average marginal effects on attention	on switching
Attention stage ($\boldsymbol{\zeta}$)				
Constant	-1.647	(0.080)		
Part D experience in $t-1$				
Gap coverage [D]	-0.851	(0.045)	-0.138	-0.044
Hit the gap [D]	0.773	(0.045)	0.148	0.042
Gap coverage & hit the gap [D]	-0.393	(0.066)	-0.067	-0.020
Hit catastrophic region [D]	0.213	(0.074)	0.039	0.011
Changes in features of the $t-1$ plan				
Plan consolidated [D]	0.214	(0.045)	0.039	0.011
Change of premium [$1000]	5.756	(0.120)	1.041	0.307
Change of deductible [$1000]	2.396	(0.252)	0.433	0.128
Change of ICL [$1000]	0.426	(0.309)	0.077	0.023
OOP cost effect of formulary change [$1000]	0.498	(0.073)	0.090	0.027
Tiers with increases in copayments [share]	-0.093	(0.043)	-0.017	-0.005
Switch between copay and coinsurance [D]	0.693	(0.045)	0.135	0.038
Health shocks and health care use in $t-1$				
Onset of costly condition [D]	-0.044	(0.039)	-0.008	-0.002
Onset of cheap condition [D]	-0.071	(0.028)	-0.013	-0.004
Five or more doctor visits [D]	0.190	(0.032)	0.034	0.010
At least one ER visit [D]	-0.054	(0.033)	-0.010	-0.003
At least one hospital stay [D]	-0.062	(0.038)	-0.011	-0.003
Determinants of acuity ($\boldsymbol{\alpha}$)				
Male [D]	-0.030	(0.008)	-0.005	-0.002
Non-white [D]	-0.171	(0.021)	-0.030	-0.009
Age 70–79 years [D]	-0.065	(0.010)	-0.012	-0.003
Age 80 years or older [D]	-0.163	(0.016)	-0.029	-0.009
PDP experience [years]	-0.044	(0.005)	-0.008	-0.002
Low education [ZIP-code share]	-0.361	(0.049)	-0.065	-0.019
Low income [ZIP-code share]	-0.139	(0.065)	-0.025	-0.007
High income [ZIP-code share]	-0.498	(0.046)	-0.090	-0.027
Ever had depression [D]	-0.055	(0.010)	-0.010	-0.003
Standard deviation of unobserved acuity (σ_c)	0.669	(0.044)		

Notes: Average marginal effects on the probability of paying attention and the probability of switching plans. [D] indicates that variable is a dummy. The attention stage also includes year dummies.

Table 10: Plan-choice stage (Model V)

	Base coefficient	SE	Interaction coefficient	SE	Average marginal effects on choosing plan	
	$\boldsymbol{\beta}^0, \gamma^0, \sigma_\gamma^2, \eta$		$\boldsymbol{\beta}^q, \gamma^q$		unconditional	cond. on attention
OOP cost [$1000]	0.794	(0.021)	0.604	(0.031)	-0.011	-0.023
Annual premium [$1000]	1.699	(0.017)	0.663	(0.029)	-0.030	-0.059
Deductible amount [$1000]	-7.584	(0.396)	1.922	(0.688)	-0.018	-0.041
ICL amount [$1000]	6.730	(0.860)	-8.132	(1.137)	0.003	0.009
No gap coverage [D]	-0.577	(0.115)	1.000	(0.202)	0.000	0.000
Default plan [D]					0.761	0.304
Coefficient mean (γ^0, γ^q)	8.582	(0.223)	-5.551	(0.366)		
Coefficient variance (σ_γ^2)	1.557	(0.057)				
Variance of OOP cost [($1000)^2]	-0.317	(0.048)	0.042	(0.071)	-0.001	-0.002
Acuity effect in error variance (η)	0.041	(0.028)				

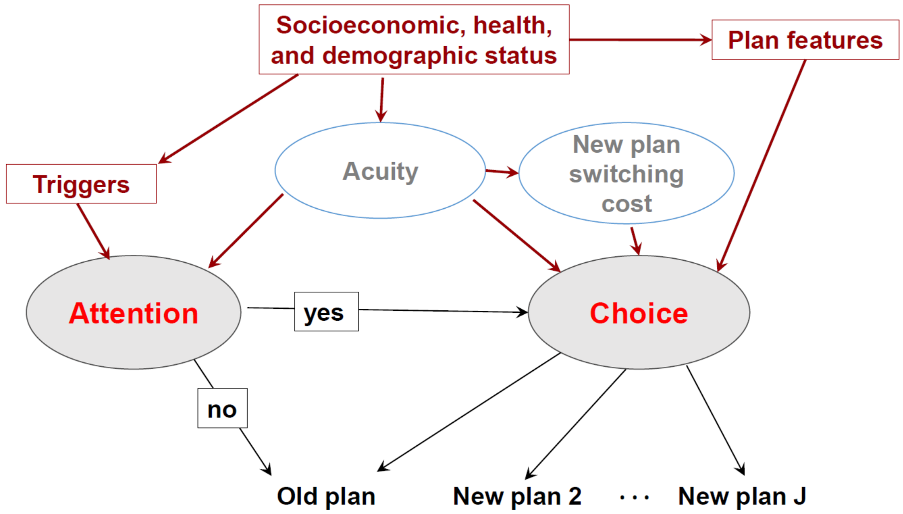
Notes: The base parameters $\boldsymbol{\beta}^0, \gamma^0, \sigma_\gamma^2$ denote the non-interacted components of the coefficients of plan-characteristics; the interaction parameters $\boldsymbol{\beta}^q, \gamma^q$ scale the acuity interaction components. The unconditional average marginal effect of the default plan dummy measures inertia. The plan-choice stage also includes five dummy variables for the largest plans.

Table 11: Quantiles of marginal effects in the plan-choice stage (Model V)

	Unconditional			**Cond. on attention**		
	5%	50%	95%	5%	50%	95%
OOP cost [$1000]	-0.0496	-0.0014	-0.0000	-0.1016	-0.0040	-0.0001
Annual premium [$1000]	-0.1302	-0.0035	-0.0000	-0.2660	-0.0101	-0.0002
Deductible amount [$1000]	-0.0838	-0.0030	-0.0001	-0.1762	-0.0093	-0.0003
ICL amount [$1000]	-0.0015	0.0003	0.0125	-0.0019	0.0015	0.0339
No gap coverage [D]	-0.0003	0.0000	0.0018	-0.0011	0.0000	0.0034
Default plan [D]	0.5154	0.7823	0.9371	0.0287	0.2674	0.7031
Variance of OOP cost [($1000)^2]	-0.0039	-0.0001	-0.0000	-0.0081	-0.0004	-0.0000

Notes: 5% percentile, median and 95% percentile of marginal effects on the probability that plan is chosen, unconditional and conditional on attention.

