Kesternich, Iris; Heiss, Florian; McFadden, Daniel und Winter, Joachim:

# Suit the action to the word, the word to the action: Hypothetical choices and real decisions in Medicare Part D

Munich Discussion Paper No. 2012-25

Department of Economics
University of Munich

Volkswirtschaftliche Fakultät
Ludwig-Maximilians-Universität München

Online at https://doi.org/10.5282/ubm/epub.14124

# Suit the action to the word, the word to the action: Hypothetical choices and real decisions in Medicare Part D

**Iris Kesternich**
University of Munich

**Florian Heiss**
University of Mainz

**Daniel McFadden**
University of California, Berkeley and University of Southern California

**Joachim Winter**
University of Munich

**October 2012**

**Abstract:** In recent years, consumer choice has become an important element of public policy. One reason is that consumers differ in their tastes and needs, which they can express most easily through their own choices. Elements that strengthen consumer choice feature prominently in the design of public insurance markets, for instance in the United States in the recent introduction of prescription drug coverage for older individuals via Medicare Part D. For policy makers who design such a market, an important practical question in the design phase of such a new program is how to deduce enrollment and plan selection preferences prior to its introduction. In this paper, we investigate whether hypothetical choice experiments can serve as a tool in this process. We combine data from hypothetical and real plan choices, elicited around the time of the introduction of Medicare Part D. We first analyze how well the hypothetical choice data predict willingness to pay and market shares at the aggregate level. We then analyze predictions at the individual level, in particular how insurance demand varies with observable characteristics. We also explore whether the extent of adverse selection can be predicted using hypothetical choice data alone.

**Keywords:** Medicare; health insurance demand; hypothetical choice experiments.

**JEL classification:** I11; C25; D12; H51; I18.

**Acknowledgements:** This research was supported 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, the Munich Center for Health Economics (MC-Health), and Deutsche Forschungsgemeinschaft (DFG) through SFB Transregio 15 and GRK 801. We thank Anna Gumpert, Arne Risa Hole, Paul Ruud, Martin Watzinger, participants of the IRDES Workshop in Applied Health Economics and the Health Econometrics Workshop (Rome), an anonymous referee as well as the Editor, Joseph P. Newhouse for helpful comments and suggestions. We thank Hendrik Brackmann and Sebastian Kohls for excellent research assistance.

# 1. Introduction

In recent years, consumer choice has become an important element of public policy. One reason is that consumers differ in their tastes and needs, which they can express most easily through their own choices. Elements that strengthen consumer choice feature prominently in the design of public insurance markets, for instance in the United States in the recent introduction of prescription drug coverage for older individuals via Medicare Part D (see, *inter alia*, McFadden et al. 2008).

For policy makers, an important practical question in the design phase of such a new program is how to deduce enrollment and plan preferences prior to its introduction. For instance, if plan sponsors are required to enroll any eligible individuals, as it is the case in Medicare Part D, will those individuals who are considered to be particularly vulnerable enroll? How severe will adverse selection be? In this paper, we investigate whether hypothetical choice experiments can serve as a tool to answer such questions in the design phase of a new program. We combine data from hypothetical and real plan choices, elicited around the time of the introduction of Medicare Part D. We first analyze how well the hypothetical choice data predict willingness to pay and market shares at the aggregate level. We then analyze predictions at the individual level, in particular how insurance demand varies with observable characteristics. We also explore whether the extent of adverse selection can be predicted using hypothetical choice data alone.

Consumer preferences for private goods provided within, or subsidized by, government programs can be deduced either from consumer behavior observed in relation to similar policies in the past or in different regions ("revealed preference"), or through "stated preference" methods. There are different ways of letting consumers state their preferences: They can be asked to directly rank or rate attributes; products can be auctioned; or consumers can be asked to choose between two or more alternatives that vary in their attributes. The latter approach is known as hypothetical choice experiment in which individuals are presented with the same type of choice situation they would be facing after the introduction of the policy. In the economics profession, there has long since been a discussion about whether stated preferences methods do accurately represent consumers' preferences. Arrow et al (1993), and Harrison and List (2004), and Bohm (2008) review stated preference methods and assess their scope and limitations. Our analysis exploits the introduction of Medicare Part D to inform this debate. Despite this being an old debate, it is still quite topical, for instance because of the design choices U.S. states face when it comes to the introduction of the health insurance exchanges mandated by the 2010 Patient Protection and Affordable Care Act (PPACA).

The hypothetical choice experiment on the demand for Medicare Part D plans we analyze was conducted in May 2006, just after the initial enrollment period for Part D had ended. The experiment was part of the Retirement Perspective Survey (RPS), an online survey of a representative sample of older Americans. Survey respondents were asked to choose between insurance contracts that differ in their level of coverage. They are also given the option not to have prescription drug coverage at all. The experiment was not designed to reflect the complexity of the enrolment decision (which we analyzed elsewhere, e.g. Winter et al., 2006; Heiss et al., 2009); rather the experiment concentrated on the desired level of plan generosity conditional on enrollment. The RPS also elicited detailed background variables, measures of prior year drug use, and data on Medicare Part D plan choices among those who could make an active decision

(Winter et al. 2006; Heiss et al. 2006). We thus can analyze hypothetical and real plan choices jointly, using a variety of discrete choice models that are tailored to the nature of the data and the substantive questions at hand.

We first analyze how well the hypothetical choice data predict willingness to pay and market shares at the aggregate level. We then analyze predictions at the individual level, in particular how insurance demand varies with observable characteristics. We also explore whether the extent of adverse selection can be predicted using hypothetical choice data alone. We report several interesting findings. Hypothetical and real data produce similar estimates of willingness-to-pay (WTP) for insurance plan attributes. Even though there are some differences in these estimates, our data suggest that hypothetical choice data can be used to predict demand on newly introduced insurance markets, perhaps with some refinements of the hypothetical choice tasks. Where we find differences between the estimates for the hypothetical and real choice data, they are likely due to a specific feature of the real market, namely, that not obtaining insurance is the default. We also find evidence for adverse selection in the analysis of both hypothetical and real choices, which suggests that hypothetical choice tasks might be used to predict the degree of adverse selection on newly introduced insurance markets as well.

The remainder of this paper is structured as follows. In section 2, we review the literature relevant to our analysis. We then describe the survey from which our data are taken and the design of the hypothetical choice experiment in section 3. The discrete-choice models of insurance plan demand and the results are described in section 4. Results are presented and discussed in Section 5. Section 6 concludes.

## 2. Related literature and implications for our study design

We begin by reviewing the literature on hypothetical choice experiments and their relation to the analysis of real decisions for the underlying products. We then present some results from the existing literature on the demand for insurance coverage in the new Medicare Part D market. We also relate the literature to the approach taken in our study.

### 2.1 Hypothetical choices and real decisions

Random utility theory has been the cornerstone of most empirical studies using observational or survey data to analyze consumer demand at the individual level (see McFadden, 2001, for a review of the approach and its history). In this framework, each product is defined as a bundle of characteristics; for example in the analysis of travel demand, each mode of travel is defined by the price and time of travel. Consumer preferences are defined over these characteristics, and the utility consumers obtain from buying a product depends both on the product's and their personal characteristics and tastes, which can only be partially observed by the researcher.

Consumer surveys provide data on the respondents' purchasing decisions in real markets. The preferences researchers deduce from this information are called revealed preferences because people reveal their preferences by making (and reporting) their choices. The advantage of deducing preferences from consumers' actual choices is that they spend real money facing their

actual budget constraints and, therefore, the researcher may be confident about making predictions based on econometric models using such data. However, making predictions from observed choices has some important limitations. First, in real markets the attributes of products often exhibit little variation. For example, in many markets, price varies very little over products as well as over time. In a statistical model, the researcher might thus wrongfully deduce that consumer decisions do not respond much to variations in price, when in reality prices do not vary much because price elasticity is very high. Second, when forecasting demand for new products, revealed preference data simply do not exist. The same holds true for certain public goods, which are not traded in the market. Third, choices might be observed for a selected group of consumers only. When drawing conclusions or making predictions, for example what happens if a certain insurance coverage is extended to a larger group of consumers, researchers may want to have variation not only in product, but also in consumer characteristics. With regard to Part D, there is another limitation to drawing conclusions from consumers' choices to their preferences. As it is an insurance market for the elderly, consumers might use choice aids, or they might be supported in making their choices by relatives and/or caregivers. In fact, in our data, of those consumers who remained uninsured, about 7.5 percent said that someone else had helped, but they had made the ultimate choice themselves; while 5.4 percent said that someone else made the choice for them. Among those who signed up for a Part D plan, these percentages are much higher: 29 percent versus 6.2 percent respectively.

