Evaluating Consumers’ Choices of Medicare Part D Plans

by

Jon Ketcham and Nick Kuminoff – ASU
Mike Keane –UNSW and Cepar
Tim Neal – UNSW and Cepar

September 2019

Background – Medicare Part D

- The Basic Medicare program provides health insurance for the 65+ population in the US
- But Medicare only covers 50-60% of costs:
  - For example, there are substantial co-pays
  - and Basic Medicare does not cover prescription drugs
- Medicare Part D was implemented in 2006 to provide prescription drug coverage
- The federal subsidy is now about $74 billion or 16% of Medicare total spending

Background – Medicare Part D

- The new Federal subsidies created a new private insurance market
- Several private insurers offer an array of Part D prescription drug plans (PDPs)
- The have different premiums and cost-sharing requirements
- In 2009 there were an average of 50 drug plans to choose from per CMS region
- **Medicare Part D created a complex new choice environment for consumers**

Quality of Decision Making

- A number of studies have analyzed the quality of PDP choices:
- Most conclude that consumers act “**confused**”:
- McFadden (2006) and Winter et al. (2006) look at 2006 data and find that most consumers could have saved money by choosing a different plan
- Abaluck and Gruber (2011, 2016) find evidence that choice behavior violates theory restrictions

Theory Restrictions on Choice Behavior

- Consider a logit model where utility of plan *j* to person *i* is given by:
- $U_{ij} = P_j\alpha + E(OOP)_{ij}\beta_1 + \sigma_{ij}^2\beta_2 + c_j\beta_3 + Q_j\beta_4 + \varepsilon_{ij}$
- P_j= premium of plan j
- $E(OOP)_{ij}$ = expected out-of-pocket cost to person *i* under plan *j*
- σ_{ij}^2 = *variance of OOP*
- c_j = *"irrelevant financial characteristics" of plan j*
- Q_j = *quality measures for plan j*

Theory Restrictions on Choice Behavior

- $U_{ij} = \boldsymbol{P_j}\alpha + \boldsymbol{E(OOP)_{ij}}\beta_1 + \boldsymbol{\sigma_{ij}^2}\beta_2 + \boldsymbol{c_j}\boldsymbol{\beta_3} + Q_j\beta_4 + \varepsilon_{ij}$
- Define “Total Cost” = $P_j + E(OOP)_{ij}$
- Abaluck and Gruber (2011) argue that once one conditions on Total Cost and **variance** ($\boldsymbol{\sigma_{ij}^2}$) then consumers should not care about the particular combination of co-pays, deductibles, etc. through which this is achieved
- “Irrelevant Financial Characteristics” should be irrelevant
- Theory restrictions: $\boldsymbol{\alpha = \beta_1}$ and $\boldsymbol{\beta_3}$ **= 0**
- We also expect $\beta_2 < 0$ if people are risk averse

Empirical Results

- A-G find that theory restrictions are violated:
- $\alpha \gg \beta_1$ implying that people care way too much about premiums
- $\beta_3 \neq 0$ implying people care about "irrelevant" attributes
- This is true in our data as well:
- We use admin data that contains a random 20% sample of all non-poor Medicare beneficiaries who enrolled in a PDP from 2006-2010
- We use 30% of that, which gives 525k people for 3.5 years each. (About 1.8M observations).
- If we run a logit model similar to A-G on these data we get the following result:

Table 1—Conditional Logit Results for Plan Choice

Conditional Logit	Coefficient	Std. Err.
Premium	-0.450	(0.001)
E(OOP)	-0.042	(0.000)
90th pct. OOP	-0.038	(0.000)
Quality	4.059	(0.012)
Top 100 Count	0.221	(0.001)
Cost Share	1.847	(0.011)
Deductible	-0.356	(0.001)
Gap Coverage	-0.100	(0.005)
Dummy for Last Choice	3.947	(0.008)
Dummy for Last Brand	1.861	(0.005)
Missed Savings in t-1 (%)	1.028	(0.018)
Pseudo R^2	0.611	
LL	-2,051,288	
AIC	4,102,598	
BIC	4,102,721	