Figure 1: Structure of the two-stage model



Notes: This figure shows the structure of the full two-stage model. We refer to this as Model V below. Model I–IV are restricted versions of the full model.

Figure 2: Acuity and attention probability (Model V)

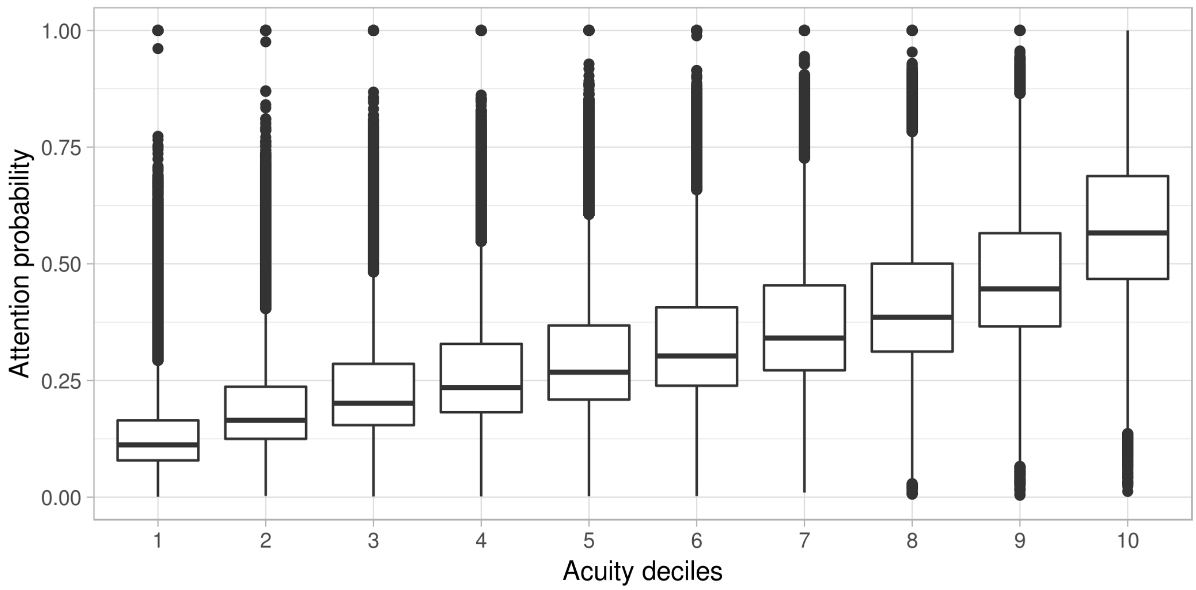
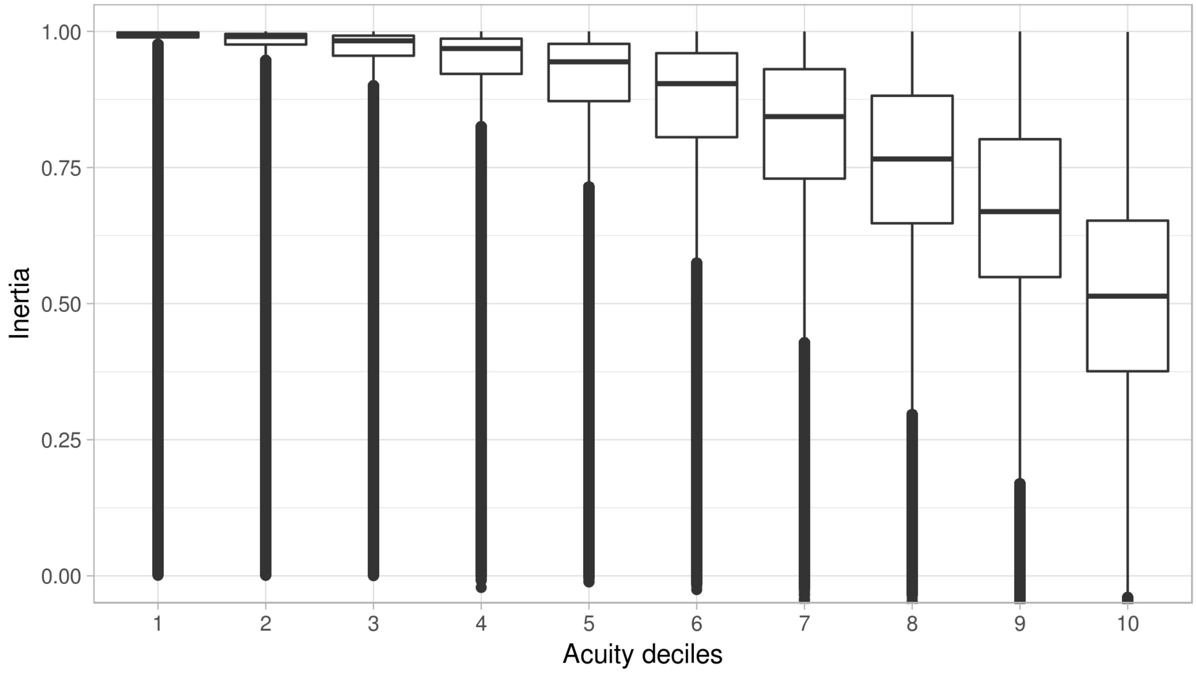
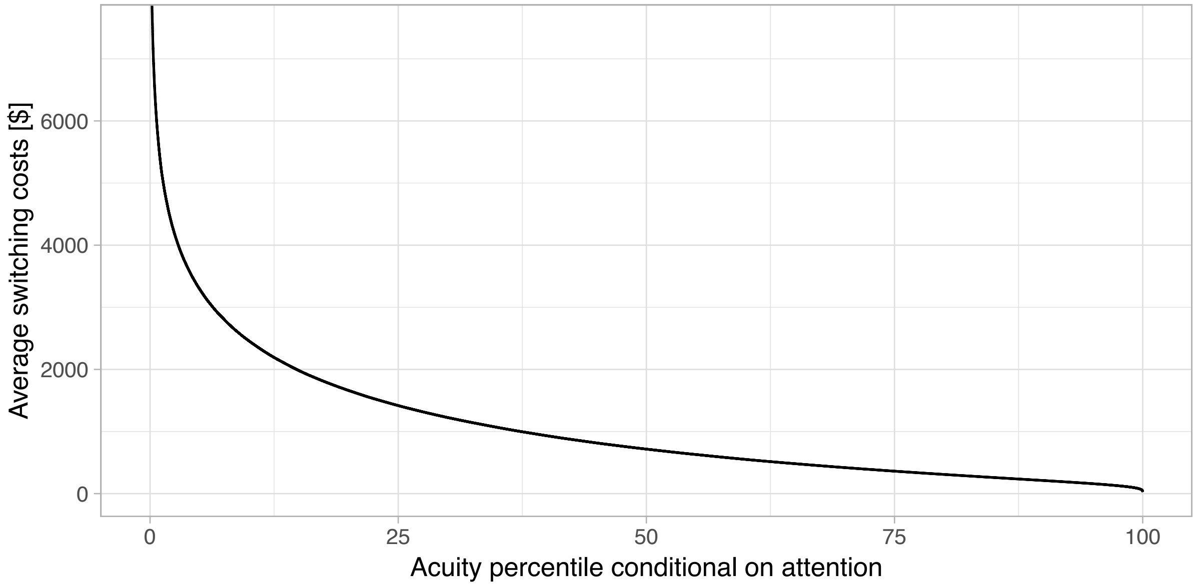