To overcome these problems, individuals have been asked to state their preferences. Preferences elicited from this type of experiments are called stated preferences in contrast to consumers' revealed preferences from their real choices. Researchers can directly ask for WTP. They can also infer WTP from individuals' hypothetical choices. In hypothetical choice tasks, respondents are asked to choose among alternatives from a choice set whose attributes have been defined by the researcher. This allows the researcher to give estimations on the overall utility of the alternatives as well as their attributes

For example, in our data respondents are asked to choose between different insurance contracts whose prices vary randomly. Hypothetical choice experiments have been used extensively in public policy analysis, in marketing and psychology. Louviere et al. (2000) and Street and Burgess (2007) provide a comprehensive guide to the design, implementation and interpretation of stated choice methods. In these experiments, sufficient variation can be created. Additionally, the researcher can hold fixed everything in the choice situation that he wants to hold fixed, and concentrate only on the product characteristics that he is really interested in. This helps to overcome the well-known problem that prices are correlated with unobserved product quality, which leads to biases in the estimate of consumer price elasticity in real data.

As in every experiment, researchers are concerned with the external validity of stated choice experiments. Carson et al. (1996) and Murphy (2005) have performed a meta-analysis comparing WTP estimates from stated and revealed preference counterparts, and they find that the stated WTP is slightly lower than real WTP. Of course, many concerns regarding hypothetical choice experiments can be avoided by carefully designing them (Mitchell and Carson 1989, Arrow et al. 1993). As in every survey or experiment, wording, sequencing and context have to be designed carefully in order for respondents to understand the situation and in order to avoid biases (through, for example, strategic response behavior.)

Until now, stated preference experiments have mainly been evaluated in the laboratory or in other group settings such as classrooms or church halls, thus in a highly artificial situation with a selected group of respondents (Murphy et al. 2005). These evaluations were either done for public goods like the avoidance of environmental damage or contributions to a charity (see the papers cited in Murphy et al. 2003) where respondents are unfamiliar with stating willingness to pay and it is difficult to argue what "true" WTP should be. Or they have been tested for private goods with very low economics stakes, usually below $10 (examples include baseball cards (List et al. 2006), insurance (Balistreri et al. 2001) and all kinds of agricultural commodities like fish, fruit and meat (for example Alfnes et al. 2006, Chang et al. 2009, Lusk and Schroeder 2004). [1]

The design of our hypothetical choice experiment addresses some of the concerns that have been brought up in the literature, and it addresses a substantive question that is, arguably, of much more general interest than those studies in some of the earlier literature.

First, respondents might strategically overstate or understate their WTP, or respond in a way that they think is socially desirably or "politically correct" (Bohm 2008). We expect this to hold true more in the public good setting and not in the insurance setting that we will apply the hypothetical choice experiments to. Also, it is perhaps more salient to a respondent that by strategically biasing a reported WTP, she can influence public policy. In a hypothetical choice experiment such as ours, with a more complicated structure and WTP only being inferred form the choice data, it is not immediately clear how the respondent might bias her responses in order to affect a policy decision. In fact, choice-based methods have been found to lead to less bias than open-ended WTP elicitation formats (Murphy et al. 2005).

The second type of concern is related to the hypothetical nature of stated choice methods. The sample drawn for experiments might not be representative of the population, the situation presented in a laboratory might be artificial, and the choices presented might be unfamiliar to respondents (Murphy et al. 2005). For example, a sample of undergraduate students might not be a representative sample or the population for eliciting WTP. In private goods markets, individual self-select into the market; in the case of public goods, the policy maker has to decide on whom to include in the relevant market. The sample of respondents we use in our data is a random sample of the relevant population, and the insurance decision they are facing in the hypothetical choice task mirrors the decision they are facing in the real market.

### 2.2 Insurance plan demand in the Medicare Part D market

Since its introduction in 1965, Medicare provides health insurance for elderly and disabled Americans. In 2008, enrollment was at about 45 million.[2] Individuals are eligible for Medicare if they are U.S. citizens or long-term legal residents of at least 65 years of age and if either they or

[1] There is, of course, a fair number of choice experiments with stakes that were economically relevant (such as the study of student versus church samples in Blackburn, Harrison, Rutström, 1994), but these experiments compared hypothetical choices of different groups of individuals, not hypothetical and real choices.

[2] There are several websites which provide information on Medicare: www.cms.hhs.gov, www.medicare.gov and www.statehealthfacts.org.

their spouses have paid Medicare taxes for at least ten years.[3] Before the introduction of Medicare Part D in 2006, only pharmaceutical treatments administered in a physician's office, in a hospital or other institution were covered by the program. This was a major drawback since only a subset of Medicare beneficiaries had prescription drug coverage from some other sources, while about 30 percent of Medicare beneficiaries had little or no prescription drug coverage (Winter et al., 2006; Neuman et al., 2007). This had serious negative consequences. Medical expenditures placed a major financial burden on the elderly.[4] Moreover, cost-related non-adherence, i.e. the discontinuation of medication because it is too expensive, was a big concern before the introduction of Part D (Madden et al., 2008).

The aims of Medicare Part D were to make drug insurance coverage affordable for the elderly with low incomes, to provide protection against catastrophic drug costs and to reduce cost-related non-adherence.[5] The key feature of Part D is that the market is administered by the government, but private companies offer their products to consumers who choose contracts and carriers. Under Part D, consumers can choose between contracts providing standardized basic coverage that is subsidized by the government, or contracts that offer more extensive coverage at additional cost. The market is designed to increase the efficiency of the allocation of health care resources by confronting consumers with the full marginal cost of the services they use. Thus, Part D gives important insights into the practicality of Consumer Directed Health Care (CDHC) at least for the elderly. Further, as with the introduction of Medicare Part D, both contracts and prices available to consumers changed, Medicare Part D can act as a natural experiment of consumer behavior in real-world decision situations that are characterized by complexity, ambiguity and important consequences. A rapidly increasing body of research has drawn important lessons about the demand for prescription drug insurance, and the design of insurance markets more generally, from Medicare Part D.

Heiss et al. (2006) and Winter et al. (2006) find that, by and large, Medicare Part D has been a success in providing a large percentage of the Medicare eligible population with prescription drug coverage. Enrollment rates were above 90 percent in the first year of Medicare Part D. Those who remained without coverage in 2006 belong to two very different consumer groups: Those in relatively good health and those potentially difficult to reach. Furthermore, the complexity of the market with its many providers and many different products may have resulted in suboptimal choices, especially among the most vulnerable – those with low income, low educational attainment, poor health or some cognitive impairment. While most consumers have made rational decisions regarding the question whether they should enroll at all, they had some problems in deciding which plan was optimal for them.

Abaluck and Gruber (2011) find that elders' decisions depart from optimization under full information: They find that actual premiums are the main driver of consumers' choices, and that

[3] Further, disabled U.S. citizens or those with end stage renal disease are eligible for the program. However, we concentrate on the elderly beneficiaries here who form the vast majority of over 80 percent of beneficiaries.

[4] According to data from the Medical Expenditure Panel Survey, per-person expenditures among Medicare recipients for prescription drugs were equal to $1789 in 2003, with more than half of this paid out-of-pocket and just about 8 percent paid for by the Medicare program (Duggan and Scott-Morton, 2008).

[5] There exists a subsidy that recipients whose incomes are at or below 135 percent of the poverty line can apply for (the so-called low-income subsidy, or LIS). Estimations for 2008 show that 12.5 million Medicare beneficiaries are eligible for LIS, with 9.4 million actually receiving it. Certain groups of Medicare recipients are automatically enrolled in the subsidy, for example those on Medicaid (The Henry J. Kaiser Family Foundation, 2008).

too little weight is placed on expected out-of-pocket costs. Further, financial characteristics of a plan (for example, providing gap coverage or no deductible) are valued beyond any impacts on their own financial expenses or risk. They use a dataset of prescription drug claims matched to information on the characteristics of the choice set. A limitation of their data is that they cover only consumers with Part D stand-alone who moreover also filed a prescription drug claim. (In our study, we presented the hypothetical plan choice task to these consumers as well.) Abaluck and Gruber (2011) estimate a (monthly) WTP of $25 for full donut hole coverage, $4.17 for generic donut hole coverage, and $6.67 for going from a deductible of $250 to a deductible of zero.

Kling et al. (2012) test whether individuals make rational decisions regarding Part D – an environment with complex choices. They can reject the null hypothesis of choice stability and thus accurately perceived prices. Consumers who are provided with personalized information on how different drug plans affect their out-of-pocket costs make different decisions to consumers who are not provided with such information. The authors conclude that consumers had misperceived the influence of drug costs on prices before the intervention. Their results are based on a group of patients from a particular hospital, which raises the question in how far their results can be generalized.

Frakt and Pizer (2010) and Lucarelli et al. (2008) use aggregate data for demand estimation using the approach of Berry (1994). Both papers use the prescription drug plan finder to generate a dataset with region-plan pairs of stand-alone Part D plans. Frakt and Pizer (2009) estimate a premium elasticity of demand for the active deciders of (-1.45). Lucarelli et al. (2008) estimate that (monthly) WTP for eliminating the deductible is $3.83, while WTP for obtaining gap coverage of branded drugs is $36.92.

