

Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program

By JASON ABALUCK AND JONATHAN GRUBER*

We evaluate the choices of elders across their insurance options under the Medicare Part D Prescription Drug plan, using a unique dataset of prescription drug claims matched to information on the characteristics of choice sets. We document that elders place much more weight on plan premiums than on expected out-of-pocket costs; value plan financial characteristics beyond any impacts on their own financial expenses or risk; and place almost no value on variance-reducing aspects of plans. Partial equilibrium welfare analysis implies that welfare would have been 27 percent higher if patients had all chosen rationally. (JEL D12, I11, J14)

The Medicare Modernization Act of 2003, better known as the legislation that added the Part D prescription drug benefit to the Medicare program, represents the single most significant expansion of public insurance programs in the US in the past 40 years. The most novel, and controversial, feature of this legislation was the use of multiple private insurance providers to deliver this new public insurance product. Unlike the traditional model of government mandated uniform insurance packages for all enrollees, under the Part D program dozens of private insurers were allowed to offer a wide range of products with varying prices and product features. Perhaps most well known was the extent to which plans covered the “donut hole,” a broad uncovered range of expenditures in the minimum mandated plan.

This unprecedented privatization of the delivery of a public insurance product raises a host of important policy questions. Primary among these is the impact of allowing choice across so many private insurance options. The typical elder in our data (described below) faces a choice of over 40 stand-alone drugs plans, and our estimates suggest that the range of cost from the most to least expensive option facing an elder is comparable to the mean of those costs. Choice is clearly meaningful in this context. Yet, to date, we know almost nothing about how elders are making these crucial choices.

This paper investigates the choices of elders for the newly formed Part D program in 2006. We analyze data that provide information on the Part D plans chosen and prescription drug utilization for a large sample of elders in the United States. These

*Abaluck and Gruber: Department of Economics, MIT, 50 Memorial Drive, Cambridge, MA 02142 (e-mail: abaluck@mit.edu, gruberj@mit.edu). We are grateful to Amy Finkelstein, Penny Goldberg, Jerry Hausman, Panle Jia, Fiona Scott Morton, and seminar participants at Boston University, Carnegie Mellon University, Duke University, MIT, and the NBER for helpful comments; to Jim Hendrix, Chris Messner, Pallavi Mudumby, Steven Pieri, and John Porell from Wolters Kluwer for providing data; to Matthew Harding for providing Matlab code, to Iuliana Pascu, Arnaldo Pereira, Charles Wu, Josephine Duh, and particularly John Graves, Natalija Novta, and Ashley Swanson, for research assistance; and to the National Science Foundation and the National Institute of Aging for financial support (NIA grants R01 AG031270 and T32 AG000186-21).

data were collected by Wolters Kluwer (WK), a “switch agent” that lies between pharmacies that fill prescriptions and the insurance companies and prescription benefit managers that pay for them. WK collects information on almost one-third of all third-party prescription drug transactions, and we will use the universe of their data for those over age 65 during 2005–2006 to examine choice of Part D plan. We match to this dataset a comprehensive set of information from the Centers for Medicare and Medicaid Services (CMS) on the Part D plans available to each person in our dataset.

Specifically, for each elder whose claims appear in our sample, we model the financial implications of each of the plans in their choice set, based on both 2005 and 2006 drug utilization and several different models of expectations. We begin by presenting the basic facts on choice, documenting that the vast majority of elders are choosing plans that are not on the “efficient portfolio” of plan choice for that elder. We then turn to more rigorous multinomial models of individual choice to incorporate nonfinancial characteristics, preference heterogeneity, and unobserved plan characteristics into our analysis.

Our findings are striking: along three dimensions, elders are making choices that are inconsistent with optimization under full information. First, elders place much more weight on plan premiums than they do on the expected out-of-pocket costs that they will incur under the plan. Second, consumers appear to value plan financial characteristics far beyond any impacts on their own financial expenses or risk. Third, consumers substantially undervalue variance-reducing aspects of alternative plans. The first two of these conclusions are robust to a variety of specifications and econometric approaches; the third is more sensitive.

Our paper proceeds as follows. Section I provides background on the Part D program and reviews the growing literature on its impacts. Section II discusses our data sources, and Section III presents initial results on choice set variation and choice behavior. Section IV describes our choice framework, and Section V presents results and robustness checks. Section VI considers issues of misspecification and measurement error, and Section VII assesses robustness to heterogeneity concerns. Section VIII concludes with a discussion of the policy implications of our findings. The appendices referenced throughout are included in the Web Appendix of this paper.

I. Background

A. *The Medicare Part D Program*

Medicare, which provides universal health insurance coverage to those over age 65 and to those on the disability insurance program, was established in 1965. The original program covered most medical needs for the elderly and disabled, including hospital and doctor costs, but it excluded coverage for prescription drugs. This omission was not perceived as a major one in the early years of the Medicare program, but in the 1990s the advancement of prescription drug treatments for common illnesses among the elderly drew attention to this gap in Medicare coverage. Medicare recipients, for example, spent an average of \$2,500 each on prescription drugs in 2003, more than twice what the average American spent on all health care in 1965.¹

¹ Data for prescription drug spending comes from the Congressional Budget Office (2002). Data for average Americans' health spending comes from the “National Health Expenditures” section of the Centers for Medicare

In 2003, the Bush administration and Congress reached agreement on a far-reaching prescription drug benefit package at a projected cost to the federal government of \$40 billion per year for its first ten years. The most noticeable innovation of the Part D plan is that this new Medicare benefit is not delivered by the government, but rather by private insurers under contract with the government. Beneficiaries can choose from three types of private insurance plans coverage of their drug expenditures. The first is stand-alone plans called Medicare Prescription Drug Plans (PDP) (a plan that just offers prescription drug benefits). In 2006, there were 1,429 total PDPs offered throughout the nation, with most states offering about 40 PDPs. The majority of PDPs are offered by a dozen national or near-national companies.

The second alternative is Medicare Advantage (MA) plans, plans that provide all Medicare benefits, including prescription drugs, such as HMO, PPO, or Private FFS plans. There were 1,314 total plans nationally in 2006. Finally, beneficiaries could retain their current employer/union plan, as long as coverage is “creditable” or at least as generous (i.e., actuarially equivalent) as the standard Part D plan, for which they would receive a subsidy from the government.

Under Part D, recipients are entitled to basic coverage of prescription drugs by a plan with a structure actuarially equivalent to the following: none of the first \$250 in drug costs each year; 75 percent of costs for the next \$2,250 of drug spending (up to \$2,500 total); 0 percent of costs for the next \$3,600 of drug spending (up to \$5,100 total, the “donut hole”); and 95 percent of costs above \$5,100 of drug spending. Over 90 percent of beneficiaries in 2006, however, were not enrolled in the standard benefit design, but rather are in plans with low or no deductibles, flat payments for covered drugs following a tiered system, or some form of coverage in the donut hole. The main requirement for plans is that they must have equal or greater actuarial value than the standard benefit.² The government also placed restrictions on the structure of the formularies that plans could use to determine which prescription medications they would ensure. Overall, Part D sponsors have great flexibility in terms of plan design. Many insurance companies sponsored multiple plans of differing levels of premiums and coverage generosity. Arranging the data into cells by plan sponsor and state, we find that only a quarter of the cells have only one plan, and 58 percent contain three plans. In those sponsor/state cells with multiple plans, most sponsors offer one standard plan, and one or two enhanced plans.

Enrollment in Part D plans was voluntary for Medicare eligible citizens, although Medicare recipients not signing up by May 15, 2006, were subject to a financial penalty if they eventually joined the program (to mitigate adverse selection in the choice of joining the program). One group, however, was automatically enrolled: low-income elders who had been receiving their prescription drug coverage through state Medicaid programs (the “dual eligibles”). These dual eligibles were enrolled in Part D plans by default if they did not choose one on their own. The Part D plans for dual eligibles could charge copayments of only \$1 for generics/\$3 for name brand drugs for those below the poverty line, and only \$2 for generics/\$5 for name brand

and Medicaid Services' *National Health Accounts*.

² Cover Memo for Medicare Part D Benefit Parameters: Annual Adjustments for Standard Benefit in 2007 (CMS).

drugs for those above the poverty line, with free coverage above the out-of-pocket threshold of \$3,600.³

Despite reluctance voiced before the legislation passed, there was enormous interest from insurers in participating in the Part D program. By November 2006, 3,032 plans were being offered to potential Part D enrollees. Every county in the nation had at least 27 plans available; the typical county had 48 plans, while some counties featured more than 70 choices, primarily due to high number of MA plans.⁴

Enrollment in the new Part D program was initially fraught with problems, but in the following months the federal government was able to iron out many of the problems that had arisen during the initial transition. As of June 2006, there were 10.4 million people enrolled in stand-alone PDP plans, 5.5 million people enrolled in MA plans, and about 6 million dual eligibles.⁵ Yet 73 percent of people over 65 felt that the Medicare prescription drug benefit was too complicated, while 91 percent of pharmacists and 92 percent of doctors expressed this concern. When asked if they agree with the statement “Medicare should select a handful of plans that meet certain standards so seniors have an easier time choosing,” 60 percent of seniors answered “Yes.”⁶

Despite these reservations, there were no signs of diminished plan choice in subsequent years. The number of PDPs increased by about 30 percent in 2007, from 1,429 to 1,875, and remained at this level in 2008.⁷

B. Issues of Elder Choice in Part D

Mark Duggan, Patrick Healy, and Fiona Scott Morton (2008) provide a detailed overview of many of the economic issues raised by the Medicare Part D Program. The use of this private delivery device, with such a multiplicity of choices, is a novel feature of the Part D legislation. Standard economic theory would suggest that this is a beneficial plan feature: allowing individuals to choose across a wide variety of plans that meet their needs, rather than constraining them to a limited set of choices being made by the government, can only increase welfare in the standard model in a partial equilibrium setting.

But there are reasons to believe that the standard model is insufficient, particularly for a population of elders. There is growing interest in behavioral economics in models where agents are better off with a more restricted choice set, as nicely reviewed in Sheena S. Iyengar and Emir Kamenica (2006). Recent theoretical work shows that the traditional “more is better” principle may be reversed in some choice set contexts, for

³ In addition, two other groups receive substantial subsidies—those found eligible for Low Income Subsidy (LIS) or for Partial Subsidy by the SSA. To qualify for LIS, beneficiaries must have income less than 135 percent of poverty and resources less than \$7,500/individual or \$12,000/couple. This group received benefits comparable to the dual eligibles with incomes above 100 percent of poverty. To qualify for Partial Subsidy, beneficiaries must have income at 135 percent–150 percent of poverty and resources less than \$11,500/individual or \$23,000/couple. This group can enroll in plans with a \$50 deductible, a 15 percent copayment up to the out-of-pocket threshold, and \$2/\$5 copayments above that point. In addition, premiums are fully paid by the government up to 135 percent of poverty, and then partially subsidized up to 150 percent of poverty.

⁴ Details on number of plans in a median county obtained from Prescription Drug Plan Formulary and Pharmacy Network Files for 2006, provided by CMS.

⁵ Enrollment data (rounded). Enrollment numbers available at <http://www.kff.org/medicare/upload/7453.pdf>.

⁶ Kaiser Family Foundation and Harvard School of Public Health (2006).