Figure 3: Acuity and inertia (Model V)



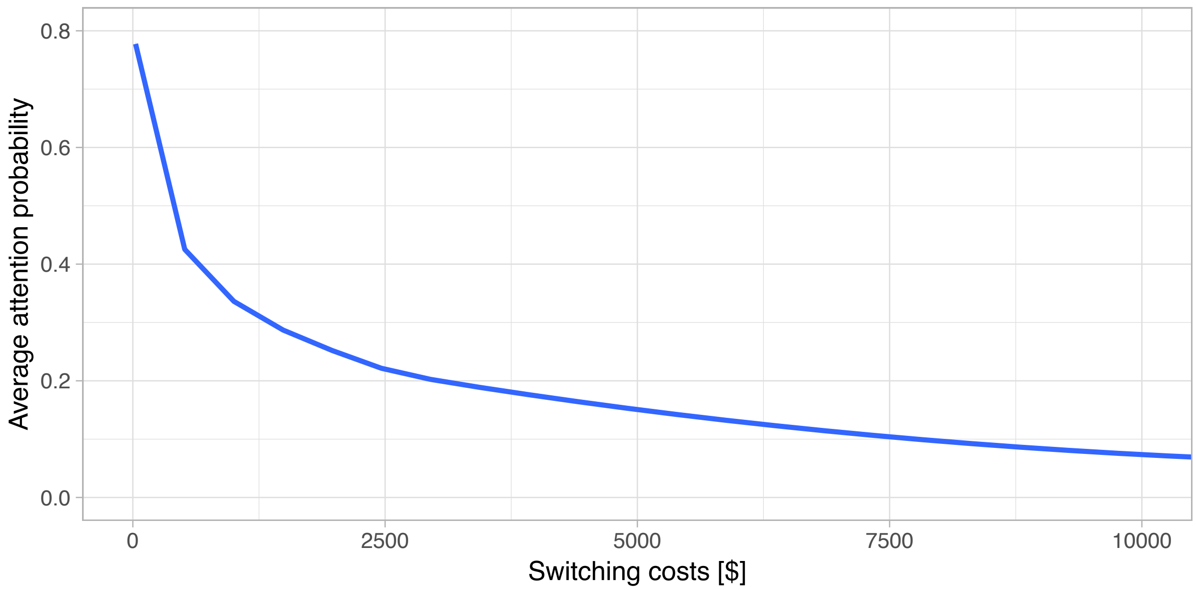
Notes: Inertia is the unconditional marginal effect of the default plan dummy. It reflects the difference between the probabilities of staying in the $t-1$ plan and choosing an otherwise identical alternative.

Figure 4: Acuity and switching costs (Model V)



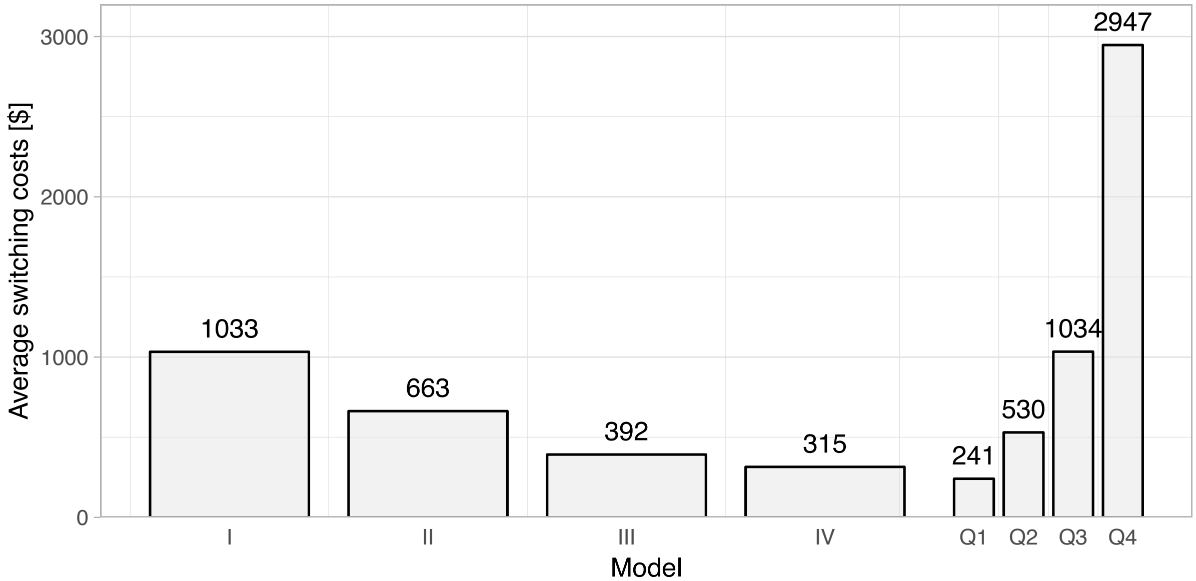
Notes: Switching cost are calculated as the ratio of the random coefficient and the coefficient of the premium, see equation (15). The figure shows the conditional average of switching costs w.r.t. acuity.

Figure 5: Switching costs and attention probability (Model V)



Notes: The figure shows the average attention probability as a function of switching costs, calculated as the ratio of the random coefficient and the coefficient of the premium, see equation (15).