Overall, the introduction of Part D is seen as a success story in the sense that more seniors are now provided with affordable coverage. Some vulnerable groups, however, have not profited from the introduction. Further, adverse selection is driving those plans with extensive coverage out of the market. Our analysis of consumer decisions will shed some additional light on these findings

# 3. Data

In this section, we present the data we use in our empirical analysis. We first describe the Retirement Perspectives Survey (RPS), and then the hypothetical choice experiment that was part of RPS 2006.

## 3.1 The Retirement Perspectives Survey

The Retirement Perspectives Survey (RPS) was conducted in 2005, 2006, 2007 and 2009 in order to elicit information on enrollment decisions, knowledge, and opinions regarding Medicare Part D. The RPS questionnaires were administered over the internet to an age-restricted, random sample drawn from a panel of individuals maintained by Knowledge Networks, a commercial survey firm. These surveys collected information on prescription drug use, health conditions, socio-economic status, and household demographic composition. Winter et al. (2006) and Heiss

et al. (2011) provide detailed descriptions of the RPS design and data. In the present paper, we concentrate on the hypothetical choice experiment that was fielded as part of the second RPS wave in May 2006, just after the initial enrollment period for Medicare Part D had ended.

Despite being an internet survey, the RPS is fairly representative of the US non-institutionalized population in the relevant age group in terms of demographics and socio-economic status. This is because members of the KN Panel are recruited offline using random-digit dialing, and those who are willing to participate but not internet users are provided by KN with internet access, hardware, and training. Table 1 compares the RPS 2006 sample we analyze in the present paper to the 2006 Health and Retirement Study (HRS) in terms of socio-economic characteristics and insurance status. The RPS seems to reasonably mirror the HRS, even more in the weighted samples. A more detailed analysis of the RPS samples can be found in Heiss et al. (2011).

For our analysis, we are only interested in those individuals eligible for Medicare, therefore we restrict the sample to respondents aged 65 and older in 2006. Since one of the main aims of this paper is to combine and compare respondents' real and hypothetical choices, we restrict the working sample for the subsequent econometric analysis to those respondents who had to make an active decision whether to enroll in Part D, and which plan to select, in the real market. Thus, we exclude all respondents who already had prescription drug coverage before the introduction of Part D from other sources such as the previous employer (about 74 percent of the RPS respondents). In the remaining sample of respondents who had no prescription drug insurance before the introduction of Part D (the "active deciders"), many also remained without insurance for 2006. An important advantage of our design is that in the hypothetical choice task, respondents could also decide to obtain no insurance. Thus, the hypothetical task mimics the real-word enrollment and (conditional on enrollment) plan generosity decisions made by active deciders.

Definitions and descriptive statistics of the variables we use in our analysis can be found in Table 2. Most variables are constructed directly from responses to survey questions, with the exception of the variable "expected drug costs". This variable was created by Winter et al. (2006) based on respondents' self-reported prescription drug use and the full price they would have paid for them over the counter. This variable approximates the drug bill an individual would have to pay if she had no insurance. The dependent variables, hypothetical and real choices in the Medicare Part D market for insurance coverage, are described below.

### 3.2 The hypothetical choice experiment

In RPS 2006, a hypothetical choice experiment was conducted in order to elicit the preferences for prescription drug coverage. As the RPS focuses on questions on Medicare Part D and as all respondents in the RPS 2006 wave, fielded in May 2006, have already answered a (similar but not identical) questionnaire on Medicare Part D in RPS 2005, fielded in November 2005, we expect respondents to be familiar with questions on insurance and the Medicare Part D market when taking part in the hypothetical choice experiment.

Consumers were provided with a short introduction, in order to place our hypothetical choice experiment in the context of Part D. For instance, they were reminded that the same late enrollment penalties would apply in the hypothetical choice situations as stipulated in the real-

world Part D regulation. Also, the introduction text to the hypothetical choice experiment was adjusted slightly to reflect the fact that there are, for our purposes, three types of consumers; those who had purchased Part D stand-alone coverage for 2006 (who might be called "active deciders"), those who had coverage for 2006 from other sources (passive demand), and those who did not have coverage for 2006. The exact wording of the survey experiment can be found in the appendix of this paper. All respondents were told that they should consider plan choice starting from a situation with no prescription drug coverage at all.

The attributes of the hypothetical plans were chosen such that the hypothetical market has the same type of plans as the real one. Under Medicare Part D, the plans insurers can offer are standardized. The standard drug benefit, as defined by the Medicare Prescription Drug Improvement and Modernization Act of 2003, is characterized by four main features[6]:

- A $250 (annual) deductible below which the insured have to pay for all costs themselves.
- An interval of drug spending between $250 and $2,250 where the plan covers 75 percent of drug costs.
- A coverage gap between $2,250 and $5,100 where the insured has to bear the full costs.
- A ("catastrophic expenses") threshold of $5,100 above which the insurance covers 95 percent of all costs. Companies can either offer the standard plans, or plans that offer more extensive coverage, either by having no deductible or by providing coverage in the coverage gap

Respondents were given a choice between four alternatives with randomly varying premiums:

1. No prescription drug coverage.
2. The basic plan, which corresponds exactly to Part D basic coverage.
3. An enhanced plan which is just like the basic plan, but without deductible and
4. A premier plan which offers gap coverage in addition to having no deductible.

Each respondent was presented with three different hypothetical choice tasks. In the first round, everybody was presented with the same hypothetical prices, and in the second and third round, prices were randomly assigned to the respondents. Table 3 shows the range of premiums that were assigned to the different types of plans in our hypothetical choice experiment, and thus our hypothetical supply prices, along with the premiums that were available in the real market for Part D plans in 2006. For the first choice, premiums were the same for all respondents. These premiums closely resemble the premiums for all plans available in the market as constructed by Heiss et al. (2009) using public data provided by the Center for Medicare and Medicaid Services (CMS). Note that the premiums for the plans actually chosen by the RPS respondents, and therefore the prices in market equilibrium, are somewhat lower than supply prices, at least for basic and enhanced coverage. For the second and third choice, premiums were randomly assigned.[7]

[6] Features of the plan have changed slightly over time. These are the features of the plan in 2006, when our first hypothetical choice experiment was conducted.

[7] Table A.1 in the appendix contains the premium values for all 26 treatments.

In the real market, respondents' choice sets consist of all plans that are available in their region, i.e., between 38 and 52 plans. The insurance plans in the real market differ in more dimensions than the hypothetical market. The plan described in the hypothetical market is described as one that "covers all prescription drugs you currently need and most of what you might need in the future", while in reality, coverage might be more limited. Plans might only provide gap coverage for generics but not for brand name drugs. There might be drug tiers, or certain drugs might not be covered or only with prior authorization. Also, in the real market, Part D stand-alone plans comprise both those with Part D basic coverage and actuarily equivalent plans; these are combined into one category in our hypothetical plans.

We chose to leave some plan features unspecified in the hypothetical choice task to keep it simple. Thus, our choice scenario is incomplete in the sense of Manski (1999). The information respondents have on the hypothetical plans is a subset of the information they would have in the real choice setting. As long as respondents have rational expectations about the dimensions of the alternatives that are missing in the hypothetical task, we nevertheless obtain consistent estimates using the hypothetical choice data.

## 4. Econometric models

In this section, we discuss the econometric models we use to analyze the data from hypothetical and real choices. In addition to the substantive question of how these data can be used to analyze the demand on a newly introduced market, we also address the methodological issues of how flexible a discrete choice model should be when data from hypothetical and real choices are to be combined. Our modeling strategy addresses three issues in hypothetical choice experiments: First, we usually observe repeated choices of one decision maker. This can cause correlation in the unobserved parts of utility and thus contradict the assumptions of the multinomial logit model. Second, there might be heterogeneity in the valuation of attributes in the population over and above the heterogeneity caused by observed characteristics like income and risk. One possibility to take into account taste heterogeneity and the fact that we observe repeated choices are mixed logit models. Third, as in the multinomial logit framework, utility coefficients are only identified relative to the variance of unobserved factors, differences in the utility coefficients of hypothetical and real choices might be wrongfully attributed to differences in valuation of attributes, when in reality, it is just the variances that differ. We investigate this by allowing for different variances of the real and hypothetical choices.

The general framework is that of random utility maximization. Let the (indirect) utility that consumer $i$ obtains from the insurance contract $j$ be

$$U_{ij} = V(a_j, p_j, c_i, h_{ij}) + e_{ij}$$

where $a_j$ is a vector of attributes of insurance contract $j$; $p_j$ is the contract's premium; $c_i$ are the drug costs of the consumers in the year before choices were made (which are a good predictor of the drug costs in the year for which choices are made, see Heiss et al., 2012); $h_{ij}$ is a dummy variable that indicates whether the choices are observed in a hypothetical task or in the real market; and $e_{ij}$ is the error term that contains the impact of all unobservable factors.