⁷ Data on 2008 plans taken from CMS 2008 PDP Landscape Source (v. 09.25.07) available at <http://www.cms.hhs.gov/prescriptiondrugcovenin/>.

example, when the presence or absence of options conveys information (Kamenica 2008; Dmitri Kuksov and J. Miguel Villas-Boas 2006) or when agents have preferences with regret (Ben Irons and Cameron Hepburn 2007; Todd Sarver 2008). And a growing body of empirical work shows that individuals are less likely to participate in markets where they face more choice; decisions to purchase a good (Iyengar and Mark R. Lepper 2000; Peter Boatwright and Joseph C. Nunes 2001), take a loan (Marianne Bertrand et al. 2005) or enroll in a 401(k) plan (Iyengar, Gur Huberman, and Wei Jiang 2004) are found to decrease when participation requires choosing from a larger set of alternatives.

Iyengar and Kamenica (2006) find that not only the decision to participate in a market, but also the nature of choice itself, is affected by the size of the option set. They investigate choice over asset allocation in both laboratory and real-world (pension plan choice) settings, and find that individuals opt for safer investments when faced with a larger range of risky choices. In particular, they find that the presence of more investment options in a 401(k) plan leads to more frequent choice of money market or bond options rather than equity investment. Iyengar and Lepper (2000) also find that satisfaction with choices made falls with the size of the choice set in several experimental settings.

Another recent literature has shown that the nature of how choices are presented can have important impacts on choice. For example, Justine S. Hastings and Lydia Tejada-Ashton (2008) examine financially illiterate individuals choosing across retirement funds in Mexico's privatized Social Security system. They find that presenting information on plan administrative fees in pesos, rather than in percentage terms, causes a significant shift in choice toward lower-fee plans. In another study, Raj Chetty, Adam Looney, and Kory Kroft (2007) find that consumers are much more sensitive to tax rates when the tax burden is included in posted prices rather than added at the register.

These issues may be paramount within the context of the elderly, given that the potential for cognitive failures rises at older ages. Salthouse (1996) shows clear evidence that the performance on a series of memory and analytic tasks declines sharply after age 60. Part of the reason for this may be the rise in incidence of dementia with age; starting at age 60, dementia rates roughly double every five years (Laura Fratiglioni, Diane De Ronchi, and Hedda A. Torres 1999). A recent study by Sumit Agarwal et al. (2006) shows that in ten different contexts, ranging from credit card interest payments to mortgages and small business loans, the elderly pay higher fees and face higher interest rates than middle-aged consumers. These types of findings raise particular concern about choice in the Part D context.

C. Previous Studies of Part D Choice

We are aware of only three previous studies of these issues in the context of Part D. The first is a set of studies by Daniel McFadden and colleagues, as summarized in Florian Heiss, Daniel McFadden, and Joachim Winter (2006). These researchers surveyed a set of elders about their plans for enrolling in Part D programs, and evaluate whether enrollment intentions in the plan were "rational" given the penalties for delay. They find that 71 percent of potential enrollees were making the appropriate decision (under various assumptions about discount rates, etc.), while 10 percent of enrollees did not intend to enroll when it would be in their interest to

do so, and 19 percent intended to enroll when it would be in their interest to delay. Thus, for most potential enrollees, the decision over whether to enroll seems to be made rationally.

Their findings are less sanguine, however, for choice of Part D plan. This survey offered individuals a choice of the standard plan described above versus alternatives that provide different levels of insurance coverage (e.g., catastrophic only, complete coverage, etc.), with corresponding actuarially fair premiums. They find that only about 36 percent of enrollees choose the cost-minimizing plan, and they do not place much value on the insurance aspects of more comprehensive plans. They conclude that “consumers are likely to have difficulty choosing among plans to fine-tune their prescription drug coverage, and do not seem to be informed about or attuned to the insurance feature of Part D plans.”

While this is an interesting set of findings, it provides only a preliminary look at the crucial issue of plan choice. These conclusions are based on data that do not contain precise details about the prescription drugs used by individuals; assumptions about utilization are made using aggregate imputations from other sources. Moreover, this is based on hypothetical choices across a set of nonexisting plans; individuals may become educated about the program when they are actually faced with plan choices. Thus, the failures of choice documented by Heiss, McFadden, and Winter (2006) may not hold when we use data on actual individual utilization and choices.

A recent paper by Claudio Lucarelli, Jeffrey Prince, and Kosali Simon (2008) uses aggregate data on plan market shares to conduct a study of how plan features affect demand and to undertake a welfare analysis of choice restrictions. They estimate sizeable welfare losses from limiting the option set facing seniors. But they do so in a framework that assumes that seniors are choosing optimally so that, by definition, restricting the choice set can only be harmful. Without individualized data on plan choices, they are unable to evaluate the underlying efficacy of plan choice.

Most closely related to our work is a recent field experiment by Jeffrey R. Kling et al. (2008). They examine how providing people with information about the relative costs of each of the available plans in 2007 computed using their 2006 claims affects their choices. They find that individuals who receive this intervention are more likely to switch plans, and more likely to end up with lower predicted and realized costs. Using our richer dataset on patient claims, we are able to model the individualized risk characteristics of plans, in addition to looking just at average costs. Our model is also more general in terms of sample and implications. While they investigate the consequences of one particular intervention on a sample of patients at a single hospital, our model allows us to calculate the potential welfare gains from reforms that change the structure of the choice set, and to do so for a large fraction of Medicare Part D enrollees.

II. Data

Our primary data source is a longitudinal sample of prescription drug records from the Wolters Kluwer (WK) Company. They are the largest “switch” operator in the prescription drug market: they collect the electronic claims from pharmacies and pass them on to the Pharmacy Benefit Managers (PBMs) and insurance companies

that will pay the claims. After adjudicating the claim, it is passed back through the switch to the pharmacy. WK performs this function for a large sample of pharmacies throughout the United States. Once pharmacies are in their sample, there is a 93 percent chance that they remain enrolled, so this is effectively a longitudinal sample of pharmacies. On average, the claims captured by the WK system represent almost 31 percent of all third-party prescription claims filled in the United States.⁸ The geographic distribution of these data is very closely representative of the geographic distribution of third party claims as well; the correlation between the WK market share and the overall third party market share across each of the states is 0.86.

WK keeps a longitudinal file that tracks prescription drug use for more than 100 million persons in the United States. They have made available to us for research purposes a longitudinal sample of prescription claims for any individuals age 65 and over in 2005.

These data are crucial because they are the only available data (of which we are aware) that contain information both on specific drug utilization by elders and on plan choice. Information about specific drug utilization is key because plan costs vary tremendously based on drug utilization, as we discuss below.

We begin with a sample of 1.53 million elders who:

- (i) Have a Part D claim with coverage of any sort (e.g., past the deductible);
- (ii) Are not employer-insured, dual eligibles, or eligible for low-income subsidies/partial subsidies;
- (iii) Have claims for only one region of the country;
- (iv) Have no claims with missing payment information;
- (v) Are in the sample of consistently reporting pharmacies; and
- (vi) Have data for both 2005 and 2006.

This data file has a rich set of information about every drug claim for individuals in the longitudinal sample, including information on: month in which the prescription was filled; county of location of the pharmacy; a depersonalized patient ID which allows longitudinal patient linkages; patient age; National Drug Code (NDC) for the drug; quantity measures (days supply, dosage, package size); patient and insurer payments; and insurer or PBM name. The total drug price is computed as the sum of patient and insurer payments; one or both of these values are blank for a small fraction of claims, and individuals with these claims are dropped. We constructed a crosswalk between the drug ID variable in the claims data and the NDC numbers listed on the formulary using a file from First Data Bank. To allow for the possibility of substitution if individuals were enrolled in a different plan, we constructed a coarser drug ID variable, which is unique only up to drug name and assumed that

⁸ Figure based on data provided by WK for 2006:III.

individuals could substitute to the cheapest drug with the same name were they enrolled in an alternative plan.

WK has created a sample for us that links longitudinally all claims from elders that fill prescriptions at a pharmacy in their sample. Thus, there are three types of attrition from the sample. First, elders may die (in which case we still observe all of their claims). Second, pharmacies may enter or leave the sample. This can be addressed by using only pharmacies that are continuously in their sample.⁹ Finally, individuals may switch pharmacies. If the switch is to a pharmacy within the WK sample, then the company does a detailed statistical match to ensure that the patient is captured and matched to other prescriptions (based on the deidentified form of data fields such as first name, last name, date of birth, year of birth, gender, health insurance ID, and zip code). If the switch is outside of the WK sample, then the individuals will be lost to this sample.

Unfortunately, there is no way to capture such transitions. So long as these transitions are not correlated with premiums, it will not bias our price elasticity estimate; so long as they are not correlated with plan cost-sharing provisions, they will lead only to a general understatement of out-of-pocket spending that will lead us to overstate that coefficient.

We can assess their importance by taking advantage of the fact that Wolters Kluwer provided us with a coverage level variable that indicates the proportion of pharmacies in each county covered by Wolters Kluwer. We have rerun our models on the 10 percent of counties where WK covers at least 40 percent of all third-party prescriptions, and our results are very similar to what we report below. This suggests that attrition is not significantly biasing our results.

A. The CMS Plans Database

We obtain information on availability of Part D plans and specific plan features directly from four files provided by CMS: the plan information file, the beneficiary cost file, the formulary file, and the geographic locator file. The plan information file lists plan names and identifiers, and regions/counties in which plans are offered. The beneficiary cost file contains copays and coinsurance rates for different tiers of each plan. The formulary file contains a list of all the drugs that are included on the formulary for each plan. The geographic locator file allows us to identify all the Social Security Administration (SSA) counties that correspond to different PDP and MA regions.

The major strength of the CMS data is that it allows us to fully parameterize any elder's plan choice set based on their location. We have used these data to build a "cost calculator" that mimics the calculator provided on Medicare's Web site. This calculator uses a given set of prescriptions for a given elder to compute their projected out-of-pocket spending in each plan available in their county.

⁹ A store is flagged as continuously enrolled provided that it does not miss more than 11 days (including weekends and holidays) of reporting in a month.

B. Matching Patients to Their Part D Plan

One challenging aspect of the WK data is that we know each patient's county and the name of the company that provides the Part D plan that is covering each prescription, but not specifically which Part D plan offered by that company is covering the prescription. For example, we know that an elder is covered by a Humana product, but not whether it is Humana Complete, Humana Enhanced, etc.

Fortunately, we can resolve this matching problem in most cases by using a combination of county code, company name, and copayment structure. For each claim and each of the plans within the same company offered in a particular county, we check if the copay that the patient paid for this claim matches any of the prescribed copays of the plan. We assign a person to a plan if most of the claims match to the same unique Part D plan. We carry out this exercise for each month. To confirm that a person has been matched to a correct Part D plan, we look at all the months together and insist that a person be consistently matched to the same plan in each month from June 2006 on, since enrollment into Part D plans was open until May 15, 2006.

Of the approximately 1.53 million individuals in our sample, 50.5 percent were matched to Part D plans. The remainder were excluded either because they had a large number of non-Part D claims (implying that they have some other form of coverage), because they had too few claims to reliably match, or because their copays were inconsistent with the copays listed for Part D plans in their region.