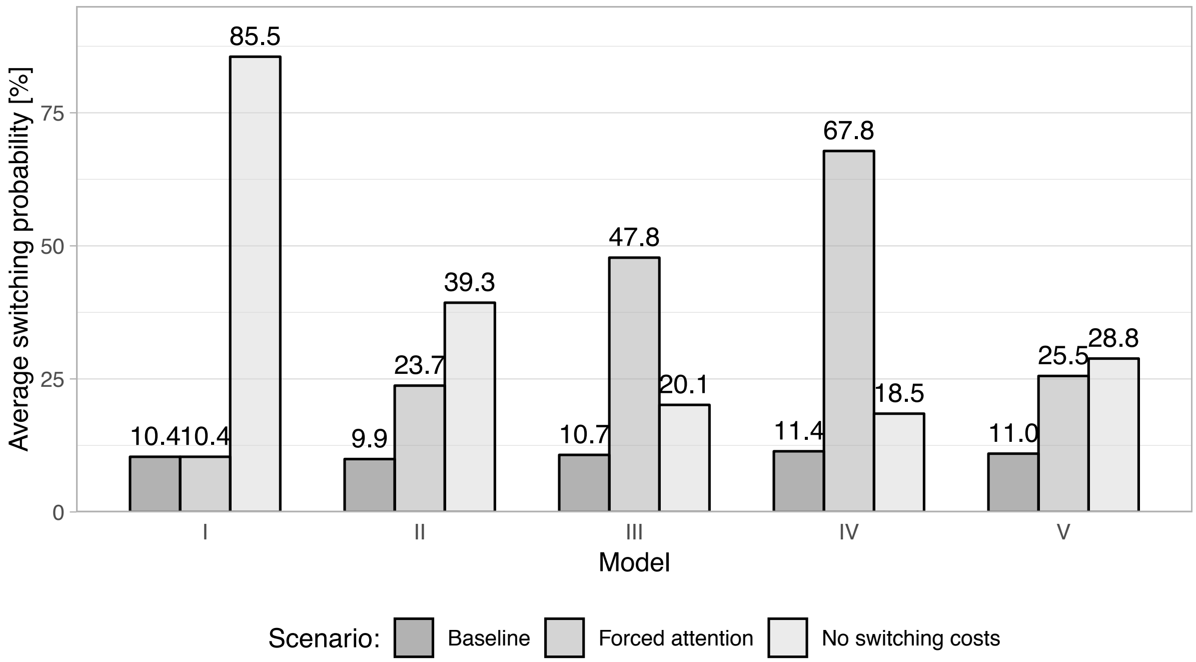
Figure 6: Comparison of switching costs across Models I–V



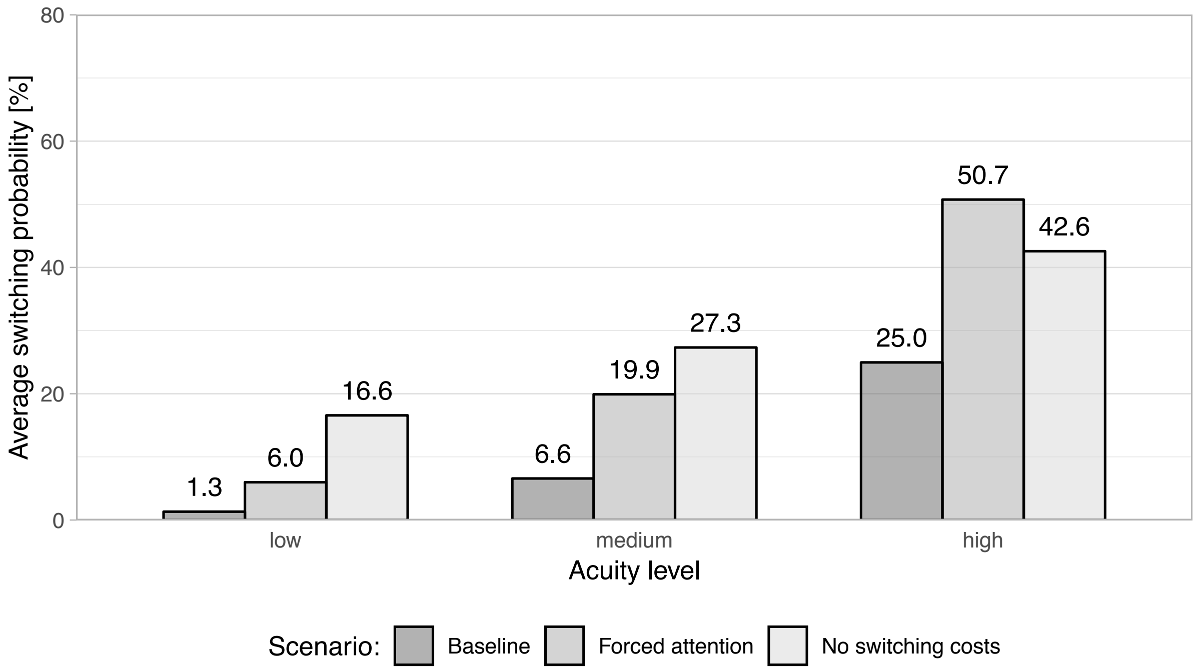
Notes: This graph shows the average switching cost estimate of Models I–IV. For Model V, which allows for heterogeneity in switching costs with respect to acuity, the figure shows the average within each quartile of the resulting switching cost distribution.

Figure 7: Counterfactual simulations: switching probabilities

(a) Comparison of Models I–V



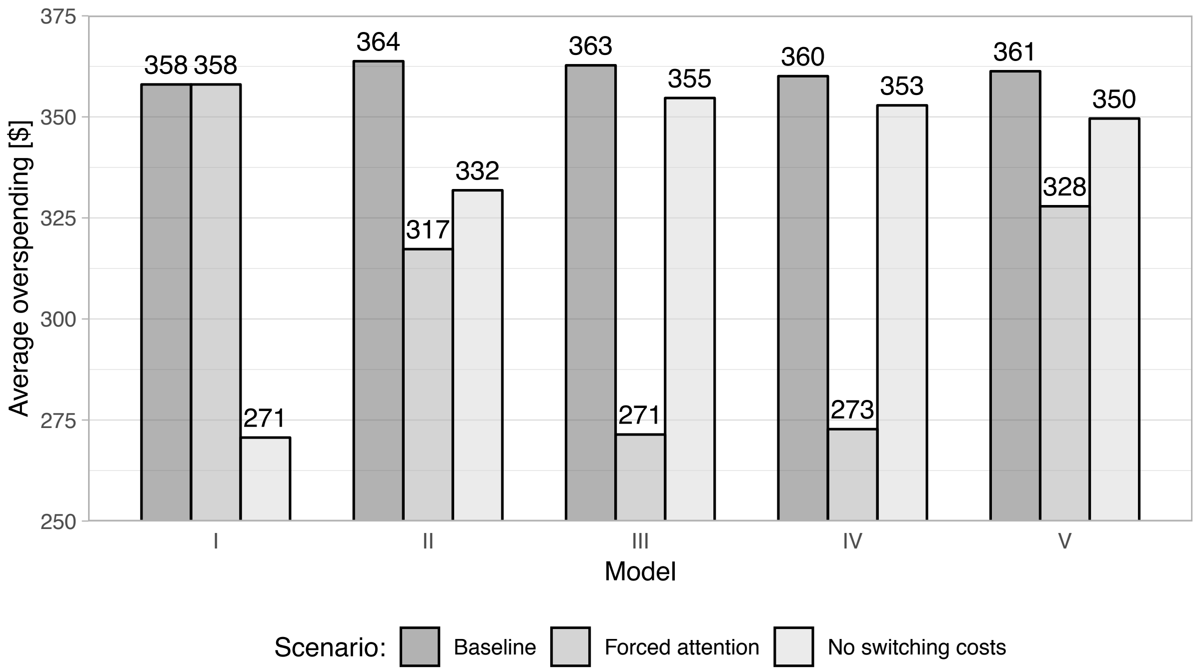
(b) Counterfactual simulations by acuity (Model V)