The attributes of the insurance contracts we consider are (see also Table 2):

- **Drug tiers:** A dummy variable indicating whether the plan divides drugs into tiers with different levels of co-payments.
- **Insurance:** A dummy variable indicating whether the contract provides prescription drug coverage with copayments, deductible and gap coverage equal to the Part D standard benefit.
- **No deductible:** A dummy variable indicating whether the contract does not have the $250 deductible of the Part D benefit.
- **Gap coverage (generics):** A dummy variable indicating whether the contract additionally provides coverage in the coverage gap for generics.
- **Gap coverage (brand name drugs):** A dummy variable indicating whether the contract additionally provides coverage in the coverage gap for generics.
- **Premium:** The plan's monthly premium in dollars.
- **Top 100 drugs uncovered:** The number of top 100 drugs missing from the formulary.
- **Top 100 drugs with authorization:** The number of top 100 drugs only covered after authorization or step therapy.

We assume that the decision maker chooses the alternative with the highest utility $U_{ij}$. We first estimate a Multinomial Logit Model (MNL) (e.g. McFadden, 1976). In this model, observable explanatory variables enter via a linear index, while the unobservables, $e_{ij}$, are assumed to be i.i.d. random variables from an Extreme Value Type I distribution with density function

$$f(e_{ij}) = \exp(-e_{ij}) * \exp(-\exp(-e_{ij})).$$

In order to be able to take into account preference heterogeneity, but also the fact that we observe several choices by the same individual, we also estimate mixed logit models. These models allow for preference heterogeneity since the utility contributions of the different attributes may vary over decision makers. To obtain choice probabilities, which are straightforward in the standard MNL model, one has to integrate out the distribution of the coefficients (for a discussion of mixed logit models, see for example McFadden and Train, 2000, or Train, 2003). Note also that we observe three hypothetical choices per respondent, which will be reflected in a panel-like structure of the models.

Our main interest lies in estimating and predicting consumers' willingness to pay (WTP) for drug insurance with different levels of coverage. WTP is defined as the amount of premium increase that exactly offsets the increase of an attribute by one unit (or in the discrete case, the amount of premium that exactly offsets being provided with the discrete attribute versus not being provided with it), so that total utility remains unaffected. In the discrete choice models we consider, where product attributes and price (the premium) enter via a linear index, under quite general conditions willingness to pay for an attribute is given by the negative ratio of its coefficient and the coefficient of the premium:

$$WTP = -(\beta_{attribute}/\beta_{premium}).$$

Within this general framework, we estimate three classes of model specifications sequentially. We describe these specifications in the remainder of this section. In the following, we distinguish between parameters estimated using data on hypothetical and real choices by using the subscripts

SP (stated preferences) and RP (revealed preferences). The results (parameter estimates and implied WTPs) are presented in section 5 for each model and specification in turn; we also present extended specifications of these models to characterize adverse selection in this market.

### 4.1 Standard multinomial logit models, estimated with hypothetical and real decision data separately

We start our by estimating standard MNL models separately with data on hypothetical and real decisions. Thus, we allow the coefficients to be different in both types of data. Moreover, the coefficients are constant over decision makers. Finally, the standard errors are clustered by respondent in the stated preference model where three hypothetical choices are observed.

The coefficient estimates are reported in Table 4, specifications (1) and (2); the implied WTPs are reported in Table 5.

### 4.2 Standard multinomial logit models, estimated with combined data on hypothetical and real decisions

Next, we present models that combine stated and revealed preference data in a joint estimation. The simplest such model is the standard MNL model; results are reported in specifications (3) in Table 7; the implied WTPs are reported in Table 7. We extend these specifications in our analysis of adverse selection; those results are reported in specification (8) reported in Table 11.

### 4.3 Multinomial logit models, estimated with combined data on hypothetical and real decisions, allowing for differences in scale factors

Estimating discrete choice models requires some type of normalization because utility is a cardinal variable. In the case of the logit model, the variance is normalized to $\pi^2/6$. Thus, the estimated $\beta$s are estimates of the "real" $\beta$s, divided by $\lambda$ which is defined by the (unknown) variance $\sigma^2$ of the unobserved factors:

$$\sigma^2 = \lambda^2\pi^2/6.$$

In other words, the "true" $\beta$s cannot be identified separately from $\sigma^2$, and when we compare the coefficients from two data sources, we will never know whether differences result from differences in the true parameters or the variance of unobserved factors.

While the coefficients in respondents' utility functions should be the same in both types of data, unobserved factors will differ in stated and revealed preference situations. We would expect the real choices of respondents to be affected by many more unobserved factors than their stated choices. On the other hand, respondents face repeated choices in the hypothetical choice tasks, which could blow up the variances.

Therefore, when estimating a joint model of the stated and revealed preferences, we want to allow for different scale factors $\lambda_{SP}$ and $\lambda_{RP}$ (see Morikawa, 1989, and Louviere et al., 2000),

assuming that the true utility parameters are the same for the two data sets. As the scale factors are unobserved and can never be identified within one source of data, it is the convention to normalize $\lambda_{RP}$ to unity such that $\lambda_{SP}$ represents the stated preference relative to the revealed preference scale factor.

The first version of these models that we estimate imposes the restriction that $ß_{SP} / \lambda_{SP} = ß_{RP}$. As before, the coefficients are constant over decision makers, and the standard errors are clustered by respondent. Results are reported in Table 7, specification (4), and the implied WTPs are reported in Table 8; the extended version that allows us to characterize adverse selection is presented in specification (9), Table 11.

A second version eases the restriction that the utility coefficients are identical for hypothetical and real choices, which implies that $ß_{SP} / \lambda_{SP}$ and $ß_{RP}$ differ. Still, the coefficients are constant over decision makers, and the standard errors are clustered by respondent. Results are reported in Table 7, specification (5), and the implied WTPs are again reported in Table 8. We also present an adverse selection specification (10) in Table 11.

### 4.4 Mixed logit models, estimated with combined data on hypothetical and real decisions

The final class of models we estimate, again with combined data on hypothetical and real decisions, allows utility coefficients to vary between respondents but restricts them to be identical for hypothetical and real choices ($ß_{SP} = ß_{RP}$). Letting the coefficients vary between, but not within, the decision makers reflects the repeated nature of the choice tasks.

We specify two versions of mixed logit models. The first one, specification (6) in Table 9, assumes a normal mixing distribution for some of the parameters. Specifically, the coefficients for the attributes "insurance", "no deductible" and "gap coverage for brandname drugs" are assumed to be normally distributed among respondents. The second one assumes that the coefficient of the attribute "insurance" is normally distributed, while those of "no deductible" and "gap coverage for brandname drugs" are lognormally distributed among decision makers. Results for this second specification are reported in specification (7) in Table 9. WTPs for both specifications are reported in Table 10.

## 5. Results

Our sample contains 410 respondents. In the hypothetical market, each respondent faces 3 repeated choices between 4 contracts (including the choice of no insurance), with randomly varying premiums. In real market every decision maker has between 38 and 52 plans to choose from, depending on the plans offered in the respective state he lives in (including the choice of no insurance). Real contracts different in more dimensions: gap coverage can be generic, drug tiers, top 100 drugs uncovered, authorization.

Tables 4 and 5 allow a first glance at our data, with separate estimates of the hypothetical and real choices (Table 4) and the respective estimates of WTP (Table 5). The reported parameters are those of a linear indirect utility function, relative to the variance of the unobserved part of utility.

These parameters are smaller in hypothetical choices. All parameter estimates have the expected signs. Consumers place a positive valuation on having insurance, and their utility is even bigger when additionally they have no deductibles and gap coverage. The probability of choosing a plan decreases in the plan's premium and also in each top 100 drug which is not covered or only covered with authorization. The choice probability also decreases when the plan has several tiers with different copayments.

We cannot reject equality of the WTPs for joint attributes, except take-up (insurance). Interestingly, the estimated WTPs for not having a deductible are quite close to $250 (the upper bound of what consumers should be willing to pay) both in the hypothetical and in the real data. The annualized estimates are $177 for the real data (Table 5, column (1)) and $267 for the hypothetical data (Table 5, column (2)). The latter is large, but also not very precisely estimated. A possible explanation for the high WTP for avoiding a deductible is that the large majority of those consumers who have insurance coverage (82 percent) have drug costs that exceed the deductible. In the following, we discuss the differences and similarities between the hypothetical and real choices more systematically, in order to shed light on what can be learned from hypothetical data.

In Table 6, we show estimates and confidence intervals of the ratios of the WTPs obtained from the hypothetical and real decisions. We see that this ratio is significantly larger than one for the take-up of insurance (we will come back to this later). A ratio of 1 is within the confidence interval for the WTP ratios for "no deductible" and "gap coverage". However, at least for the deductible choice, the confidence interval for the ratio is quite large, ranging from hypothetical WTP being at about 2/3 of real WTP to hypothetical WTP being twice as large as real WTP.