Of the matched individuals, 57.1 percent were uniquely matched to a Part D plan, and 42.9 percent were multiply matched (meaning that more than one Part D plan was consistent with their copays). While the unique matches are clear, excluding multiple matches leads us to misstate the proportion of enrollment in some plans. This problem is especially severe among Humana plans because Humana offers several plans, which differ only in the deductible and donut hole coverage, and thus cannot generally be distinguished on the basis of copays. While comprising 20 percent of all matches, Humana plans are only 10 percent of unique matches. To deal with this problem, we include both unique and multiple matches, with multiple matches randomly assigned to one of the plans to which they are matched with probability equal to the proportion of total national enrollment in that plan in 2006.¹⁰

C. Construction of Out-of-Pocket Cost Variables

The total enrollee costs of Part D can be decomposed into premiums, which are known for certain at the time of plan choice, and the distribution of out-of-pocket costs given the information available at the time when plans are chosen. Our focus is on estimating the distribution of costs given all the information potentially available to individuals at the time they make their choice. There are three reasons that estimating this distribution is challenging: first, we observe only realized out-of-pocket costs for the plan in which an individual is enrolled; second, we observe only a single realization of out-of-pocket costs for each individual (making it impossible

¹⁰ Regional enrollment figures are not available at the plan level in 2006 for most plans.

to compute a variance measure); and, third, we do not observe all the information available to individuals at the time they make their choice.

To handle the first difficulty, we assume that the set of 2006 claims is fixed and would remain constant had the individual in question chosen a different plan; that is, we assume no moral hazard. This assumption allows us to use the calculator to determine what each individual's realized costs would be for each plan in their choice set. Given typical estimates of the elasticity of prescription drug utilization in the range of 0.2 to 0.5, and considering that this would affect our results only to the extent that individuals have sufficient foresight to take into account future utilization effects in their plan choices, this is a fairly innocuous assumption, as shown in Web Appendix A.

To handle the second difficulty, we sample realized costs from 200 individuals who are "identical" to the individual in question at the time when the plan choice is made. In practice, we define "identical" as individuals with the same decile of 2005 drug expenditures, 2005 days supply of branded drugs, and 2005 days supply of generic drugs; after extensive searching, we found that this combination provided the best prediction of 2006 prescription drug spending based on 2005 characteristics. We therefore assign each individual to one of 1,000 cells demarcated by the interacted deciles of these measures. We restrict our sample to individuals for whom there are at least 200 other individuals in their cell, and we use these 200 individuals in each cell to compute both our rational expectations measure of utilization in 2006 (described below) and our variance measure.

The third difficulty is that individuals may actually know more than can be predicted given 2005 costs at the time when they make their plan choices. Intuitively, we can attempt to determine whether individuals know more than can be predicted given 2005 costs by analyzing whether their choices are sensitive to the component of the variation in realized costs across plans that cannot be predicted given 2005 characteristics. We discuss a model of this type in Section V.

D. Final Sample Creation

Under Part D individuals could enroll not only in a stand alone PDP plan, but also in a more comprehensive MA plan; we distinguish between individuals matched to MA and those matched to PDP plans based on copay, and exclude the former. We focus just on PDP plans (and therefore, just on individuals who chose PDP plans) because MA plans involve broader health care decisions which are beyond the scope of our data (e.g., regarding HMOs and fee-for-service plans). This exclusion is justified by the "independence of irrelevant alternatives" assumption that underlies our logit modeling, as discussed (and tested) further below. We also exclude individuals who have fewer than 500 observations in their state or fewer than 100 observations in their brand/state cell to increase the speed of estimation of the model by reducing the required number of brand/state fixed effects; this restriction has no effect on our final results.

Our final sample consists of 477,393 individuals. The typical patient in this sample is almost 75 years old, three-fifths are female, and they have an average of 34 claims per year. Their total prescription drug spending averages \$1,711 per year. While some individuals were enrolled in Part D for the full year, others enrolled as late as

May. The average total premiums paid after enrollment was \$287 and the average out-of-pocket costs paid out over the same period was \$666. This is the sample used in the efficient frontier analysis below. In our conditional logit models, we randomly subsample 20 percent of these individuals for computational reasons. We estimate the more computationally demanding random coefficients models on a randomly chosen subsample of 15,000 patients.

The distribution of enrollees across Part D plans is highly correlated in this final sample with the national facts on PDP enrollment provided by CMS. The correlation between the share by brand in our sample and the CMS sample is 0.98, and the correlation between the share of our sample in the top 10 plans is correlated with the CMS reported share in those plans at 0.89 (the correlation for the top 100 plans is 0.91).

III. Facts on Plan Choice

To motivate our regression framework, Figure 1 shows the basic facts on the relationship of plan choice to total plan costs. For each individual in the data, we estimate the total cost of enrolling in each PDP plan in their county, adding both premiums and expected out-of-pocket costs. We then estimate the difference in total costs between the plan chosen by that individual and the lowest-cost plan in their county. For this exercise, we use a perfect foresight model of expectations, using actual 2006 expenditures to estimate the costs that individuals face in each plan.

As Figure 1 shows, only 12.2 percent of individuals choose the lowest-cost plan in their state. Indeed, on average, individuals could save 30.9 percent of their total Part D spending by choosing the lowest-cost plan rather than the plan they chose. If we redo these calculations using actual 2005 expenditures, or predicted 2006 expenditures based on 2005 expenditures rather than actual 2006 expenditures, we find even stronger deviations from the lowest-cost plan.¹¹

Of course, individuals are not simply choosing a fixed payment stream when choosing a Part D plan; individuals who are highly risk averse may explicitly be choosing plans with higher mean expenditure to protect themselves against variance in expenditure. Yet this does not seem to be the case. Even if we include *only* plan choices where the variance is nonincreasing, over 70 percent of enrollees could have chosen a lower-cost plan, and the typical enrollee could have saved 23.3 percent of their Part D expenditures without raising their variance.¹²

The explanation for these facts is in Figure 2, which shows the choice set for individuals in California. The *x*-axis in this graph is the mean of total costs for each plan, and the *y*-axis is the average standard deviation in costs (where the standard deviation is computed using the 1,000 cell method, and the average is taken across individuals). In this graph, there is a clear “efficient frontier” of plans that dominate

¹¹ It appears that some plans may have offered low premiums in 2006 in order to entice consumers to choose their plan in the first year of the Part D program before raising their premiums in subsequent years. This behavior should not affect our analysis except insofar as there are large switching costs because consumers have the option to switch plans after each year, but one might still wonder to what extent the results above are driven by such plans. To assess this issue, we repeated the analysis above using the 2007 premiums for all plans and found that the average potential cost savings fell slightly from 30.9 percent to 25 percent.

¹² The fact that this number is smaller than the 30.9 percent number is because we are searching for cost savings over a small set of plans, not because individuals are especially sensitive to risk, a point we document further below.

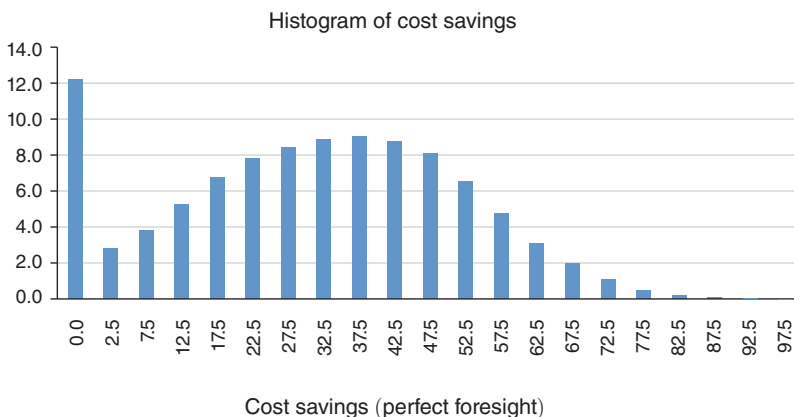


FIGURE 1. HISTOGRAM OF COST SAVINGS FROM SWITCHING TO LOWEST-COST PLAN (Percents)

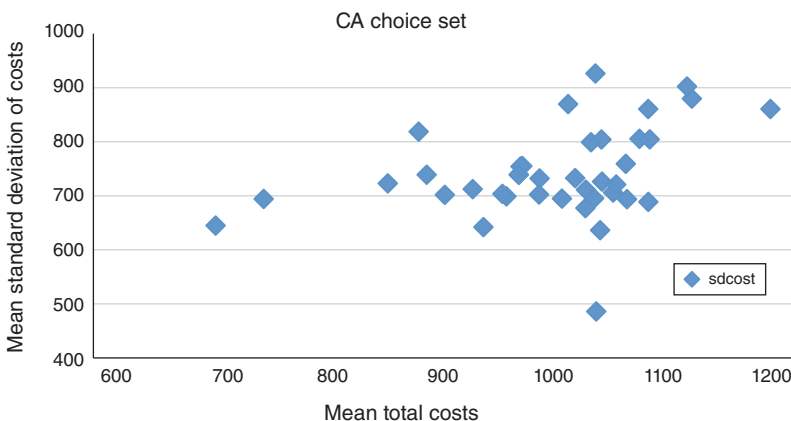


FIGURE 2. AVERAGE MEAN AND STANDARD DEVIATION FOR EACH PDP PLAN IN CALIFORNIA

others in terms of both cost and variance. This graph masks considerable heterogeneity across individuals: different plans lie on the efficient frontier for different individuals, so the fact that a plan lies off the efficient frontier in this graph does not imply that it is suboptimal for each individual. Nonetheless, most of the plans are well off the efficient frontier, meaning individuals could have lowered either their mean costs or their variance by picking a different plan.

As we will document below, one reason for the large amount of choice off the efficient frontier is that individuals consider plan characteristics in making their choices—but not how those plan characteristics matter for themselves. This is perhaps best illustrated by a simple examination of the decision to choose a plan with donut hole coverage. Figure 3 shows the probability of choosing donut hole coverage, and the financial implications of doing so, sorted by 2006 spending percentiles; the results are once again similar for other measures such as 2005 actual spending or 2006 predicted spending. The bottom line shows the percentage of the population at each percentile choosing donut hole coverage; the top line shows the average savings

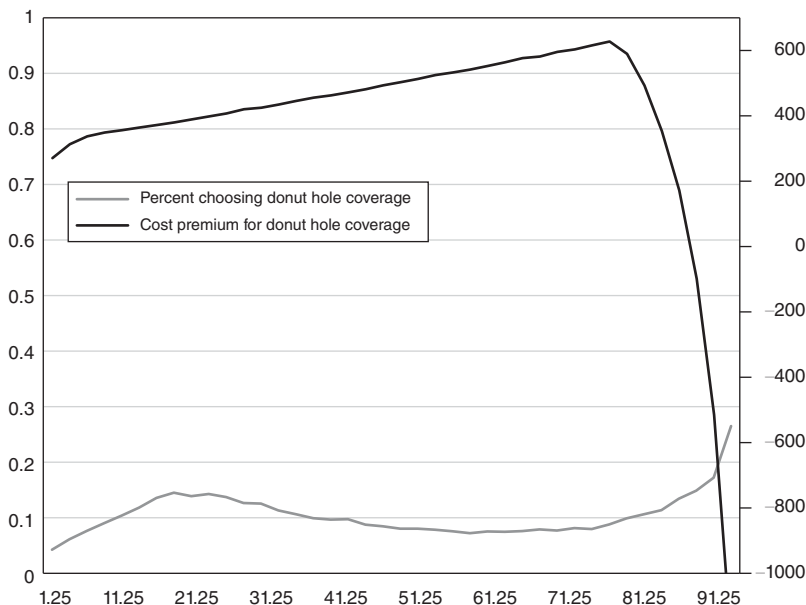


FIGURE 3. PERCENT CHOOSING DONUT HOLE COVERAGE AND ADDED COST BY EXPENDITURE QUANTILE

for individuals in that quantile from switching from the lowest-cost plan in their region that offers donut hole coverage to the lowest cost plan that does not.