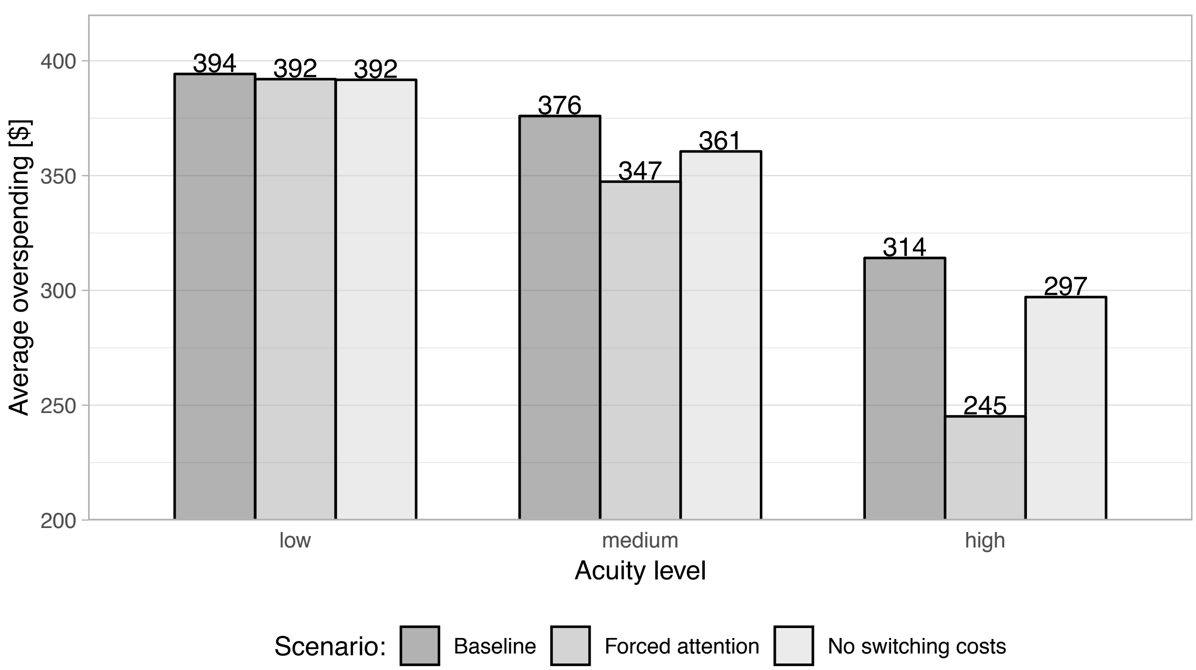
Notes: In the "forced attention" scenario, the probability of attention is set to 1 for everyone; in the "no switching costs" scenario, switching costs are set to zero for everyone. Acuity levels represent terciles of acuity.

Figure 8: Counterfactual simulations: average overspending

(a) Comparison of Models I–V

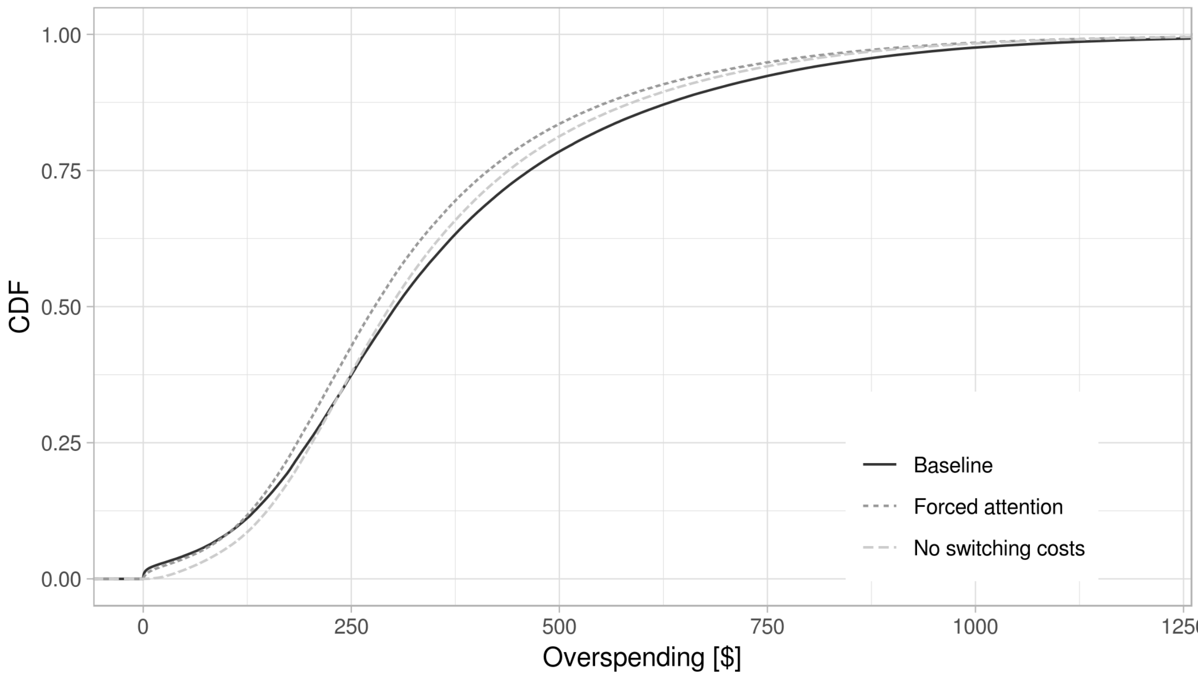


(b) Counterfactual simulations by acuity (Model V)