In the remainder of this section, we consider the models that combine the data from hypothetical and real choices. The emphasis here is what can be learned using hypothetical and real choice data jointly to learn about consumers' valuation of the different insurance attributes. The choice structure and the product attributes can be controlled hypothetical choice experiment. However, these design choices also need to be reflected in the econometrics analysis. In the following, we discuss the following issues: defaults and implicitly defined attributes, differences in variances, the repeated nature of the hypothetical choice tasks and, related to this, taste heterogeneity.

An important aspect of the Medicare Part D enrollment and plan selection process is the default option – all eligible individuals without prior coverage, the active deciders, who do not take a decision, remain uninsured. In our data, in the real market 94 of our 410 respondents remained uninsured. In our hypothetical choice task, we did not implement this default. Rather, all respondents had to pick, in each of the three choice tasks, one of four alternatives, which included remaining uninsured. Thus, in our hypothetical choice experiment, the choice between different insurance contracts should be better reflected than the enrollment process per se.

Table 7 shows the coefficient estimates obtained from models of the combined hypothetical and real choice data. Model 3 imposes equality of both unobserved variances and coefficients between the two types of data. Significance and direction of the coefficients are like in the separate choices; to interpret their magnitudes, it helps looking at WTP (Table 8). WTP for the joint model again shows sensible magnitudes in general. While WTP for insurance *per se* is

estimated to be quite low due to low take-up when only analyzing real choices, the joint model predicts higher take-up.

Next, we want to analyze differences in the variance of unobserved variables between real and hypothetical choices. We estimate a multinomial logit model in which the variance of the error term in the latent utility model differs by data source. The estimated heteroskedasticity parameter is significantly different from zero, so the variances in the hypothetical and real choices appear to differ. We estimate a relative scale factor of exp(heteroskedasticity parameter) = exp(-0.76) = 0.47. This means that the ratio of the variance from hypothetical to real choice is $1/\exp(-0.76)^2 = 4.57$.

Note that the estimated WTPs are different in the hypothetical and real data. The different roles that the "no insurance" option plays – it is an active choice in the hypothetical and the default in real decisions – can explain the rejection of equality of the WTPs for takeup of insurance. We return to this issue below.

Model (4) allows not only for variance differences between the hypothetical and real choices, but it also allows the utility coefficients to vary between these two type of data sources. Thus, this model helps us to analyze whether up and above the differences in variances in the two data sets, there are also differences in the estimated parameters (and thus WTP). First note that all estimated coefficients from the hypothetical and real choices have the same signs. There are no significant differences between the coefficients (and therefore WTP) are not significantly for the no deductible and gap coverage option, but significantly higher in the hypothetical choices regarding the insurance option. A potential explanation is that the real world, the default option was to remain uninsured, while in our hypothetical choices one has to actively decide against insurance. Thus, in our specific case, one can learn about choice between different insurance options more than about the take-up choice. The conclusion we can draw from this is that defaults are not only important in policy, but also in designing HCEs.

Next, consider the mixed logit estimates reported in Table 9, model (6) which allow for taste heterogeneity. The standard deviations of the attributes are highly significant and almost of the same magnitude as the coefficient themselves which shows that valuation for insurance does show a lot of variation in our population. The estimated coefficients are larger than the coefficients of the multinomial logit model (Model 3), except for the gap coverage for brand-name drugs. A larger coefficient for the mixed logit model is expected as a part of the error is now explicitly modeled, so the remaining error should be smaller and the estimated coefficients larger. The relative difference of the mixed logit and logit coefficients gives an indication of the importance of the error coefficients. In our case, it is the coefficient for insurance, which rises most strongly in comparison to the multinomial model.

The normal distribution of the attributes' coefficients carries over to WTP with the standard deviation being equal to the standard deviation of the respective normal distribution divided by the premium coefficient, which we modeled to be constant. Table 10 shows estimated means and standard distributions WTP. Willingness to pay for gap coverage shows great variation within the population. Graph 1 shows the predicted distribution of WTP. A disadvantage of using a normal mixing distribution is that the support of a normal distribution is from minus infinity to plus infinity. This means that a part of the population will be predicted to have negative WTP. 18

percent of the population are predicted to have negative WTP for basic insurance, 23 percent for having no deductible and 43 for having coverage in the coverage gap.

To avoid the unrealistically high percentage of the population that is predicted to have negative WTPs (in particular for gap coverage), we finally estimate a model in which a subset of the coefficients is log normally distributed in Model 7.

Table 10 and Graph 1 show the distribution of WTPs. The variance is largest for the insurance coefficient. We see that the mean valuation for insurance is much higher than previously estimated, but that the valuation varies considerably in the population. This is what we expected as we have a considerable number of consumers who remained uninsured even in this highly subsidized market. Mean valuation for gap coverage is quite low, but there are consumers who value this extended coverage very highly.

Next, we will analyze whether these differences in valuation are due to expected risk. Thus, we turn to the question whether hypothetical choice experiments are a useful tool for analyzing observable differences of decision makers. This is important, because policy makers want to learn about vulnerable groups, such as consumers with low income, and about the importance of adverse selection in an insurance markets. In addition, there might be substantial heterogeneity in the population which is important in generalizing the findings from an experiment.

We thus conclude our analysis of hypothetical and real choices with specifications that reveal whether there is adverse selection in the Medicare Part D market (Table 11). Specifically, we test whether the demand for prescription drug insurance (both hypothetical and real) in the year when Part D was introduced depends on prescription drug use in the prior year. For this purpose, we add previous year drug use as an additional predictor in three specifications. Specifically, we interact a dummy variable that takes on the value 1 for "high cost" respondents with the generosity features of the offered plans (insurance, no deductible, gap coverage). These interactions are generally statistically significant (joint $F$-test, $p < 0.001$), in particular the interaction of the "high cost" dummy with the "insurance" dummy. Thus, respondents with higher levels of drug costs in the previous year are more likely to buy more generous coverage in the current year, which indicates that there is adverse selection. Importantly, further interactions with dummies for the hypothetical choices are not significant (specification 10), which indicates that adverse selection can also be detected by using the data based on the hypothetical choice tasks alone.

Finding that respondents choose their drug plans according to their expected drug use has different implications for this market. The bright side is that the elderly consumers in our sample seem to be able to identify the plans that are best for them, both in the hypothetical and in the real markets. This could actually be welfare increasing. On the other hand, adverse selection has well-known negative consequences for the supply side, and in extreme cases can lead to the complete break-down of the market (Rothschild and Stiglitz 1976). However, in Medicare Part D, the negative consequences of adverse selection on the supply side are mitigated through risk adjustment by Medicare. Adverse selection does not seem to have had severe negative consequences on the market, with the exception of plans with full coverage in the donut hole which have disappeared within two year, presumably due to strong adverse selection that was not compensated by risk adjustment (Heiss, et al. 2009).

Although this was not the intention when we designed these hypothetical choice experiments, these findings suggest that it might be possible to predict adverse selection on a new insurance market prior to its introduction by presenting a sample of individuals who are potential consumers (for instance, eligible individuals in the case of publicly sponsored health insurance) with appropriately designed hypothetical choice tasks.

## 6. Conclusions

In this paper, we analyzed data from hypothetical and real choices in the market for Medicare Part D prescription drug insurance. We presented and estimated several discrete choice models that offer different degrees of flexibility in combining such data in a joint estimation.

One insight from our analysis is that hypothetical choice experiments can be useful tools for analyzing and predicting the demand on insurance markets since behavior is clearly related to actual behavior. Significance and signs of the estimated coefficients are always the same in the real and hypothetical data. Even magnitudes are not statistically different with the exception of the take-up decision which in this specific application can be attributed to the role of defaults, an issue to which we return shortly. Another difference between our results for the two markets is that the hypothetical data predict higher willingness to pay for basic insurance and thus higher take-up than the real data. Does this suggest that respondents spend their hypothetical money more easily than their real money? Not necessarily: In the real world, the default was to remain uninsured, while in our hypothetical choices, "no insurance" required an active decision. This is supported by the fact that we do not find significant differences in hypothetical and real choices between different attributes of the insurances ("gap coverage" and "no deductible").

A special feature of the market we look at is that it consists of consumers who had no prescription drug insurance before the introduction. Making these consumers take-up insurance might require providing strong incentives from the side of the policy maker – having to opt out of insurance might be one option. Indeed, a high share of our respondent remains without insurance in the real market, even though the standard plans offered are highly subsidized by the government. Thus, one part of our respondents seems to have a very low valuation for insurance in general.