The plans that offer donut hole coverage actually have slightly inferior coinsurances relative to the lowest cost non-donut hole plans in the initial coverage range, and so the cost of donut hole coverage is rising with expenditures until the point when individuals become likely to enter the donut hole.

The results here are striking: the percentage choosing donut hole coverage is virtually flat throughout the spending distribution at around 10 percent. Even if individuals are willing to pay extra in mean costs for the protection provided by donut hole coverage, it is hard to rationalize the fact that the same proportion of individuals in the tenth and eighty-fifth percentile of the spending distribution choose donut hole coverage.

IV. Base Model of Part D Plan Choice

In this section, we extend the efficient frontier analysis presented above by considering several discrete choice models. These models serve three general purposes in our setting. First, they allow us to control for additional plan characteristics such as plan quality. Second, they allow us to understand more precisely how preferences combine with choice set characteristics so we can forecast how individuals might choose in counterfactual choice environments. Third, they allow us to quantify the welfare consequences of choices.

We begin by specifying a constant absolute risk aversion (CARA) utility model with a normally distributed cost distribution:

$$(1) \quad U(C) = -\exp(-\gamma(W - C)) \text{ where } C \sim N(\mu, \sigma^2).$$

In this case, indirect utility is given by

$$(2) \quad u(\mu, \sigma^2) = EU(C) = -\alpha \exp\left(\gamma\mu + \frac{1}{2}\gamma^2\sigma^2\right),$$

where $\alpha = -\exp(\gamma W)$ is a constant. A first-order Taylor expansion about the point (μ', σ'^2) yields

$$(3) \quad u(\mu, \sigma^2) \approx u(\mu', \sigma'^2) - \alpha\gamma u(\mu', \sigma'^2)(\mu - \mu') - \frac{1}{2}\alpha\gamma^2 u(\mu', \sigma'^2)(\sigma^2 - \sigma'^2).$$

We can write total costs as $C = \pi + OOP$, and since π is known for any given plan, $\text{var}(C) = \text{var}(OOP) = \sigma^2$ and $\mu = E(C) = \pi + E(OOP) = \pi + \mu^2$. Adding an error term (and dropping constant terms), we can rewrite equation (3) as

$$(4) \quad u = -\alpha\gamma u(\mu', \sigma'^2)(\pi + \mu^*) - \frac{1}{2}\alpha\gamma^2 u(\mu^*, \sigma'^2)\sigma^2 + \epsilon.$$

This maps into a conditional logit model of plan choice where the utility of individual i from choosing plan j is given by:

$$(5) \quad u_{ij} = \pi_j\beta_0 + \mu_{ij}^*\beta_1 + \sigma_{ij}^2\beta_2 + \mathbf{x}_j\boldsymbol{\lambda} + \mathbf{q}_{b(j)}\boldsymbol{\delta} + \epsilon_{ij},$$

with $\beta_0 = \beta_1 = -\alpha\gamma u(\mu^*, \sigma'^2)$ and $\beta_2 = -\frac{1}{2}\alpha\gamma^2 u(\mu^*, \sigma'^2)$. In this equation \mathbf{x} represents any financial plan characteristics that affect choice, $\mathbf{q}_{b(j)}$ represents plan quality ratings and other nonfinancial aspects of plans (which vary only across brands), and ϵ_{ij} are i.i.d. type I extreme value random variables.

This gives us $\gamma = 2\beta_2/\beta_1$, which allows us to map the ratio of the coefficients on the variance of costs and the coefficient on the mean of costs into the coefficient of absolute risk aversion. This assumes that wealth is constant across all states of the world: the only risk facing individuals is uncertainty about the distribution of out-of-pocket costs. The same expression would hold if we added idiosyncratic risk that was uncorrelated with prescription drug expenditures, but it is not implausible that there would be correlated risks: in states where prescription drug expenditures are higher, other medical expenditures are higher as well. Such correlated risks would tend to bias upward our already low estimates of risk aversion.

We include in our model several financial plan characteristics beyond premiums, out-of-pocket costs, and the variance of out-of-pocket costs. These are: the deductible of the plan; a dummy for whether the plan covers all donut hole expenditures; a dummy for whether the plan covers generic expenditures in the donut hole only; and a cost-sharing index. The cost sharing index is calculated for each plan as the average percentage of expenditures covered by the plan between the deductible and the donut hole. This variable differs from expected out-of-pocket costs in that it has the same value for everyone in the sample for each plan, and because it is not directly impacted by whether plans have deductibles or donut hole coverage. We also include two measures of plan quality: the share of the top 100 drugs used by elders that is included in the plan's formulary and a quality index. The quality index is computed by CMS on a 1–5 scale by aggregating consumer ratings at the brand

level collected along 17 dimensions which are categorized as “customer service,” “drug pricing information” (availability/rate of price changes), and “using your plan to get your prescription filled.”

Identification is a natural concern in this context. All of the plan characteristics included in our model may be endogenous due to unobserved demand factors, and they may be biased by correlation with unobserved plan characteristics. To address this concern, we observe and include in our model all of the publicly available information that might be used by individuals to make their choices. We also consider models where we control for a full set of brand dummies, as well as a full set of interactions of state dummies with brand dummies. When we include brand dummies, the coefficient on the quality index (which is measured at the brand level) is no longer separately identified, although it can be recovered by a generalized least squares (GLS) regression of these dummies on the quality variable. When brand-state dummies are included, coefficients on plan characteristics such as the premium, deductible, and donut hole coverage are identified by the variation across plans offered by the same brands in a given state.¹³

Even with these fixed effects, it is possible that premiums are endogenous because they are set based on brand–state specific assessments of demand conditions. If premiums are higher in regions where insurers anticipate more demand for their particular plan (relative to other plans offered by the same insurer), our estimate of the coefficient on premiums will be biased toward zero, since individuals will appear to be less averse to higher premiums. To the extent that these factors make high premiums appear less undesirable than they actually are, our conclusion that premiums are overweighted relative to out-of-pocket costs would be strengthened, as would our estimates of the welfare loss due to consumer mistakes.¹⁴

A. Restrictions on Preferences

The model laid out above suggests three natural restrictions on preferences which extend the efficient frontier concept to the discrete choice setting.

RESTRICTION 1: $\beta_0 = \beta_1$.

This restriction states that the coefficient on premiums should equal the coefficient on expected out-of-pocket costs. Controlling for the risk characteristics of plans, individuals should be willing to pay exactly one dollar in additional premiums for coverage, which reduces expected out-of-pocket costs by one dollar. If this restriction fails to hold, individuals are not choosing on the efficient frontier: they

¹³ For instance, in many states Humana offers a Standard Plan with lower premiums but limited coverage, an Enhanced Plan with higher premiums but no deductible, and a Complete Plan, which offers superior cost sharing and full donut hole coverage at much higher premiums.

¹⁴ We did attempt to estimate the models reported below using two instruments and the control function approach: these were the average premium for a given plan in all states where the plan is offered (designed to avert local demand shocks) and a “marginal cost” instrument constructed using the average covered expenditures for individuals enrolled in the plan. In both cases, the magnitude of the coefficient on premiums increased in the IV models. We are not confident that the exclusion restriction is satisfied for either of these instruments, so below we continue to estimate the model without an instrument.

could switch to alternative plans with comparable risk characteristics but lower total costs.

RESTRICTION 2: $\lambda = 0$.

This restriction states that financial plan characteristics other than premiums, expected out-of-pocket costs and the variance of out-of-pocket costs do not affect choices. Individuals should not care about deductibles, donut hole coverage, or copays *per se*; they should care only about these factors to the extent that they affect the distribution of out-of-pocket costs. Once we control for this distribution, these factors should be redundant.

RESTRICTION 3: $\beta_2 < 0$.

This restriction states that individuals should be risk averse.

While these restrictions follow naturally from utility maximization with full information and standard preferences, the model from which they are derived makes several important functional form assumptions: we assume that the distribution of out-of-pocket costs can be summarized by its mean and variance, that indirect utility is a linear function of this mean and variance, and that the errors are i.i.d. type I extreme value. In Appendix A, we show that the restrictions assumed in the previous section still hold, even when these functional form assumptions are weakened.¹⁵ Of course, it is always possible to write down preferences that would violate the restrictions above, but these restrictions are generally compatible with commonly used expected utility functions given the observed cost distributions.

V. Results

In this section we present the results from the conditional logit model described above. At the outset, it is important to mention that the results we present here are not very sensitive to misspecification or measurement error. Online Appendices A to C (available at <http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.4.1180>) investigate these issues extensively via simulation. We take the distribution of realized costs observed in the data for each of the 1,000 cells, assume it is the true distribution, and simulate individuals' choices using known utility functions by assuming individuals' maximize expected utility. We then investigate to what degree the restrictions above on choice are violated if we add various types of known misspecification or measurement error. For example, if the true utility function is CARA with risk aversion 1, does estimating the model in the logit framework with linearized indirect utility generates choice inconsistencies? What if we only observe a noisy measure of out-of-pocket costs which contains attrition and measurement error?

¹⁵ In particular, we simulate choices using the actual distribution of costs and several commonly used utility functions (constant relative risk aversion (CRRA), CARA) with varying levels of risk aversion. In some cases, the restrictions do not hold exactly, but the violations are much smaller in magnitude than we observe when we estimate the model using actual choices.

The upshot of this analysis is that with very large risk aversion or substantial amounts of measurement error, we do sometimes observe statistically significant choice inconsistencies, but these are always smaller in magnitude than we report below and have inconsistent signs. In our discussion below, we make this comparison more explicit.

A. Base Results

Table 1 reports the results from several conditional logit models. Model (1) includes only the premium, realized out-of-pocket costs, the variance of out-of-pocket costs, and the quality variables. As noted in the discussion of the cost variables, expected out-of-pocket costs—meaning the individual’s expectation of out-of-pocket costs at the time of plan choice—are not directly observed, so we use realized costs as a proxy for expected out-of-pocket costs. This proxy has noise (where “noise” includes the component of realized costs unknown to the individual at the time when the choice is made) and so its coefficient is biased downward. We address this problem at length in the next section and show that it does not much impact our conclusions.