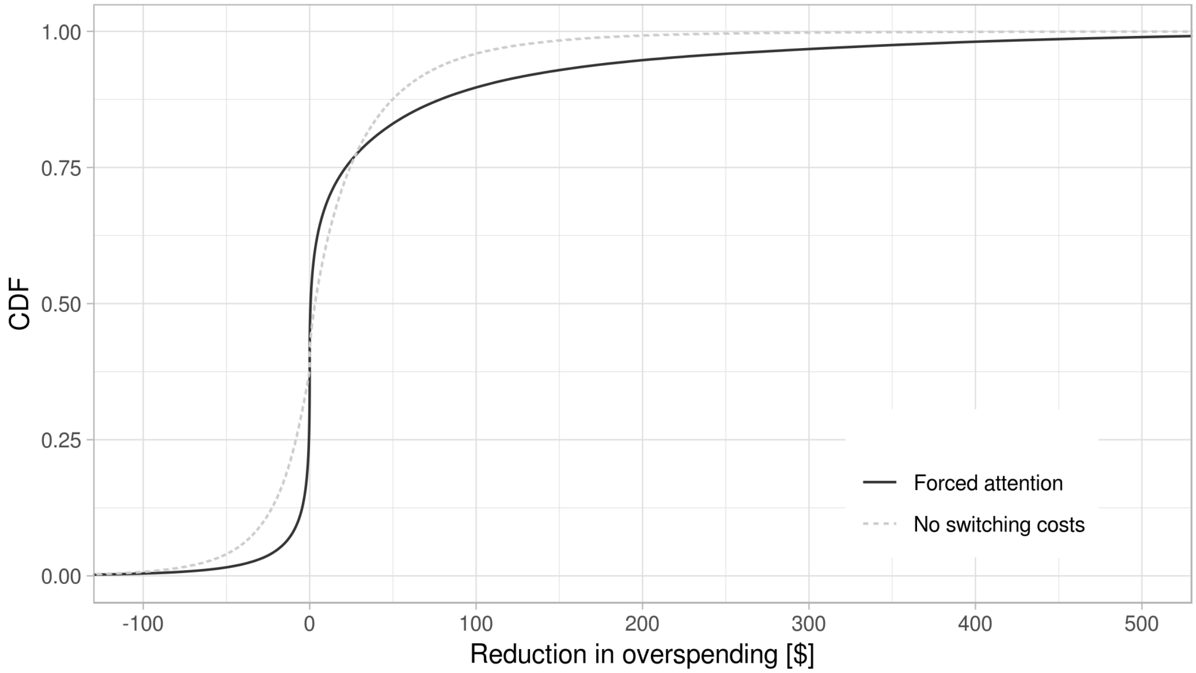
Notes: In the "forced attention" scenario, the probability of attention is set to 1 for everyone; in the "no switching costs" scenario, switching costs are set to zero for everyone. Acuity levels represent terciles of acuity.

Figure 9: Counterfactual simulations: distribution of overspending (Model V)



Notes: Empirical CDFs of simulated overspending based on Model V. In the "forced attention" scenario, the probability of attention is set to 1 for everyone; in the "no switching costs" scenario, switching costs are set to zero for everyone.

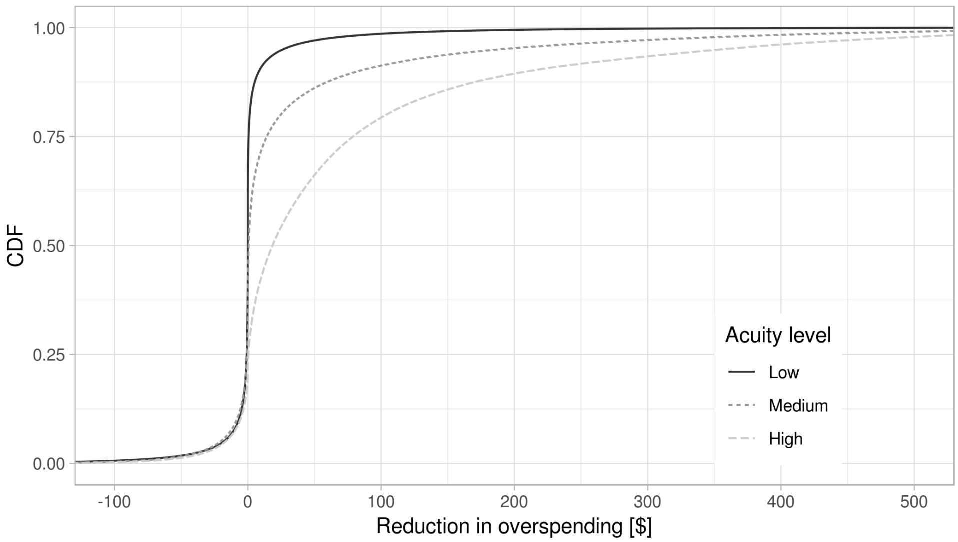
Figure 10: Counterfactual simulations: reduction in overspending (Model V)



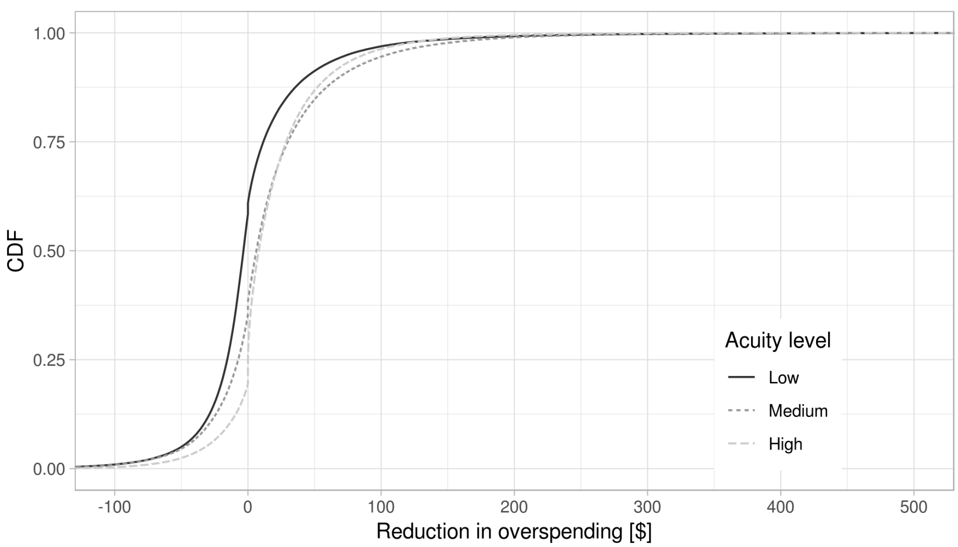
Notes: Empirical CDFs of simulated reduction in overspending (i. e., overspending at baseline – counterfactual overspending) based on Model V. In the "forced attention" scenario, the probability of attention is set to 1 for everyone; in the "no switching costs" scenario, switching costs are set to zero for everyone.