Conditional Logit Results

- The coefficient on premiums, -.450, is much larger than that on E(OOP), which is only -.042
- The irrelevant attributes of cost sharing, deductibles and gap coverage are all highly significant
- There is also a great deal of inertia:
  - Lagged Plan (3.947) and lagged brand (1.861)

Preference Heterogeneity

- A concern is that a model with homogenous coefficients may be mis-specified:
  - Theory restrictions may hold at the individual level for most or even all people, but fail to hold in the aggregate
  - We may have spurious state dependence
- But if we estimate a heterogeneous logit model we obtain similar results:

Table 2 - Mixed Logit Model Results

Mixed Logit	Mean		Std. Dev.	
Premium	-0.832	(0.002)	0.531	(0.002)
E(OOP)	-0.228	(0.001)	0.144	(0.002)
90th pct. OOP	-0.037	(0.001)	0.080	(0.003)
Quality	8.646	(0.030)	0.038	(0.082)
Top 100 Count	0.179	(0.001)	0.101	(0.002)
Cost Share	0.471	(0.019)	0.580	(0.174)
Deductible	-0.371	(0.002)	0.007	(0.005)
Gap Coverage	-0.048	(0.006)	0.014	(0.012)
Dummy for Last Choice	1.877	(0.012)	0.866	(0.022)
Dummy for Last Brand	3.906	(0.014)	1.967	(0.015)
Missed Savings in t-1 (%)	1.498	(0.023)	0.124	(0.073)
LL	-1,373,296			
AIC	2,746,636			
BIC	2,746,882			

Heterogeneous Logit Results

- Estimates change, but basic story is still the same:
- The coefficient on premiums, -.832, is still much larger than that on E(OOP), which is only -.228
- The irrelevant attributes are still highly significant
- There is still a great deal of inertia:
  - Lagged Plan (1.877) and lagged brand (3.906)
- But ***heterogeneity is clearly very important***:
  - *Ln L* improves from -2.05M to -1.37M

Behavioral and Preference Heterogeneity

- Our goal is not simply to test if consumers behave rationally or not, but to better understand the decision rules that consumers actually use
- Given an understanding of how consumers actually choose, we can test whether interventions in the market improve welfare
- These interventions include limiting the size of the choice set and other measures to try and mitigate confusion

Behavioral and Preference Heterogeneity

- Our approach:
- We develop a model of behavioral heterogeneity that contains three consumer types:
  - A type that obeys theory restrictions (Type 1)
  - A type that may violate the $\alpha = \beta_1$ restriction (e.g., they may place too much weight on premiums)
  - A type that violates both the $\alpha = \beta_1$ and $\beta_3 = 0$ restrictions (e.g., they may care about irrelevant attributes)
- Within each type we allow for preference heterogeneity (e.g., some consumers may put more weight on cost, risk or quality relative to others).
- An ordered logit lets type depend on covariates
  - Geweke and Keane (JE, 2007) “Smoothly Mixing Reg” model

Table 3 – Mixed-Mixed Logit Model

	Type 1		Type 2		Type 3	
	Mean	Std. Dev.	Mean	Std. Dev.	Mean	Std. Dev.
Premium	-0.818	0.497	-1.646	0.320	-0.640	0.397
E(OOP)	-0.818	0.497	-0.213	0.114	-0.168	0.091
90th pct. OOP	-0.115	0.190	0.033	0.080	-0.052	0.129
Quality	4.409	3.797	5.566	3.451	11.549	0.042
Top 100 Count	-0.053	0.005	0.122	0.004	0.380	0.209
Cost Share					0.986	2.982
Deductible					-0.466	0.002
Gap Coverage					-0.044	0.031
Dummy for Last Choice	1.288	1.073	-0.147	0.011	2.849	0.059
Dummy for Last Brand	2.601	0.179	3.268	1.231	6.223	2.792
Missed Savings in t-1 (%)	2.289	0.663	1.023	0.047	0.236	0.101
Type Probabilities and Other Statistics						
Alzheimer's Disease	0.234					
Depression	0.190					
Age	-0.010					
cut_1	-2.911					
cut_2	-2.014					
Posterior Type Share	0.098		0.114		0.787	
	LL	AIC	BIC			
Model Selection	-1,336,413	2,672,940	2,673,577			