The cost variables—premiums and out-of-pocket costs—are measured in hundreds of dollars. Model (1) therefore shows that a \$100 increase in premiums leads to a 32 percent reduction in the probability that a given plan is chosen, implying an average elasticity of -0.75 .¹⁶ There are two ways to interpret the remaining coefficients. First, we can divide by the premium coefficient in order to compute the willingness to pay in dollars for a one unit increase in the characteristic. Second, the coefficient itself can be interpreted as the percentage increase in the probability that a plan is chosen from a one unit increase in the characteristic provided that probability is small (as it is for most plans). When we compare models estimated using actual choices with simulations, we compare the implied dollar value of plan characteristics computed by dividing by the coefficient on premiums. This is because the scale of the coefficients is determined by the proportion of choices explained by the included variables, and in the simulations, this scaling factor is arbitrarily set by whatever standard deviation we assume for the structural error term. We can and do use the simulations to compare ratios of coefficients with the actual results, but nothing substantive can be inferred from the absolute magnitude of the coefficients in the simulation. Two points about the model (1) results are noteworthy. First, the coefficient on out-of-pocket costs is only about $\frac{1}{2}$ as large as the coefficient on premiums, violating Restriction 1. Second, the coefficient on the variance term is negative and significant, but extremely small, implying risk aversion substantially less than we obtained in the simulations with $\text{CRRA} = 1$.

Model (2) adds additional covariates to control for deductibles, donut hole coverage, average cost sharing, formulary coverage, and plan quality. Many of these covariates enter the model with significant coefficients. When we add plan characteristics, the coefficient on premiums increases, suggesting that it was initially biased downward due to omitted variable bias. The coefficient on the variance term drops even farther once we add a control for the number of the most popular 100 drugs

¹⁶ The implied elasticity varies across plans based on premium level and market share. The “32 percent” number given in the text is derived from the equation $(\partial \log(p_{ij}) / \partial x_{ij}) = (1 - p_{ij})\beta$. Thus, for $p_{ij} \approx 0$ which holds for a large number of plans, we can interpret β as the percentage change in p_{ij} associated with a one unit change in x_{ij} .

TABLE 1—CONDITIONAL LOGIT RESULTS

	(1)	(2)	(3)	(4)
Premium (hundreds)	-0.4330*** (0.0029)	-0.7663*** (0.0038)	-0.4990*** (0.0061)	-0.5218*** (0.0069)
OOP costs (realized) (hundreds)	-0.2127*** (0.1520)	-0.1172*** (0.0015)	-0.0961*** (0.0015)	-0.0967*** (0.0016)
Variance (times 10 ⁶)	-0.0189*** (0.0027)	-0.0004 (0.0007)	-0.0006 (0.0007)	-0.0005 (0.0007)
Deductible (hundreds)	x	-0.2899*** (0.0049)	-0.1628*** (0.0067)	-0.1674*** (0.0072)
Donut hole	x	3.023*** (0.0181)	1.762*** (0.0277)	1.865*** (0.0303)
Generic coverage	x	0.4203*** (0.0140)	0.3004*** (0.0175)	0.2700*** (0.0177)
Cost sharing	x	3.282*** (0.0538)	1.189*** (0.0741)	1.057*** (0.0778)
Number of top 100 on form	x	0.0937*** (0.0007)	0.0587*** (0.0017)	0.0644*** (0.0018)
Average quality	0.4091*** (0.0032)	0.7398*** (0.0039)	x	x
Brand dummies	No	No	Yes	No
Brand-state dummies	No	No	No	Yes
Risk index	10	0	0	0
Number of patients	95,742	95,742	95,742	95,742
Number of plans	702	702	702	702
Number of states	47	47	47	47
Number of brands	36	36	36	36

Notes: The table shows conditional logit results from estimating the model given in equation (6) by maximum likelihood. Each column shows coefficients from a single regression. The coefficients reported are the parameters of the utility function, not marginal effects. Standard errors are in parentheses. The first column includes only premium, realized out-of-pocket costs, and the variance measure. The second column adds controls for the indicated plan characteristics, the third column adds brand fixed effects, and the fourth column adds brand-state fixed effects. Premiums, out-of-pocket costs, and deductibles are in hundreds of dollars and the variance term is in millions. The cost-sharing variable is computed as the average value of covered expenditures divided by total drug expenditures for individuals in the choice set. The average quality variable is a normalized version of the “average rating” index provided by CMS. The risk index is twice the coefficient on the variance divided by the coefficient on premiums scaled by 100. In the model in the text, this value equals one million times the coefficient of absolute risk aversion.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

included in the plan’s formulary. One explanation is that, while individuals prefer plans that cover more drugs, they do not have sufficient foresight to choose plans that cover drugs they (or at least people in their cell) might need in the future but are not already taking. Alternatively, it may be that there is substantial measurement error in the variance term, and that number of top 100 drugs is a proxy for the variance. Models (3) and (4) add brand dummies and brand-state dummies, respectively. The coefficient on premiums actually shrinks once we include brand-dummies, but the effects of the premium remain large; a \$100 increase in annual premiums leads to a 50 percent reduction in the probability that a plan is chosen, corresponding to an average elasticity of -1.17 . The coefficient on out-of-pocket costs has similar magnitude across all of the models, which reflects the fact that it is identified based on individual variation. In columns 3 and 4, the coefficient on the premium is more than five times as large as the coefficient on out-of-pocket costs.

The coefficients on plan characteristics are also very large in all specifications. Controlling for the out-of-pocket cost consequences, model (4)—which has the smallest plan characteristics—suggests that individuals are willing to pay over \$300 for full donut hole coverage, \$50 for generic donut hole coverage, about \$80 to go from a deductible of 250 to a deductible of 0, about \$80 to go from the plan with the least cost sharing (25 percent) to the plan with the most cost sharing (65 percent), and \$12 for each of the top 100 drugs that appear on the formulary. These numbers are not enormous, but they are an order of magnitude larger than the results in the simulations,¹⁷ and have nontrivial consequences for the welfare evaluation of plan choice, as we investigate in the welfare analysis section.

It is important to underscore the fact that these numbers are not the full hedonic value of those plan characteristics—these are the willingness to pay above and beyond the implications of those plan characteristics for out-of-pocket costs.¹⁸ Because individuals appear to be underweighting the individualized component of out-of-pocket costs, we can interpret these numbers as saying: individuals are willing to pay a price in premiums for desirable plan characteristics, but this price is insufficiently sensitive to their individual circumstance.

We noted above that in some of the Appendix simulations, plan characteristics had statistically significant coefficients, even controlling for out-of-pocket costs, due to (imposed) measurement error or misspecification of the utility function. We might worry that the results using actual choices likewise reflect these factors rather than choice inconsistencies. The coefficients on plan characteristics estimated using actual choices imply larger dollar values for plan characteristics than do the simulations, however. For example, our estimate of the implied value of donut hole coverage controlling for out-of-pocket costs is larger than the average out-of-pocket cost savings from donut hole coverage observed in our data, so even with infinite measurement error in these costs, the simulations could not match what we observe.

Thus, formal modeling of choice reveals a violation of all three of the preference restrictions we laid out above. The coefficient on premium is an order of magnitude larger than the coefficient on out-of-pocket expenditures; generalized plan characteristics enter the model highly significantly, even conditional on individual out-of-pocket risk; and individuals are not willing to pay more for plans with lower variance in expected spending.

One potential shortcoming of our last conclusion, however, is the reliability of our variance measure. We compute the variance by assessing the variability in

¹⁷ For comparison, in the simulations where true utility is CRRA with risk aversion of 10 (misspecification is increasing in risk aversion), we estimate—controlling for out-of-pocket costs—a \$9 value for full donut hole coverage, a -\$8 value for generic donut hole coverage, a \$33 value of moving from a \$250 deductible to a 0 deductible, a \$32 value of going from the plan with the least cost sharing to the most cost sharing, and -\$1 for each of the top 100 drugs that appear on the formulary. (Since these values are driven entirely by misspecification in the simulations, there is no reason the signs should be sensible.)

¹⁸ To recover the hedonic value of plan characteristics, it would be necessary to add the values reported below to the values of plan characteristics implicit in our out-of-pocket cost measure. We can attempt to recover these values by regressing our out-of-pocket cost measure on plan characteristics controlling for individual fixed effects. This procedure will give biased results to the extent that plan characteristics not included in the regression affect out-of-pocket costs, so we try not to lean heavily on this exercise when interpreting the results above. Nonetheless, we report the results of this regression for reference: \$88 increase in out-of-pocket costs for a \$250 deductible, \$99 decrease in out-of-pocket costs for full donut hole coverage, \$12 decrease for generic donut hole coverage, \$810 decrease moving from cost sharing of 0 percent to cost sharing of 100 percent (or \$324 moving from the 25th percentile plan to the 75th percentile), and a \$1.3 decrease for each of the top 100 drugs covered.

out-of-pocket spending across a sample of similar individuals; we have tried several alternative specifications of risk preferences (such as including quantiles in the right-tail of the distribution of costs), and this does not appear to alter our results. We have also constructed alternative “cells” for a subset of the sample, which take into account information on the particular drugs each individual uses, and have found that our current method captures more than 90 percent of the variation in the variance term across plans. It is not a foregone conclusion that the coefficient on our variance variable would be zero even if elders do not explicitly consider the cost of alternative plans under a range of hypothetical outcomes: if patients with a greater risk of getting sicker in the following year also chose plans with more coverage, we might expect this to show up in the variance variable, even after we control for the average value of plan characteristics. The result appears to buttress our finding that individuals are insensitive to the individualized consequences of plan choice. Nevertheless, we place less weight on this last choice inconsistency because of concerns about the appropriate specification and measurement of risk in our setting, some of which are discussed further in the robustness section.

B. Robustness

While our unique dataset makes this analysis possible, the data do have a number of shortcomings. In this section we show the robustness of our basic conclusions to efforts to address these shortcomings.

First, we are able to match only 50.5 percent of our sample of 1.53 million individuals to Part D plans, partly because we use very stringent criteria designed to minimize false matches. The cost of such an approach is that our matched sample may not be representative of the full sample of 1.53 million; in particular, the individuals in our sample have more claims because that makes it easier to match them to a plan. We therefore consider a less stringent matching strategy; in this less stringent match, we consider a copay as matched for a given claim if the appropriate copay appears *anywhere* on the claims formulary, even if the copay listed for drugs in that particular tier and days supply is not correct. We also accept matches if just 50 percent of overall claims are matched rather than requiring that this threshold be exceeded in every month after the first Part D claim is observed. Using this strategy, we are able to match 1.28 million of our 1.53 million individuals, or 84 percent.

The first column of Table 2 shows the robustness of our findings to this alternative measure. None of our main conclusions is altered. Our coefficients of interest are somewhat smaller than in the original sample, which may be because of noise introduced by being less stringent in our willingness to accept matches, but the fundamental choice inconsistencies persist.