Figure 11: Counterfactual simulations: reduction in overspending by acuity (Model V)

(a) Forced attention



(b) No switching costs



Notes: Empirical CDFs of simulated reduction in overspending (i. e., overspending at baseline – counterfactual overspending) based on Model V for terciles of acuity (low, medium, high). In the "forced attention" scenario, the probability of attention is set to 1 for everyone; in the "no switching costs" scenario, switching costs are set to zero for everyone.

Appendix

Table A1: Plan-choice stage (Model I)

	Base coefficient $\boldsymbol{\beta}^0, \gamma^0, \sigma^2_\gamma, \eta$	**SE**	**Average marginal effects on choosing plan**	
			unconditional	**cond. on attention**
OOP cost [$1000]	1.012	(0.009)	-0.009	-0.009
Annual premium [$1000]	2.050	(0.006)	-0.026	-0.026
Deductible amount [$1000]	-5.649	(0.052)	-0.019	-0.019
ICL amount [$1000]	1.221	(0.155)	0.004	0.004
No gap coverage [D]	-0.763	(0.030)	-0.003	-0.003
Default plan [D]			0.407	0.407
Coefficient mean (γ^0)	8.024	(0.024)		
Coefficient variance (σ^2_γ)	2.490	(0.013)		
Variance of OOP cost [($1000)^2]	-0.287	(0.010)	-0.001	-0.001
Acuity effect in error variance (η)	–	(–)		

Notes: The base parameters $\boldsymbol{\beta}^0, \gamma^0, \sigma^2_\gamma$ denote the non-interacted components of the coefficients of plan-characteristics; the model does not include interaction terms with acuity. The unconditional average marginal effect of the default plan dummy measures inertia. As Model I does not include an attention stage, unconditional marginal effects and marginal effects conditional on attention are identical. The plan-choice stage also includes five dummy variables for the largest plans.

Table A2: Attention stage and determinants of acuity (Model II)

			Average marginal effects	
	Coefficient	**SE**	**on attention**	**on switching**
Attention stage ($\boldsymbol{\zeta}$)				
Constant	-0.558	(0.062)		
Part D experience in $t-1$				
Gap coverage [D]	-1.809	(0.038)	-0.338	-0.085
Hit the gap [D]	0.546	(0.043)	0.118	0.027
Gap coverage & hit the gap [D]	-0.277	(0.059)	-0.059	-0.013
Hit catastrophic region [D]	0.014	(0.066)	0.003	0.001
Changes in features of the $t-1$ plan				
Plan consolidated [D]	0.450	(0.045)	0.097	0.022
Change of premium [$1000]	4.282	(0.105)	0.924	0.210
Change of deductible [$1000]	2.977	(0.279)	0.643	0.146
Change of ICL [$1000]	0.627	(0.329)	0.135	0.031
OOP cost effect of formulary change [$1000]	0.498	(0.068)	0.107	0.024
Tiers with increases in copayments [share]	-0.045	(0.041)	-0.010	-0.002
Switch between copay and coinsurance [D]	0.866	(0.044)	0.185	0.043
Health shocks and health care use in $t-1$				
Onset of costly condition [D]	-0.050	(0.038)	-0.011	-0.002
Onset of cheap condition [D]	-0.054	(0.027)	-0.012	-0.003
Five or more doctor visits [D]	0.130	(0.031)	0.028	0.006
At least one ER visit [D]	-0.050	(0.031)	-0.011	-0.002
At least one hospital stay [D]	-0.065	(0.037)	-0.014	-0.003
Determinants of acuity ($\boldsymbol{\alpha}$)				
Male [D]	-0.116	(0.026)	-0.025	-0.006
Non-white [D]	-0.541	(0.048)	-0.114	-0.026
Age 70–79 years [D]	-0.258	(0.033)	-0.055	-0.012
Age 80 years or older [D]	-0.520	(0.038)	-0.109	-0.024
PDP experience [years]	-0.035	(0.023)	-0.008	-0.002
Low education [ZIP-code share]	-0.835	(0.137)	-0.180	-0.041
Low income [ZIP-code share]	-0.767	(0.204)	-0.166	-0.038
High income [ZIP-code share]	-1.394	(0.098)	-0.301	-0.068
Ever had depression [D]	-0.032	(0.030)	-0.007	-0.002
Standard deviation of unobserved acuity (σ_c)	–	(–)		

Notes: Average marginal effects on the probability of paying attention and the probability of switching plans. [D] indicates that variable is a dummy. The attention stage also includes year dummies.

Table A3: Plan-choice stage (Model II)

	Base coefficient $\boldsymbol{\beta}^0, \gamma^0, \sigma_\gamma^2, \eta$	SE	Average marginal effects on choosing plan: unconditional	Average marginal effects on choosing plan: cond. on attention
OOP cost [$1000]	1.154	(0.010)	-0.010	-0.020
Annual premium [$1000]	2.230	(0.006)	-0.029	-0.060
Deductible amount [$1000]	-6.121	(0.059)	-0.019	-0.039
ICL amount [$1000]	1.202	(0.170)	0.004	0.008
No gap coverage [D]	-0.474	(0.035)	-0.002	-0.003
Default plan [D]			0.692	0.351
Coefficient mean (γ^0, γ^q)	6.170	(0.034)		
Coefficient variance (σ_γ^2)	-2.369	(0.027)		
Variance of OOP cost [($1000)^2]	-0.291	(0.011)	-0.001	-0.002
Acuity effect in error variance (η)	–	(–)		

Notes: The base parameters $\boldsymbol{\beta}^0, \gamma^0, \sigma_\gamma^2$ denote the non-interacted components of the coefficients of plan-characteristics; the model does not include interaction terms with acuity. The unconditional average marginal effect of the default plan dummy measures inertia. The plan-choice stage also includes five dummy variables for the largest plans.

Table A4: Attention stage and determinants of acuity (Model III)