On the methodological side, we addressed the repeated nature of choices and taste heterogeneity, by estimating mixed logit models – and indeed, there is substantial taste heterogeneity especially in take-up, but also in the most comprehensive types of insurance. It this heterogeneity caused by risk? Thus, is there adverse selection, and how well does our HCE predict adverse selection? Since in our data, we observe drug use in the year before the introduction of Part D, we can perform an informal test of adverse selection. After controlling for the other covariates, including plan attributes, prior year drug costs predict choices both in hypothetical and real choices, which we interpret as an indication of adverse selection. Since this effect shows up both in the hypothetical and real choices, we would suggest that hypothetical choice tasks can be used to predict the degree of adverse selection on newly introduced insurance markets – this is an idea that warrants more investigation in future research. Also, our study shows the value of collecting rich background data in addition to those that are generated by standard hypothetical choice tasks.

We conclude with some thoughts about the limitations of our design. Combining data on hypothetical and real decisions allowed us to assess whether hypothetical choice data alone can be used as a tool to analyze and predict demand in a new insurance market that is characterized by heavy government subsidies and regulations. Our overall conclusion is that hypothetical choice data are useful in this respect, in particular when they are obtained from representative samples obtained from the population that faces an (active) decision in the new market. One might object that the hypothetical choice experiment was conducted after the initial enrollment period ended in May 2006 and that therefore respondents already had some experience with this new insurance market. While we acknowledge that this is a concern, we would also argue that if a hypothetical choice task is to be used to predict market demand, it is important that consumers are well-informed. Thus, our set-up in which consumers have some experience does not seem to be entirely unrealistic. Since new programs of the scale of Part D are introduced only very infrequently, a controlled experiment that systematically varies the availability of information on a market before presenting hypothetical choice tasks might provide additional insights into this issue; we leave this to future research.

Our analysis also revealed some important practical concerns. Most importantly, our hypothetical choice tasks treated the "no enrollment" option as an active decision (since we were substantively interested in its predictors), whereas in the real market where it is the default. We anticipate that similar concerns arise in other applications of hypothetical choice tasks as well. A second limitation of our study, which is more fundamental, is that even in a relatively simple choice situation such as plan selection in Medicare Part D, it is impossible to describe all product attributes (insurance plan design features) that might be relevant for a decision maker. Manski (1999), who formulated this critique at a general level, suggested using probabilistic choice tasks to address the fundamental uncertainty associated with the choice options in hypothetical tasks. We anticipate that both problem – the role of defaults and unspecified product attributes – will be addressed in future methodological studies and implementations of hypothetical choice tasks in the context of public policy design.

## References


**Alfnes, F., A. G Gutterson, G. Steine and K. Kolstad (2006):** Consumers' willingness to pay for the color of salmon: A choice experiment with real economic incentives. *American Journal of Agricultural Economics*, 88, 1050–1061.

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

**Arrow, K., R. Solow, P. R. Portney, E. E. Learner, R. Radner and H. Schuman (1993):** Report of the NOAA Panel on contingent valuation. *Federal Register*, 58, 4601–4614.

**Balistreri, E., G. McClelland, G. Poe and W. Schulze (2001):** Can hypothetical questions reveal true values? A laboratory comparison of dichotomous choice and open-ended contingent values with auction values. *Environmental and Resource Economics*, 18, 275–292.

**Berry, S. (1994):** Estimating discrete-choice models of product differentiation. *Rand Journal of Economics*, 25, 242–262.

**Blackburn, M., G. W. Harrison, and E. E. Rutström (1994):** Statistical bias functions and informative hypothetical surveys. *American Journal of Agricultural Economics* 76, 1084–1088.

**Bohm, P. (2008):** Field test elicitations of demand for public goods. In C. R. Plott and V. L. Smith (eds.): *Handbook of Experimental Economic Results*, Volume 1, 736–741. Elsevier.

**Carson, R. T., N. E. Flores, K. M. Martin, and J. L. Wright (1996):** Contingent valuation and revealed preference methodologies: Comparing the estimates for quasi-public goods. *Land Economics*, 72, 80–99.

**Chang, J. B., J. L. Lusk and F. B. Norwood, (2009):** How closely do hypothetical surveys and laboratory experiments predict field behavior? *American Journal of Agricultural Economics*, 91, 518–534.

**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.

**Frakt, A. B. and S. D. Pizer (2010):** Beneficiary price sensitivity in the Medicare prescription drug plan market. *Health Economics*, 19(1), 88–100.

**Harrison, G. W. and J. A. List (2004)**: Field experiments. *Journal of Economic Literature*, 62, 1009–1055.

**Heiss, F., D. McFadden, and J. Winter (2006):** Who failed to enroll in Medicare Part D, and why? Early results. *Health Affairs*, 25, 344–354.

**Heiss, F., D. McFadden, and J. Winter (2009):** Regulation of private health insurance markets: Lessons from enrollment, plan type choice, and adverse selection in Medicare Part D. NBER Working Paper No.15392.

**Heiss, F., D. McFadden, and J. Winter (2011):** The demand for Medicare Part D prescription drug coverage: Evidence from four waves of the Retirement Perspectives Survey. In: D. A. Wise, ed., Explorations in the Economics of Aging, 159–182. Chicago University Press.

**Heiss, F., A. Leive, D. McFadden, and J. Winter (2012):** Plan selection in Medicare Part D: Evidence from administrative data. 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, 199–235.

**List, J.A., P. Sinha, and M.H. Taylor (2006):** Using choice experiments to value non-market goods and services. *Advances in Economic Analysis and Policy*, 6, 1–37.

**Louviere, J., D. Hensher, and J. Swait (2000):** *Stated Choices Methods*. Cambridge University Press.

**Lucarelli, C., J. Prince and K. Simon (2008):** The welfare impact of reducing choice in Medicare Part D: A comparison of two regulation strategies. NBER Working Paper No. 14296.

**Lusk, J. L. and T. C. Schroeder (2004):** Are choice experiments incentive compatible? A test with quality differentiated beef steaks. *American Journal of Agricultural Economics*, 86, 467–482.

**Madden, J., A. Graves, F. Zhang, A. Adams, B. Briesacher, D. Ross-Degnan, J. Gurwitz, M. Pierre-Jacques, D. Safran, G. Adler and S. Soumerai (2008):** Cost related medication nonadherence and spending on basic needs following implementation of Medicare Part D, *Journal of the American Medical Association*, 299, 1922–1928.

**Manski, C. F. (1999):** Analysis of choice expectations in incomplete scenarios. *Journal of Risk and Uncertainty*, 19, 49–65.

**McFadden, D. (1976):** The revealed preferences of a government bureaucracy: Empirical evidence. *Bell Journal of Economics*, 7(1), 55-72.

**McFadden, D. (2001):** Economic choices. *American Economic Review*, 91, 351–378.

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

**McFadden, D., J. Winter, and F. Heiss (2008):** Consumer-directed health care: Can consumers look after themselves? *Swiss Journal of Economics*, 144, 285–307.

**Mitchell, R. and R. Carson (1989):** *Using Surveys to Value Public Goods: The Contingent Valuation Method*. RFF Press, Washington, D.C.

**Morikawa, T. (1989):** *Incorporating Stated Preference Data in Travel Demand Analysis*. Ph.D. Thesis, Massachusetts Institute of Technology.

**Murphy, J. J., P. G. Allen, T. H. Stevens, and D. Weatherhead (2005):** A Meta-Analysis of Hypothetical Bias in Stated Preference Valuation. *Environmental and Resource Economics*, 30, 313–325.

**Neuman, P., M. Strollo, S. Guterman, W. Rogers, A. Li, A. Rodday, and D. Safran (2007):** Medicare prescription drug benefit progress report: Findings from a 2006 national survey of seniors. *Health Affairs*, 26, 630–643.

**Rothschild, M., and Stiglitz, J. (1976):** Equilibrium in competitive insurance markets: an essay on the economics of imperfect information. *Quarterly Journal of Economics*, 90(4), 630-649.

**Street, D.J. and L. Burgess (2007):** *The Construction of Optimal Stated Choice Experiments: Theory and Methods*. John Wiley & Sons Inc.

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

**Winter, J., R. Balza, F. Caro, B.-H. Jun, F. Heiss, R. Matzkin, and D. McFadden (2006):** Medicare prescription drug coverage: Consumer information and preference. *Proceedings of the National Academy of Sciences of the United States of America*, 103, 7929–7934.