Second, a limitation of our approach is that we exclude individuals with low numbers of claims, since our matching algorithm requires enough claims to identify the plan copayment structure. This clearly leads to bias to our variance measure, which is constructed by selecting matched individuals in the same cell; this is yet another reason why we have less confidence in the variance results than in our results for other forms of choice inconsistency. But there is no reason why this should lend a particular bias to our other results; if anything, we might think that failing to correctly specify the variance would make low out-of-pocket cost plans seem more desirable (since

TABLE 2—ROBUSTNESS CHECKS

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Premium (hundreds)	-0.4099*** (0.0070)	-0.5717*** (0.0076)	-0.5099*** (0.0074)	-0.4935*** (0.0072)	-0.5517*** (0.0078)	-0.5220*** (0.0070)	-0.7422*** (0.0043)
OOP costs (realized) (hundreds)	-0.0508*** (0.0016)	-0.0808*** (0.0018)	-0.0846*** (0.0018)	-0.0977*** (0.0016)	-0.0986*** (0.0016)	-0.0944*** (0.0015)	-0.1100*** (0.0015)
Variance (times 10 ⁶)	-0.0009 (0.0010)	-0.0067*** (0.0029)	-0.0011 (0.0014)	-0.0003 (0.0007)	0.0002 (0.0012)	-0.0007 (0.0008)	-0.0004 (0.0008)
Deductible (hundreds)	-0.0456*** (0.0072)	-0.2680*** (0.0070)	-0.2179*** (0.0068)	-0.4378*** (0.0095)	-0.4876*** (0.0100)	-0.1723*** (0.0072)	-0.2730*** (0.0051)
Donut hole	1.224*** (0.0302)	2.023*** (0.0299)	1.907*** (0.0296)	2.872*** (0.0346)	3.956*** (0.0380)	1.837*** (0.0304)	2.935*** (0.0204)
Generic coverage	-0.0241 (0.0158)	0.1168*** (0.0173)	0.1345*** (0.0296)	0.7570*** (0.0233)	0.8362*** (0.0265)	0.2597*** (0.0178)	0.4206*** (0.0141)
Full cost sharing	0.9973*** (0.1313)	2.129*** (0.1316)	1.319*** (0.1295)	1.357*** (0.0838)	-0.0986 (0.0897)	1.076*** (0.0781)	3.633*** (0.0555)
Number of top 100 on form	0.0737*** (0.0013)	0.0664*** (0.0015)	0.0599*** (0.0014)	0.1173*** (0.0023)	0.1126*** (0.0022)	0.0654*** (0.0018)	0.0952*** (0.0007)
Customer service	x	x	x	x	x	x	0.1644*** (0.0066)
Prescription filling	x	x	x	x	x	x	0.5666*** (0.0037)
Pricing avail./changes	x	x	x	x	x	x	0.4142*** (0.0054)
Missing quality dummies	No	No	No	No	No	No	Yes
Brand dummies	No	No	No	No	No	No	No
Brand-state dummies	Yes	Yes	Yes	Yes	Yes	Yes	No
Risk index	0	0	0	0	0	0	0
Number of patients	99,999	99,981	99,981	99,988	99,999	99,994	95,742
Number of plans	888	861	861	702	702	702	702
Number of states	49	49	49	47	47	47	47
Number of brands	48	42	42	36	36	36	36

Notes: Table shows conditional logit results from estimating the model given in equation (6) by maximum likelihood on different subsamples to check robustness. Each column shows coefficients from a single regression. The coefficients reported are the parameters of the utility function, not marginal effects. Standard errors are in parentheses. Premiums, out-of-pocket costs and deductibles are in hundreds of dollars and the variance term is in millions. The cost-sharing variable is computed as the average value of covered expenditures divided by total drug expenditures for individuals in the choice set. The first column estimates the model on a random sample of 100,000 patients selected from the sample of individuals who were matched to a PDP plan in the “Lax match” discussed in the text. The second column estimates the model on the original sample used in Table 1 plus individuals with zero claims in 2005. The third column uses the same sample as the second column, but dropping individuals with zero claims in 2005, and more than 12 claims in 2006. The fourth column includes only unique matches and multiple matches which could be assigned with 95 percent certainty. The fifth column includes only unique matches. The sixth column uses the original sample but restricting to individuals for which more than 50 percent of their claims were non-dual. The seventh column disaggregates the quality variable. Note that the seventh column also does not include brand or brand-state dummies, since these are collinear with the quality variables. The risk index is twice the coefficient on the variance divided by the coefficient on premiums scaled by 100. In the model in the text, this value equals one million times the coefficient of absolute risk aversion.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

they also have lower variance), thus biasing upward the coefficient on out-of-pocket costs. If this were the case, our results would be too conservative in reporting mistakes.

The exclusion of individuals with a small number of claims also raises selection issues, since these individuals may make better choices or have systematically different price elasticities. To address this problem, we have reestimated our model including additional individuals whom we have excluded thus far. First, assuming some serial correlation in claims behavior, we can mimic the inclusion of low claims

individuals in 2006 by including those with no claims in 2005 (we have excluded them thus far to allow for the creation of our variance measure and the rational expectations measure used below). We reran our base-case discrete choice model using a random sample drawn from the larger sample, including individuals with zero claims in 2005. In our original sample, the variance variable was constructed by assigning each individual to one of 1,000 cells based on 2005 claims. Because all these individuals are identical in 2005, we assigned all of them to a single cell, and computed the variance by running 200 randomly chosen individuals in that cell through every plan. The results of this analysis are shown in column 2, and they do not differ much from our original results.

In column 3, we extend this analysis further by restricting the analysis to those who have zero claims in 2005 and fewer than 12 claims in 2006; this is the closest we can get to the zero claim sample in 2006 while still matching plans. The results are once again quite similar.

Third, we include in our analysis both unique matches and multiple matches, imputing the latter based on market shares. This adds a degree of noise to our estimation that could plausibly bias the results. We address this issue in two ways in Table 2. First, we reestimate the model including only those multiple matches where we can identify the plan with 95 percent certainty based on the relative enrollment given by CMS for the matched plans in the patient's state (e.g., one of the plans has at least 19 times the enrollment of the other matched plans combined). Including unique matches, this was 75 percent of the original sample. Going further, in column 5, we reestimate the model only with those observations for which we can make a unique match. Neither of these changes alters any of our main conclusions.

Fourth, we do not know with certainty which individuals in our sample are dual eligibles. If we mistakenly include dual eligibles in our sample, we will measure them as having a lot of variation in out-of-pocket costs when they really have none, so we wrongly interpret them as insensitive to these costs. We are fairly confident, however, that we are excluding dual eligibles from our sample because they have such a limited range of possible copayments. For example, for 90 percent of the observations we use, more than one-quarter of their claims have copayments that are inconsistent with being a dual eligible (e.g., more than \$5). To further ensure that problems identifying dual eligibles were not biasing our results, we have reestimated our model only on individuals where at least 50 percent of their claims are inconsistent with the copayment rates for duals. As we show in column 6 of Table 2, this has little impact on our results.

Finally, in column 7, we show the impacts of decomposing further the aggregate quality index that we have used in our work thus far. We decompose the index into its three primary components. Doing so, we find that choice is positively associated with each of these quality components. The most important characteristic of quality to consumers appears to be the ease in filling prescriptions. Most importantly, decomposing the quality measure has no impact on our results.

VI. Misspecification and Measurement Error

In the previous section, we presented results from a conditional logit model of plan choice and identified three apparent irregularities in choices. Our interpretation

is that these results reflect consumer errors—plan characteristics are more salient than are their implications for the distribution of out-of-pocket costs, and individuals are unable to compute the individualized risk characteristics of the alternative plans. In this section we consider two related alternative explanations: that we have misspecified out-of-pocket costs because we have failed to appropriately model the information available to individuals at the time when they make their plan choice; and that our findings are driven by measurement error in out-of-pocket costs.

A. Modeling Private Information

Thus far we have measured out-of-pocket costs using the realized cost measure constructed from 2006 claims. We label an alternative measure that we consider in this section our “rational expectations” measure. Recall that to create our variance measure we classified all individuals into 1,000 cells defined by deciles of 2005 total spending, generic prescriptions, and branded prescriptions, and ran the 2006 claims of 200 persons in each cell through the cost calculator for that plan. This procedure generates a distribution of costs for each patient and plan. Our rational expectations measure is defined as the mean of this distribution. Under the strong assumptions discussed above (CARA utility and a normal distribution of costs), the mean and the variance would completely summarize the impact of the cost distribution on utility; our simulations in Appendix A show that they summarize this distribution well, even if these assumptions are relaxed.

It is useful to compare this rational expectations measure to the perfect foresight/realized costs measure we have been using. The latter measure is “too broad” in the sense that it includes information not available to individuals at the time when they choose (provided, that is, that they do not know exactly what their drug needs and drug prices will be for the coming year). The former measure is “too narrow” in the sense that individuals may have private information at the time they choose beyond what can be inferred from their 2005 costs. If a patient learns he has cancer just prior to choosing the 2006 plan, he would correctly forecast that his drug needs would likely exceed the average of those with similar 2005 spending.

We address these concerns by developing a model with which we can identify the information available to consumers at the time when they choose. The intuition behind this model is that we can determine if individuals know more than we can predict given just their 2005 spending by evaluating whether their plan choices are responsive to the component of 2006 spending which is not known in 2005.

The formal derivation of this model is presented in Appendix B. In summary, we augment our model in two ways. The first is to include a normally distributed term that captures the degree of private information: the difference between actual out-of-pocket spending in 2006 and what we would have predicted for 2006 based on our rational expectations model. If we were estimating a linear model, this would be comparable to estimating our model by instrumental variables, where we instrument the perfect foresight level of costs with our rational expectations cost measure, which is independent of private information. In our nonlinear setting, the comparable correction is to include this noise term, which essentially amounts to estimating a random coefficients model with one extra parameter to identify the degree of private information.

TABLE 3—RESULTS WITH PRIVATE INFORMATION

	Final sample	Restricted sample	Private info
Percent private information			0.5818*** (0.0618)
Premium (hundreds)	-0.7383*** (0.0038)	-0.7156*** (0.0094)	-0.7489*** (0.0132)
OOP costs (realized) (hundreds)	-0.1169*** (0.0016)	-0.1040*** (0.0039)	-0.1687*** (0.0094)
Variance (times 10 ⁶)	-0.0026 (0.0014)	-0.1103** (0.0517)	-0.8574*** (0.2947)
Deductible (hundreds)	-0.2677*** (0.0014)	-0.3079*** (0.1257)	-0.2767*** (0.0137)
Donut hole	2.823*** (0.0181)	2.805*** (0.0490)	2.870*** (0.0478)
Generic coverage	0.3066*** (0.0143)	0.4743*** (0.0341)	0.4784*** (0.0347)
Full cost sharing	2.990*** (0.0546)	3.391*** (0.1417)	2.829*** (0.1743)
Number of top 100 on form	0.0939*** (0.0007)	0.0995*** (0.0019)	0.1005*** (0.0021)
Average quality	0.7167*** (0.0039)	0.7418*** (0.0098)	0.7512*** (0.0095)
Brand dummies	No	No	No
Brand-state dummies	No	No	No
Risk index	1	31	229
Number of patients	95,742	15,001	15,001
Number of plans	702	702	702
Number of states	47	47	47
Number of brands	36	36	36

Notes: The table compares conditional logit results with results from estimating the random coefficients model given in equations (11) and (14) using the Laplace approximation to the likelihood function developed by Matthew C. Harding and Jerry A. Hausman (2007) with bootstrapped standard errors. Each column shows coefficients from a single regression. The coefficients reported are the parameters of the utility function, not marginal effects. Standard errors are in parentheses. The first column is identical to the second column of Table 1. The second column estimates the same model on a random subsample of 15,000, and the third column estimates the random coefficients model on this same subsample. Variable definitions are identical to Table 1. The “percent private information” field corresponds to the variable T_{frac} in the model in the text.