	Coefficient	SE	Average marginal effects on attention	Average marginal effects on switching
Attention stage ($\boldsymbol{\zeta}$)				
Constant	-2.557	(0.067)		
Part D experience in $t-1$				
Gap coverage [D]	-1.224	(0.050)	-0.111	-0.057
Hit the gap [D]	0.503	(0.038)	0.055	0.027
Gap coverage & hit the gap [D]	-0.331	(0.068)	-0.033	-0.017
Hit catastrophic region [D]	0.043	(0.081)	0.005	0.002
Changes in features of the $t-1$ plan				
Plan consolidated [D]	0.148	(0.046)	0.016	0.008
Change of premium [$1000]	6.655	(0.116)	0.702	0.346
Change of deductible [$1000]	2.637	(0.230)	0.278	0.137
Change of ICL [$1000]	0.272	(0.332)	0.029	0.014
OOP cost effect of formulary change [$1000]	0.647	(0.075)	0.068	0.034
Tiers with increases in copayments [share]	-0.250	(0.044)	-0.026	-0.013
Switch between copay and coinsurance [D]	0.748	(0.042)	0.086	0.042
Health shocks and health care use in $t-1$				
Onset of costly condition [D]	-0.055	(0.040)	-0.006	-0.003
Onset of cheap condition [D]	-0.063	(0.029)	-0.007	-0.003
Five or more doctor visits [D]	0.105	(0.034)	0.011	0.005
At least one ER visit [D]	-0.055	(0.033)	-0.006	-0.003
At least one hospital stay [D]	-0.063	(0.039)	-0.007	-0.003
Determinants of acuity ($\boldsymbol{\alpha}$)				
Male [D]	-0.130	(0.030)	-0.013	-0.007
Non-white [D]	-0.629	(0.064)	-0.060	-0.030
Age 70–79 years [D]	-0.255	(0.036)	-0.026	-0.013
Age 80 years or older [D]	-0.594	(0.043)	-0.057	-0.028
PDP experience [years]	0.014	(0.024)	0.001	0.001
Low education [ZIP-code share]	-1.141	(0.158)	-0.120	-0.059
Low income [ZIP-code share]	-0.590	(0.233)	-0.062	-0.031
High income [ZIP-code share]	-1.411	(0.117)	-0.149	-0.073
Ever had depression [D]	-0.073	(0.036)	-0.008	-0.004
Standard deviation of unobserved acuity (σ_c)	1.920	(0.035)		

Notes: Average marginal effects on the probability of paying attention and the probability of switching plans. [D] indicates that variable is a dummy. The attention stage also includes year dummies.

Table A5: Plan-choice stage (Model III)

	Base coefficient $\boldsymbol{\beta}^0, \gamma^0, \sigma^2_\gamma, \eta$	SE	Average marginal effects on choosing plan: unconditional	Average marginal effects on choosing plan: cond. on attention
OOP cost [$1000]	1.197	(0.009)	-0.010	-0.039
Annual premium [$1000]	2.259	(0.006)	-0.029	-0.112
Deductible amount [$1000]	-5.855	(0.061)	-0.018	-0.068
ICL amount [$1000]	1.090	(0.172)	0.003	0.013
No gap coverage [D]	-0.626	(0.036)	-0.002	-0.009
Default plan [D]			0.810	0.209
Coefficient mean (γ^0, γ^q)	3.750	(0.052)		
Coefficient variance (σ^2_γ)	1.272	(0.062)		
Variance of OOP cost [($1000)^2]	-0.275	(0.011)	-0.001	-0.003
Acuity effect in error variance (η)	–	(–)		

Notes: The base parameters $\boldsymbol{\beta}^0, \gamma^0, \sigma^2_\gamma$ denote the non-interacted components of the coefficients of plan-characteristics; the model does not include interaction terms with acuity. The unconditional average marginal effect of the default plan dummy measures inertia. The plan-choice stage also includes five dummy variables for the largest plans.

Table A6: Attention stage and determinants of acuity (Model IV)

	Coefficient	SE	Average marginal effects on attention	on switching
Attention stage ($\boldsymbol{\zeta}$)				
Constant	-3.165	(0.061)		
Part D experience in $t-1$				
Gap coverage [D]	-0.820	(0.044)	-0.068	-0.042
Hit the gap [D]	0.681	(0.034)	0.066	0.041
Gap coverage & hit the gap [D]	-0.490	(0.065)	-0.042	-0.026
Hit catastrophic region [D]	0.271	(0.072)	0.026	0.016
Changes in features of the $t-1$ plan				
Plan consolidated [D]	-0.088	(0.043)	-0.008	-0.005
Change of premium [$1000]	7.165	(0.109)	0.654	0.403
Change of deductible [$1000]	2.245	(0.216)	0.205	0.126
Change of ICL [$1000]	0.805	(0.326)	0.074	0.045
OOP cost effect of formulary change [$1000]	0.579	(0.070)	0.053	0.033
Tiers with increases in copayments [share]	-0.383	(0.042)	-0.035	-0.022
Switch between copay and coinsurance [D]	0.677	(0.038)	0.067	0.042
Health shocks and health care use in $t-1$				
Onset of costly condition [D]	-0.065	(0.039)	-0.006	-0.004
Onset of cheap condition [D]	-0.091	(0.027)	-0.008	-0.005
Five or more doctor visits [D]	0.124	(0.032)	0.011	0.007
At least one ER visit [D]	-0.019	(0.032)	-0.002	-0.001
At least one hospital stay [D]	-0.024	(0.037)	-0.002	-0.001
Determinants of acuity ($\boldsymbol{\alpha}$)				
Male [D]	-0.115	(0.033)	-0.010	-0.006
Non-white [D]	-0.652	(0.074)	-0.054	-0.033
Age 70–79 years [D]	-0.255	(0.038)	-0.022	-0.014
Age 80 years or older [D]	-0.716	(0.046)	-0.059	-0.036
PDP experience [years]	-0.051	(0.021)	-0.005	-0.003
Low education [ZIP-code share]	-1.347	(0.174)	-0.123	-0.076
Low income [ZIP-code share]	-0.491	(0.259)	-0.045	-0.028
High income [ZIP-code share]	-1.508	(0.128)	-0.138	-0.085
Ever had depression [D]	-0.216	(0.040)	-0.019	-0.012
Standard deviation of unobserved acuity (σ_c)	2.262	(0.039)		

Notes: Average marginal effects on the probability of paying attention and the probability of switching plans. [D] indicates that variable is a dummy. The attention stage also includes year dummies.

Table A7: Plan-choice stage (Model IV)

	Base coefficient $\boldsymbol{\beta}^0, \gamma^0, \sigma^2_\gamma, \eta$	SE	Average marginal effects on choosing plan	
			unconditional	**cond. on attention**
OOP cost [$1000]	0.985	(0.011)	-0.011	-0.041
Annual premium [$1000]	2.041	(0.009)	-0.031	-0.118
Deductible amount [$1000]	-4.340	(0.059)	-0.017	-0.066
ICL amount [$1000]	0.712	(0.132)	0.003	0.011
No gap coverage [D]	-0.473	(0.030)	-0.002	-0.009
Default plan [D]			0.814	0.129
Coefficient mean (γ^0)	2.421	(0.037)		
Coefficient variance (σ^2_γ)	0.979	(0.033)		
Variance of OOP cost [($1000)^2]	-0.188	(0.009)	-0.001	-0.003
Acuity effect in error variance (η)	-0.329	(0.007)		

Notes: The base parameters $\boldsymbol{\beta}^0, \gamma^0, \sigma^2_\gamma$ denote the non-interacted components of the coefficients of plan-characteristics; the model does not include interaction terms with acuity. The unconditional average marginal effect of the default plan dummy measures inertia. The plan-choice stage also includes five dummy variables for the largest plans.