## Appendix: **The hypothetical choice experiment in RPS 2006**

Introduction (part 1) for respondents without prescription drug coverage:

*At the end of this year, you will be able to make new choices about your prescription drug coverage for the next year.*

Introduction (part 1) for respondents with Part D plans (whether stand-alone or HMO/ Medicare+Choice):

*At the end of this year, you will be able to make new choices about your prescription drug coverage for the next year. You may stay in your current plan, you may switch to another plan, or you may even elect to unsubscribe and not select any plan.*

Introduction (part 1) for respondents with other prescription drug coverage, i.e. through their employer or union, the Veteran's Administration, private insurance or from some other source:

*Even though you have prescription drug insurance from other sources, we would like to know what your choices would be if the only coverage you could get would be through a Part D plan.*

Introduction (part 2), identical for all respondents:

*We are now going to show you some plans that have realistic features and premiums. We are interested in what plan you would choose if these were your only options. Specifically, on each of the following pages we will show you three plans that differ in coverage and premiums. On each page, please report which of these options is the most attractive and which is the least attractive. You will always have the option to choose none of these three plans and thus have no prescription drug coverage (but then you would have to pay higher premiums if you enroll later, according to current Medicare Part D regulations.*

*Since 2006, Part D of Medicare provides coverage for prescription drugs of older Americans. Plans under Part D are also known as Medicare Rx plans. Once you are eligible for Medicare, you can enroll in one of the new prescription drug plans under Medicare Part D.*

Introduction (part 2, last sentence) for respondents with stand-alone Part D plans:

*You have told us earlier that you are enrolled in such a plan.*

Introduction (part 2, last sentence) for all other respondents:

*You have told us earlier that you are not enrolled in such a plan - either because you have coverage from other sources or because you decided not to enroll in a Medicare Rx plan.*

The choice task was as follows:

*Please consider a situation in which you would have no prescription drug coverage from any other source. Imagine that these were the only three description drug plans that you could choose from. You can also choose not to have coverage at all.*

- *Basic Plan premium: $ PB This plan covers all prescription drugs you currently use and most of what you might need in the future. It has a deductible of $250, pays 75 percent of costs above $250 up to $2250, provides no additional benefit until costs reach $5100, and pays 95 percent of costs above that level.*

- *Enhanced Plan premium: $ PE This plan is equivalent to the Basic Plan but has no deductible. This means that the 75 percent coverage begins at the first dollar you spend on description drugs, up to $2250. Like the Basic Plan, there are no additional benefits until costs reach $5100. The Enhanced Plan pays 95 percent of costs above that level.*

- *Premier plan premium: $ PP This plan is equivalent to the Enhanced Plan, but is does not impose a coverage gap between $2250 and $5100. So it pays 75 percent of all costs up to $5100 and for 95 percent above that amount.*

- *No prescription drug insurance at all*

The three treatment variables (dollar premiums) *PB*, *PE*, and *PP* (for the basic, enhanced, and premium plans, respectively) were assigned randomly according. There were 26 treatments, each with equal probability. The values of the treatment variables are listed in Table A.1.

**Table 1:** Socio-economic characteristics of respondents in HRS 2006 and RPS 2006

| | | HRS 2006 | | RPS 2006 | |
|---|---|---|---|---|---|
| | | unweigthed | weigthed | unweigthed | weigthed |
| Gender | Female | 57.3% | 56.8% | 55.8% | 57.2% |
| | Male | 42.7% | 43.2% | 44.2% | 42.8% |
| Race | White | 83.5% | 89.3% | 87.6% | 83.3% |
| | Non-white | 16.5% | 10.7% | 12.4% | 16.7% |
| Age | 61 – 70 | 35.8% | 33.7% | 39.3% | 35.9% |
| | 71 – 80 | 40.3% | 41.6% | 46.9% | 47.9% |
| | 81 – 90 | 20.4% | 22.0% | 12.9% | 15.1% |
| | >90 | 3.5% | 2.7% | 0.9% | 1.1% |
| Education | Less than HS | 31.5% | 28.3% | 12.9% | 26.1% |
| | High school | 32.6% | 33.4% | 41.5% | 36.5% |
| | More than HS | 36.0% | 38.4% | 45.6% | 37.5% |
| Income | <$20K | 33.2% | 31.2% | 23.4% | 28.9% |
| | $20K – $60K | 46.2% | 46.9% | 58.2% | 52.6% |
| | >$60K | 20.6% | 21.9% | 18.4% | 18.5% |
| SRHS | excellent | 8.5% | 9.1% | 6.1% | 5.6% |
| | very good | 26.3% | 27.5% | 32.2% | 27.8% |
| | good | 31.6% | 32.4% | 39.5% | 41.8% |
| | fair | 23.3% | 22.2% | 18.1% | 19.8% |
| | poor | 10.3% | 8.8% | 4.0% | 4.9% |
| Number of observations | | 11399 | | 1666 | |

**Table 2:** Variable description and descriptive statistics

| Variable | Definition | Mean | Std. dev. | N |
|---|---|---|---|---|
| **Hypothetical choices** | | | | |
| Attributes of plans chosen by RPS respondents | | | | |
| Insurance | = 1 if Part D stand-alone coverage | 0.85 | 0.35 | 1525 |
| No deductible | = 1 if Insurance without deductible | 0.68 | 0.47 | 1525 |
| Gap coverage | = 1 if Insurance with gap coverage | 0.36 | 0.48 | 1525 |
| Premium | Refer to table 4 | | | |
| **Real choices** | | | | |
| Attributes of plans chosen by RPS respondents | | | | |
| Insurance | = 1 if Part D stand-alone coverage | 0.79 | 0.41 | 470 |
| No deductible | = 1 if Plan offers benefits without the $250 deductible of the standard plan | 0.49 | 0.50 | 470 |
| Gap coverage (generics) | = 1 if Generic drugs covered in the coverage gap of the standard plan | 0.07 | 0.26 | 470 |
| Gap coverage (brand-name drugs) | = 1 if In addition to generics, brand-name drugs also covered in the coverage gap | 0.03 | 0.18 | 470 |
| **Health and prescription drug spending** | | | | |
| Drug costs | Total drug costs in 2005, constructed from RPS 2005 (see Winter et al., 2006) | 2554.30 | 3118.45 | 1569 |

**Table 3:** Monthly plan premiums in the hypothetical and real markets (in dollars)

| **Hypothetical market (set by design)** | | **Basic** | **Enhanced** | | **Premium** |
|---|---|---|---|---|---|
| First choice | Fixed premium | 30.79 | 37.68 | | 50.33 |
| Second and third choice | Lowest premium | 15.39 | 18.94 | | 25.16 |
| | Highest premium | 40.02 | 49.25 | | 65.43 |
| **Real market** | | **Basic** | **Enhanced** | **Generics** | **Generics & brand** |
| All available plans (market average) | | 30.75 | 37.92 | 48.13 | 61.88 |
| Plans chosen by RPS respondents (sample average) | | 17.00 | 26.60 | 46.10 | 60.80 |

**Table 4:** Multinomial logit models estimated separately for real and hypothetical choice data

| | (1) | (2) |
|---|---|---|
| VARIABLES | Real | Hypothetical |
| Premium | -0.0762*** | -0.0323*** |
| | (0.00609) | (0.00777) |
| Insurance | 0.980*** | 0.964*** |
| | (0.206) | (0.278) |
| No deductible | 1.123*** | 0.718*** |
| | (0.200) | (0.130) |
| Gap coverage (generics) | 0.184 | |
| | (0.277) | |
| Gap coverage (brand-name drugs) | 1.565*** | 0.489*** |
| | (0.396) | (0.135) |
| Drug tiers | -0.927*** | |
| | (0.231) | |
| Top 100 drugs uncovered | -0.109*** | |
| | (0.0132) | |
| Top 100 with authorization | -0.0789*** | |
| | (0.0126) | |
| Observations | 18,185 | 4,812 |
| Number of groups | 410 | 1203 |
| Log likelihood | -1144 | -1613 |

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

**Table 5:** Monthly WTP estimates obtained from separate MNL models (using the delta method)

| | (1)<br>Real | (2)<br>Hypothetical |
|---|---|---|
| Insurance | 12.86*** | 29.86*** |
| | (2.93) | (4.35) |
| No deductible | 14.73*** | 22.25*** |
| | (2.97) | (5.30) |
| Gap coverage (generics) | 2.41 | |
| | (3.58) | |
| Gap coverage (brand-name drugs) | 20.53*** | 15.15*** |
| | (5.10) | |
| Drug tiers | -12.16*** | |
| | (3.51) | |
| Top 100 drugs uncovered | -1.44*** | |
| | (0.22) | |
| Top 100 with authorization | -1.04*** | |
| | (0.18) | |

*** p<0.01, ** p<0.05, * p<0.1, standard errors in parenthesis

**Table 6:** Ratios of WTPS for hypothetical and real decisions

| | WTP(Hypo)/WTP(Real) | [95% confidence interval] |
|---|---|---|
| Insurance | 2.32 | [1.34 , 3.30] |
| No deductible | 1.51 | [0.66 , 2.37] |
| Gap coverage (brand-name drugs) | 0.74 | [0.27 , 1.20] |