*** Significant at the 1 percent level.

** Significant at the 5 percent level.

Second, the measured variance from the 1,000 cell exercise overstates the true variance in costs because some of this variation represents variation in realized costs which is unpredictable based on 2005 costs but is known to the individual at the time when they choose. We develop a correction for the variance based on the estimated degree of private information.

Table 3 reports the results from estimating this model. For computational reasons, we estimate this model on a much smaller sample by randomly selecting 15,000 patients from our earlier sample. Column 1 reports our earlier results, column 2 reports the earlier specification on the new sample, and column 3 the results from adding the term for private information and correcting the variance. The model is estimated using the Laplace approximation developed in Harding and Hausman (2007) with bootstrapped standard errors, including controls for the various plan characteristics.

The results in Table 3 suggest that there is substantial private information: individual choices take into account about 60 percent of the variation in out-of-pocket costs that cannot be predicted given their cell. We also continue to find that the coefficient on realized costs is well below that on premiums, and that financial plan characteristics such as the donut hole and deductible continue to enter highly significantly in this model. Therefore, two of the major choice inconsistencies persist even when we model private information.

This model implies, however, that individuals know much of what their costs will be to each plan in their choice set in the coming year so there is little insurance motive. Under this interpretation, the variance in out-of-pocket costs is small for all plans because there is little uncertainty. This means that any measured response to the variance term would imply high levels of risk aversion, and that the standard errors in our estimates of risk aversion are much larger than we concluded in the model ignoring private information. The risk index in these models (obtained by dividing the variance coefficient by the premium coefficient and multiplying by 200) is comparable to what we obtained in our Appendix B simulations for $CRRA = 3$ with $wealth = 17,000$.¹⁹

The bottom line from our models of private information is that our conclusions about the gap between the premium and out-of-pocket expenditure coefficients, and the powerful role for general plan financial characteristics in driving choice, are robust to a wide variety of specifications of out-of-pocket spending risk. Our conclusion about the low degree of estimated risk aversion, however, is more sensitive to the precise specification of the model.²⁰

B. More General Measurement Error

The private information model can also be interpreted as correcting for a specific form of measurement error in our model: that arising from idiosyncratic variation across individuals in their knowledge about expected out-of-pocket costs at the time they choose their Part D plan. Our private information model is the nonlinear equivalent to a linear model that addresses measurement error in 2006 realized costs by instrumenting them by predicted costs based on 2005 characteristics. The fact that our conclusions are robust to controlling for private information is therefore equivalent to saying that instrumenting for idiosyncratic measurement error across individuals does not change our conclusions. However, idiosyncratic measurement error is only one of several types of measurement error in our out-of-pocket cost coefficient. In this section we consider robustness to alternative forms of measurement error.

¹⁹ As we highlight in Abaluck and Gruber (2009), the results reported in Table 2 are also consistent with an alternative model of information where individuals are not using all available information, but rather are paying attention only to a part of their prescription drug expenditures. For that portion to which they are attentive, individuals are rationally weighting premiums and out-of-pocket costs in the same way in making their decision. Yet individuals do not respond to variation in out-of-pocket costs beyond that portion. Under this interpretation, we find that the degree of private information is smaller and the coefficient on the risk measure once again becomes very small.

²⁰ To further explore robustness here, we consider an alternative measure of predicted out-of-pocket costs: predicting those costs based only on use of "regular drugs." A regular drug is defined as any drug for which the individual consumed at least 90 days' supply in 2005. In our new measure, we construct the out-of-pocket cost variable assuming that drug use in 2006 will consist only of regular drugs in 2005. Our results are very similar using this alternative measure.

To model the impact of measurement error, we draw on the simulation model developed in Appendix A, which captures the “no inconsistency” baseline. Without measurement error, this model illustrates that choices under a variety of specifications of risk would not demonstrate the inconsistencies we see in our data. We can augment that analysis by adding measurement error in out-of-pocket costs to this simulation model, and using predicted costs as our measure of out-of-pocket costs in the simulation. We consider three alternative specifications of measurement error, and present the detailed results of our analysis in Appendix C.

The first is purely idiosyncratic individual-specific measurement error in forming expectations of out-of-pocket costs. Consistent with the discussion above, even with very large error of this form, we find that our simulated out-of-pocket and premium coefficients are similar, and the coefficients on plan-specific plan characteristics are very small; that is, our predicted out-of-pocket cost is effectively acting as an instrument for measurement error in this case.

We then consider a form of multiplicative error designed to mimic what might be observed if there were attrition due to patients having claims at pharmacies not included in our data. Our simulations then show that even sizeable attrition bias causes only a small upward bias to the out-of-pocket coefficient, and causes only very modest coefficients on the plan characteristics, an order of magnitude smaller than what we observe in our logit models.

The final simulation we consider is one in which there is a systematic plan-specific error, perhaps because of errors in assigning individuals to the correct plans.

We consider a multiplicative specification to capture the fact that the impact of such errors on out-of-pocket costs would likely be proportional to the number of claims an individual possessed. In most cases, this once again leads to an upward-biased out-of-pocket cost coefficient, although if the error becomes large enough the bias becomes slightly downward (but much less than in our regressions). In this case we do estimate some sizeable coefficients on plan characteristics, with the coefficient on full donut hole coverage rising to two-thirds of what we observe in our regressions, but the plan characteristic coefficients are not consistently signed; we estimate a large positive coefficient on the deductible, for example, and a large negative coefficient on generic donut hole coverage.

Thus, our simulations do not provide any evidence to suggest that our consistent pattern of a small out-of-pocket cost coefficient and large plan characteristics coefficients are due to measurement error. Rather, they appear to correspond to true choice inconsistencies.

VII. Heterogeneity

The independence of irrelevant alternatives assumption that underlies the conditional logit model places strong restrictions on how elasticities vary across plans and will lead to inconsistent estimates if preferences are heterogeneous across the population. To address this concern, we assess the robustness of our model to heterogeneity driven by both observed and unobserved factors. We first note that our model already allows for a substantial amount of individual variation: we estimated the coefficients on individualized out-of-pocket cost parameters. Nonetheless, it may still be the case that preferences vary in ways not included in our model. In terms

of observed heterogeneity, we have reestimated our model for a number of separate samples: by gender, by age, and by tercile of the 2005 prescription drug expenditure distribution. In every case, we find that our results are very similar across all samples. In particular, each of these samples illustrates the three choice inconsistencies documented thus far: the premium coefficient is many multiples of the out-of-pocket cost coefficient; financial plan characteristics enter significantly; and the estimated degree of risk aversion is very low.

We therefore turn to considering unobserved heterogeneity. We use the Laplace approximation developed by Harding-Hausman (2007) to estimate a model with normally distributed random coefficients on all included characteristics. Our goal here is primarily a robustness check: does accounting for heterogeneity change any of our qualitative conclusions?

Table 4 shows the results of this analysis. As before, column 1 is the original model on a small sample. Column 2 adds random coefficients on premium, perfect foresight out-of-pocket, variance, and quality, while column 3 adds random coefficients on all variables. Again, we see that the choice inconsistencies are present even after accounting for unobserved heterogeneity. Further, the magnitude of the coefficients estimated in the model without heterogeneity (which correspond to the mean of the random coefficients in this model) is not much affected. We do estimate significant heterogeneity in the coefficients on premium, quality, the deductible, and the generic donut hole term; allowing for this heterogeneity turns out not to have a significant impact on the welfare results we report below.

We can also interpret the results in Table 4 as a test of the IIA assumption. To the extent that any of the coefficients are significant, this suggests that the IIA assumption does not hold exactly. Nonetheless, the fact that the magnitude of the coefficients does not change substantially once we allow for random coefficients suggests that this assumption is not altering our conclusions.

VIII. Conclusions and Implications

The new delivery mechanism for a public insurance benefit introduced by the Medicare Part D program is a radical departure from the traditional public insurance model—and an exciting opportunity to understand the role of choice in the delivery of public insurance. Using a unique dataset, we have provided the first evidence on the efficacy of the choices made by individuals under Part D. While individual choices are consistent with maximizing behavior such as preferring plans with lower premiums, lower out-of-pocket exposure, and higher quality, they are inconsistent with the standard model in three important respects: individuals underweight out-of-pocket spending relative to premiums; they overweight plan characteristics beyond their own circumstances; and they do not fully appreciate the risk-reducing aspects of plans for themselves.

Our conclusions do imply that the distribution of health insurance plan coverage would be quite different if there were no choice inconsistencies. We estimate that the share of our sample with some coverage in the donut hole gap would fall by 40 percent if these inconsistencies were corrected. This would have led to a major shift away from Humana, the insurer that offered the most generous donut hole coverage, toward other insurers.

TABLE 4—RANDOM COEFFICIENTS RESULTS

	(1)	(2)	(3)
Premium (hundreds)	-0.7156*** (0.0094)	-0.7677*** (0.0111)	-0.7354*** (0.0841)
Standard deviation of premium	x	0.2940*** (0.0560)	0.2659*** (0.0330)
OOP costs (realized) (hundreds)	-0.1040*** (0.0039)	-0.1091*** (0.0052)	-0.1193*** (0.0171)
Standard deviation of OOP costs	x	0.0001 (0.0003)	0.0226 (0.0506)
Variance (times 10 ⁶)	-0.1103*** (0.0517)	-0.1486 (0.0926)	-0.0866 (0.1237)
Standard deviation of variance	x	0.2035*** (0.0685)	0.2603 (0.1769)
Deductible (hundreds)	-0.3079*** (0.1257)	-0.2524*** (0.0107)	-0.2354 (0.1903)
Standard deviation of deductible	x	x	0.2922*** (0.1452)
Donut hole	2.805*** (0.0490)	2.775*** (0.0630)	2.523*** (0.3538)
Std. deviation of donut hole	x	x	0.8078 (0.5997)
Generic coverage	0.4743*** (0.0341)	0.4845*** (0.0427)	-0.0037 (0.2320)
Standard deviation of generic coverage	x		1.106*** (0.5108)
Full cost sharing	3.391*** (0.1417)	2.083*** (0.1852)	1.829 (1.304)
Standard deviation of cost share	x		.1290 (.4828)
Number of top 100 on form	0.0995*** (0.0019)	0.0992*** (0.0023)	0.1405*** (0.0078)
Standard deviation of top 100	x		0.1071 (0.0790)
Average quality	0.7418*** (0.0098)	0.7729*** (0.0090)	0.7622*** (0.0358)
Standard deviation of quality	x	0.3753*** (0.0146)	0.3503 (0.2182)
Brand dummies	No	No	No
Brand-state dummies	No	No	No
Risk index	124	124	124
Number of patients	15,001	15,001	15,001
Number of plans	702	702	702
Number of states	47	47	47
Number of brands	36	36	36

Notes: Table shows results from estimating the random coefficients model discussed in the heterogeneity section, estimated using the Laplace approximation developed in Harding-Hausman (2007) with bootstrapped standard errors. This model is identical to the model in equation (6), adding the normally distributed noise term (which is a function of $C_{ij} - \mu_{ij}$) and the variance adjustment. Each column shows coefficients from a single regression. The coefficients reported are the parameters of the utility function, not marginal effects. Standard errors are in parentheses. In columns 2 and 3, each set of four rows reports the mean and standard deviation of random coefficient (and their standard errors). The first column estimates the conditional logit model on the subsample of 15,000 (and is identical to the second column of Table 2). The second column adds random coefficients on financial characteristics and quality, and the third column adds random coefficients on all variables. Variable definitions are otherwise identical to Table 1.

***Significant at the 1 percent level.

**Significant at the 5 percent level.

Yet we also note that our results are not inconsistent with those of Heiss, McFadden, and Winter (2009), who use survey data of individuals to document significant adverse selection in plan choice. The fact that we estimate a nonzero coefficient on out-of-pocket costs in our logit models is consistent with some adverse selection: controlling for premiums and plan characteristics, individuals prefer plans that offer better coverage for the drugs they plan to take. Our estimated degree of adverse selection is lower than that estimated by Heiss, McFadden, and Winter (2009), however. They estimate that among those with more than three drug claims in a year, the odds of choosing some gap coverage is 10.5 percent higher than for those with three or fewer claims; that difference is only 2.7 percent in our data. They also estimate that among those with more than \$2,250 in prescription drug spending in a year, the odds of choosing some gap coverage is 8.8 percent higher than for those with less spending; that difference is 7.4 percent in our data. The difference in results between our analyses is partially driven by the fact that they have individuals with zero claims in their data, while we do not in ours; although, as discussed in the robustness section, our conclusions do not appear sensitive to that exclusion.

One means of assessing the implications of these findings is to consider the partial equilibrium welfare gains that would occur were individuals making fully informed and rational decisions about plan choice (ignoring, for now, any supply-side considerations or computation costs; these are discussed later). If individuals were fully informed, their choices would be given by the model estimated above, but would satisfy three additional restrictions: the coefficient on premiums is equal to that on expected out-of-pocket costs; financial plan characteristics other than premiums are excluded from the utility function once we control for the individual's expected out-of-pocket costs; and individuals exhibit risk aversion in their plan choice. We assume that the coefficient on premiums represents the marginal utility of a dollar if individuals were fully informed (this, in turn, determines the dollar value of quality variables and risk characteristics).

We define a normative utility function to include premiums and out-of-pocket costs (equally weighted), variance, and quality, and value the latter characteristics in terms of dollars of premiums. We then ask: if individuals had chosen the plan that maximizes this normative utility function rather than the plan they did in fact choose, by how much would utility be improved when assessed according to the normative utility function?²¹ The answer in this model is about 27 percent of total costs—this is comparable to the 30.9 percent we found when we looked only at cost savings. The small difference is due to the fact that the lowest-cost plans also have slightly lower quality ratings on average.²² We can interpret the 27 percent number as telling us the scope of the potential partial equilibrium utility gains. If there were some intervention that would make individuals fully informed and fully rational, this is the amount by which their utility could be improved (in partial equilibrium).

²¹ Appendix D of Abaluck and Gruber (2008) contains the formula used to make this calculation, as well as a derivation.

²² The 27 percent number uses the measured coefficient on the variance of costs, which is close to zero. We can alternatively impose a coefficient that corresponds to a coefficient of absolute risk aversion of 0.0003 (roughly $CARA = 3$ with wealth of 17,000). In that case, the number rises to 27.6 percent. The difference is small because the lowest-cost plans offer comparable risk protection to the plans that are actually chosen.

This large effect suggests that policymakers consider reforms that realize some of these gains. Some possibilities include directly providing individualized information about costs (as in Kling 2008) or appointing surrogates such as doctors or pharmacists to play some role in plan choice.

A more difficult question is whether these findings justify actual restrictions in the choice set facing seniors. As discussed in Abaluck and Gruber (2008), if policymakers are able to restrict choice to the plans on the efficient frontier, there are sizeable welfare gains for seniors (in partial equilibrium). It is unclear, however, whether policymakers would be able to effectively distinguish such plans; simply restricting choice to a random subset of plans does not raise welfare.

A full modeling of policy implications must also consider the general equilibrium implications. For example, restricting the size of the choice set may lower competitive pressure on the supply side. Of course, there are possible reforms that would preserve the competitive nature of the bidding process while reducing the number of plans ultimately offered to consumers, such as first-stage bidding across plans to offer one of a limited set of plan structures. On the demand side, this analysis assumes that the estimated choice process is fixed. We assume that individuals choose according to the same positive utility function regardless of the size of the choice set—any utility increases from smaller choice sets arise because there is less scope for error. If individuals are in fact better able to evaluate alternatives in a smaller choice set, then our analysis would understate the potential gains. Moreover, surveys indicate that elders spend an average of three hours selecting their Part D plan (Kling et al. 2008), so the dollar value of the hours saved by dramatically simplifying the choice process may be nontrivial as well.

Our models do not distinguish between the case of boundedly rational consumers choosing plans they trust as a heuristic, given the time-costs of fully evaluating choices, and the case where consumers simply err in underweighting out-of-pocket costs due to a lack of cognitive ability. While this distinction is important for evaluating the potential efficacy of providing consumers with additional information, it is less relevant to considering the welfare impact of altering the choice set: in either case, our estimates imply that consumers would be better off if there were less scope for choosing the wrong plan.

REFERENCES

- Agarwal, Sumit, John C. Driscoll, Xavier Gabaix, and David Laibson. 2007. *The Age of Reason: Financial Decisions over the Lifecycle*: National Bureau of Economic Research Working Paper 13191.
- Abaluck, Jason, and Jon Gruber. 2009. "Choice Inconsistencies Among the Elderly: Evidence from Plan Choice in the Medicare Part D Program." National Bureau of Economic Research Working Paper 14759.
- Abaluck, Jason and Jonathan Gruber. 2011. Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program: Dataset. *American Economic Review*. <http://www.aeaweb.org/articles.php?doi=10.1257/aer.101.4.1180>.
- Bertrand, Marianne, Dean Karlan, Sendhil Mullainathan, Eldar Shafir, and Jonathan Zinman. 2005. "What's Psychology Worth? A Field Experiment in the Consumer Credit Market." Economic Growth Center Working Paper 918.
- Boatwright, Peter, and Joseph C. Nunes. 2001. "Reducing Assortment: An Attribute-Based Approach." *Journal of Marketing*, 65(3): 50–63.
- Centers for Medicare and Medicaid Services. 2007. "Cover Memo for Medicare Part D Benefit Parameters: Annual Adjustments for Standard Benefit in 2007." https://www.cms.gov/MedicareAdvtg-SpecRateStats/downloads/2007_Part_D_Parameter_Update.pdf (accessed April 28, 2011).

- Centers for Medicare and Medicaid Services.** 2007. "Prescription Drug Plan Formulary and Network Files." Available for purchase from CMS. More details are available at: http://www.resdac.org/medicare/data_file_descriptions.asp (accessed April 28, 2011).
- Centers for Medicare and Medicaid Services.** 2007 and 2008. "2007-2010 PDP, MA, and SNP Landscape Files." <http://www.cms.gov/PrescriptionDrugCovGenIn/> (accessed April 28th, 2011).
- Centers for Medicare and Medicaid Services.** 2009. "National Health Expenditure Accounts" https://www.cms.gov/NationalHealthExpendData/02_NationalHealthAccountsHistorical.asp (accessed April 28, 2011).
- Chetty, Raj, Adam Looney, and Kory Kroft.** 2007. *Saliency and Taxation: Theory and Evidence*. National Bureau of Economic Research Working Paper 13330.
- Congressional Budget Office.** 2002. "Projections of Medicare and Prescription Drug Spending." <http://www.cbo.gov/ftpdocs/33xx/doc3304/03-07-MedicareSpending.pdf> (accessed April 28, 2011).
- Fratiglioni, Laura, Diane De Ronchi, and Hedda A. Torres.** 1999. "Worldwide Prevalence and Incidence of Dementia." *Drugs and Aging*, 15(5): 365.
- Harding, Matthew C. and Jerry A. Hausman.** 2007. "Using a Laplace Approximation to Estimate the Random Coefficients Logit Model by Nonlinear Least Squares." *International Economic Review*, 48(4): 1311–28.
- Hastings, Justine S., and Lydia Tejada-Ashton.** 2008. "Financial Literacy, Information, and Demand Elasticity: Survey and Experimental Evidence from Mexico." National Bureau of Economic Research Working Paper 14538.
- Heiss, Florian, Daniel McFadden, and Joachim Winter.** 2007. "Mind the Gap! Consumer Perceptions and Choices of Medicare Part D Prescription Drug Plans." National Bureau of Economic Research Working Paper 13627.
- Heiss, Florian, Daniel McFadden, and Joachim Winter.** 2009. "Regulation of Private Health Insurance Markets: Lessons from Enrollment, Plan Type Choice, and Adverse Selection in Medicare Part D." National Bureau of Economic Research Working Paper 15392.
- Irons, Ben, and Cameron Hepburn.** 2007. "Regret Theory and the Tyranny of Choice." *Economic Record*, 83(261): 191–203.
- Iyengar, Sheena S., Gur Huberman, and Wei Jiang.** 2004. "How Much Choice Is Too Much? Contributions to 401(K) Retirement Plans." In *Pension Design and Structure: New Lessons from Behavioral Finance*, ed. Olivia S. Mitchell and Stephen P. Utkus, 83–95. New York: Oxford University Press.
- Iyengar, Sheena S. and Emir Kamenica.** 2006. "Choice Overload and Simplicity Seeking." Unpublished.
- Iyengar, Sheena S. and Mark R. Lepper.** 2000. "When Choice is Demotivating: Can One Desire Too Much of a Good Thing?" *Journal of Personality and Social Psychology*, 79(6): 995–1006.
- Kaiser Family Foundation.** 2006. "Prescription Drug Coverage Among Medicare Beneficiaries." <http://www.kff.org/medicare/upload/7453.pdf> (accessed April 28, 2011).
- Kaiser Family Foundation and the Harvard School of Public Health.** 2006. "Survey Finds Few Seniors Expect To Switch Medicare Drug Plans For 2007." <http://benefitslink.com/pr/detail.php?id=40246> (accessed April 28, 2011).
- Kamenica, Emir.** 2008. "Contextual Inference in Markets: On the Informational Content of Product Lines." *American Economic Review*, 98(5): 2127–49.
- Kling, Jeffrey R., Sendhil Mullainathan, Eldar Shafir, Lee Vermeulen, and Marian V. Wrobel.** 2008. "Misperception in Choosing Medicare Drug Plans" Unpublished.
- Kuksov, Dmitri, and J. Miguel Villas-Boas.** 2006. "When More Alternatives Lead to Less Choice." Unpublished.
- Lucarelli, Claudio, Jeffrey Prince, and Kosali Simon.** 2008. "Measuring Welfare and the Effects of Regulation in a Government-Created Market: The Case of Medicare Part D Plans." National Bureau of Economic Research Working Paper 14296.
- Sarver, Todd.** 2008. "Anticipating Regret: Why Fewer Options May Be Better." *Econometrica*, 76(2): 263–305.
- Salthouse, T. A.** 1996. "The Processing-Speed Theory of Adult Age Differences in Cognition." *Psychological Review*, 103(3): 403–28.
- Wolters Kluwer.** 2006. "Medicare Part D Claims Data from Wolters Kluwer." Data are proprietary.