**Table 7:** Multinomial logit models estimated with combined real and hypothetical choice data

| | (3) | (4) | (5) |
|---|---|---|---|
| Premium | -0.0545*** | -0.0647*** | -0.0762*** |
| | (0.00429) | (0.00603) | (0.00600) |
| Insurance | 1.416*** | 1.378*** | 0.980*** |
| | (0.187) | (0.240) | (0.229) |
| Insurance * Hypothetical | | | 1.296*** |
| | | | (0.338) |
| No deductible | 1.002*** | 1.555*** | 1.123*** |
| | (0.116) | (0.220) | (0.201) |
| No deductible * Hypothetical | | | 0.574 |
| | | | (0.445) |
| Gap coverage (generics) | 0.147 | 0.0989 | 0.184 |
| | (0.190) | (0.197) | (0.262) |
| Gap coverage (brand-name drugs) | 0.754*** | 1.163*** | 1.565*** |
| | (0.104) | (0.292) | (0.392) |
| Gap coverage (brand-name) * Hypothetical | | | -0.410 |
| | | | (0.451) |
| Drug tiers | -1.372*** | -1.666*** | -0.927*** |
| | (0.121) | (0.178) | (0.217) |
| Top 100 drugs uncovered | -0.136*** | -0.119*** | -0.109*** |
| | (0.0177) | (0.0179) | (0.0165) |
| Top 100 with authorization | -0.104*** | -0.0982*** | -0.0789*** |
| | (0.0138) | (0.0153) | (0.0147) |
| Heteroskedasticity Parameter | | -0.762*** | -0.860*** |
| | | (0.237) | (0.252) |
| Observations | 22,997 | 22,997 | 22,997 |
| Number of groups | 1,613 | 1,613 | 1,613 |
| log likelihood | -2781 | -2771 | -2757 |

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table 8:** Monthly WTP estimates obtained from MNL models with combined hypothetical and real choice data (using the delta method)

| | (3) | (4) |
|---|---|---|
| Insurance | 25.96*** | 21.29*** |
| | (2.40) | (3.58) |
| No deductible | 18.40*** | 24.03*** |
| | (2.25) | (3.14) |
| Gap coverage (generics) | 2.70 | 1.53 |
| | (3.48) | (3.03) |
| Gap coverage (brand-name drugs) | 13.83*** | 17.97*** |
| | (1.83) | (4.03) |
| Drug tiers | -25.15*** | -25.74*** |
| | (3.06) | (3.40) |
| Top 100 drugs uncovered | -2.49*** | -1.84*** |
| | (0.42) | (0.39) |
| Top 100 with authorization | -1.90*** | -1.52*** |
| | (0.24) | (0.24) |

*** p<0.01, ** p<0.05, * p<0.1

**Table 9:** Mixed logit models estimated with combined real and hypothetical choice data

| VARIABLES | (6) Mean | (6) SD | (7) Mean | (7) SD |
|---|---|---|---|---|
| Premium | -0.0686*** | | -0.0796*** | |
| | (0.00515) | | (0.00506) | |
| Gap coverage (generics) | -0.382 | | -0.366 | |
| | (0.257) | | (0.281) | |
| Drug tiers | -1.402*** | | -1.322*** | |
| | (0.148) | | (0.148) | |
| Top 100 drugs uncovered | -0.117*** | | -0.111*** | |
| | (0.0135) | | (0.0133) | |
| Top 100 with authorization | -0.102*** | | -0.109*** | |
| | (0.0129) | | (0.0136) | |
| Insurance | 3.073*** | 3.389*** | 3.263*** | 3.222*** |
| | (0.327) | (0.379) | (0.321) | (0.315) |
| No deductible | 1.376*** | 1.835*** | -0.340 | 2.027*** |
| | (0.146) | (0.180) | (0.235) | (0.297) |
| Gap coverage (brand-name drugs) | 0.456** | 2.450*** | -0.620*** | 1.443*** |
| | (0.187) | (0.209) | (0.191) | (0.130) |
| Observations | 22,997 | 22,997 | 22,997 | 22,997 |
| Number of groups | 1,613 | 1,613 | 1,613 | 1,613 |
| log likelihood | -2402 | -2402 | -2439 | -2439 |

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

**Table 10:** WTP estimates obtained from mixed logit models with combined hypothetical and real choice data

| | Coef. | Std. Err. | p | Distribution |
|---|---|---|---|---|
| Insurance | | | | N(44.80, 49.40) |
| No deductible | | | | N(20.05, 26.75) |
| Gap Coverage (brandname) | | | | N(6.65, 35.71) |
| Gap Coverage (generics) | -5.58 | 3.88 | 0.15 | |
| Drug Tiers | -20.45 | 2.85 | 0.00 | |
| Top 100 drugs uncovered | -1.70 | 0.25 | 0.00 | |
| Top 100 drugs with authorization | -1.49 | 0.20 | 0.00 | |

**Table 11:** Extended specification testing for adverse selection

| Explanatory variables | (8) | (9) | (10) |
|---|---|---|---|
| Premium | -0.0550*** | -0.0643*** | -0.0763*** |
| | (0.00434) | (0.00583) | (0.00603) |
| Insurance | 1.208*** | 1.052*** | 0.590** |
| | (0.204) | (0.260) | (0.259) |
| Insurance * Highcost | 0.930*** | 1.516*** | 1.444*** |
| | (0.294) | (0.418) | (0.346) |
| No deductible | 0.934*** | 1.432*** | 1.163*** |
| | (0.131) | (0.230) | (0.212) |
| No deductible * Highcost | 0.220 | 0.166 | -0.0997 |
| | (0.198) | (0.242) | (0.238) |
| Gap coverage (generics) | 0.139 | 0.0601 | 0.184 |
| | (0.190) | (0.202) | (0.262) |
| Gap coverage (brand-name drugs) | 0.492*** | 0.541** | 0.972* |
| | (0.131) | (0.234) | (0.559) |
| Gap coverage (brand-name) * Highcost | 0.623*** | 1.271*** | 1.120* |
| | (0.200) | (0.416) | (0.634) |
| Drug tiers | -1.390*** | -1.629*** | -0.926*** |
| | (0.124) | (0.174) | (0.217) |
| Top 100 drugs uncovered | -0.136*** | -0.120*** | -0.109*** |
| | (0.0178) | (0.0178) | (0.0165) |
| Top 100 with authorization | -0.105*** | -0.0994*** | -0.0790*** |
| | (0.0140) | (0.0153) | (0.0147) |
| Insurance * Hypothetical | | | 1.341*** |
| | | | (0.371) |
| Insurance * Highcost * Hypothetical | | | 0.137 |
| | | | (0.797) |
| No deductible * Hypothetical | | | 0.221 |
| | | | (0.444) |
| No deductible * Highcost * Hypothetical | | | 0.956 |
| | | | (0.645) |
| Gap coverage (brand-name) * Hypothetical | | | -0.399 |
| | | | (0.607) |
| Gap coverage (brand-name) * Hypothetical * Highcost | | | 0.212 |
| | | | (0.777) |
| Heteroskedasticity parameter | | -0.653*** | -0.825*** |
| | | (0.174) | (0.254) |
| Observations | 22,997 | ,997 | 22,997 |
| log likelihood | -2728 | -2718 | -2701 |
| Number of groups | 1,613 | 1,613 | 1,613 |
| Robust standard errors in parentheses | | | |
| *** p<0.01, ** p<0.05, * p<0.1 | | | |

**Table A.1:** Values of the treatment variables in the hypothetical choice experiment in RPS 2006

| Treatment | PB<br>basic plan premium | PE<br>enhanced plan premimum | PP<br>premium plan premium |
|---|---|---|---|
| 1 | 15.39 | 18.94 | 25.16 |
| 2 | 15.39 | 18.94 | 31.38 |
| 3 | 15.39 | 18.94 | 35.12 |
| 4 | 15.39 | 22.49 | 28.71 |
| 5 | 15.39 | 22.49 | 34.93 |
| 6 | 15.39 | 22.49 | 38.67 |
| 7 | 15.39 | 24.62 | 30.84 |
| 8 | 15.39 | 24.62 | 37.06 |
| 9 | 15.39 | 24.62 | 40.80 |
| 10 | 30.79 | 34.34 | 40.56 |
| 11 | 30.79 | 34.34 | 46.78 |
| 12 | 30.79 | 34.34 | 50.52 |
| 13 | 30.79 | 37.88 | 44.1 |
| 14 | 30.79 | 37.88 | 54.06 |
| 15 | 30.79 | 40.01 | 46.23 |
| 16 | 30.79 | 40.01 | 52.45 |
| 17 | 30.79 | 40.01 | 56.19 |
| 18 | 40.02 | 43.57 | 49.79 |
| 19 | 40.02 | 43.57 | 56.01 |
| 20 | 40.02 | 43.57 | 59.75 |
| 21 | 40.02 | 47.12 | 53.34 |
| 22 | 40.02 | 47.12 | 59.56 |
| 23 | 40.02 | 47.12 | 63.3 |
| 24 | 40.02 | 49.25 | 55.47 |
| 25 | 40.02 | 49.25 | 61.69 |
| 26 | 40.02 | 49.25 | 65.43 |

**Figure 1:** Distributions of WTP estimation from mixed logit models with normal distribution