Description of Types

- Types 2s place much more weight on premium than OOP and they are not risk averse
- We call them “present biased” or “certainty biased,” as they are averse to known up-front premium costs, while being less sensitive to uncertain future OOP costs
- Type 3s have highly significant coefficients on plan financial characteristics that should be irrelevant once we condition on E(OOP) and risk
- We call the type 3s “Confused”

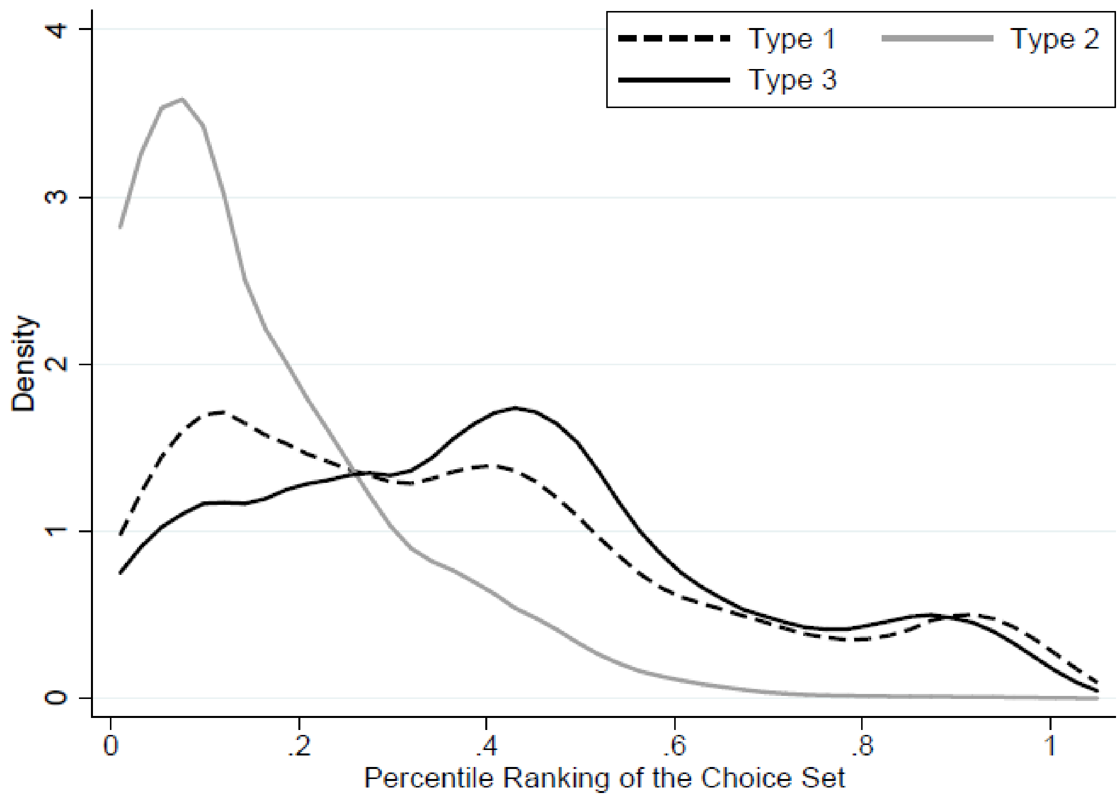
Description of Types

- According to the estimates:
- Only 9.8% of consumers are the “rational” type that obeys the theory restrictions.
- 11.4% are “Present-biased”
- 78.7% are “Confused”
- People with Alzheimer’s or Depression are more likely to be the “Confused” type

Description of Types

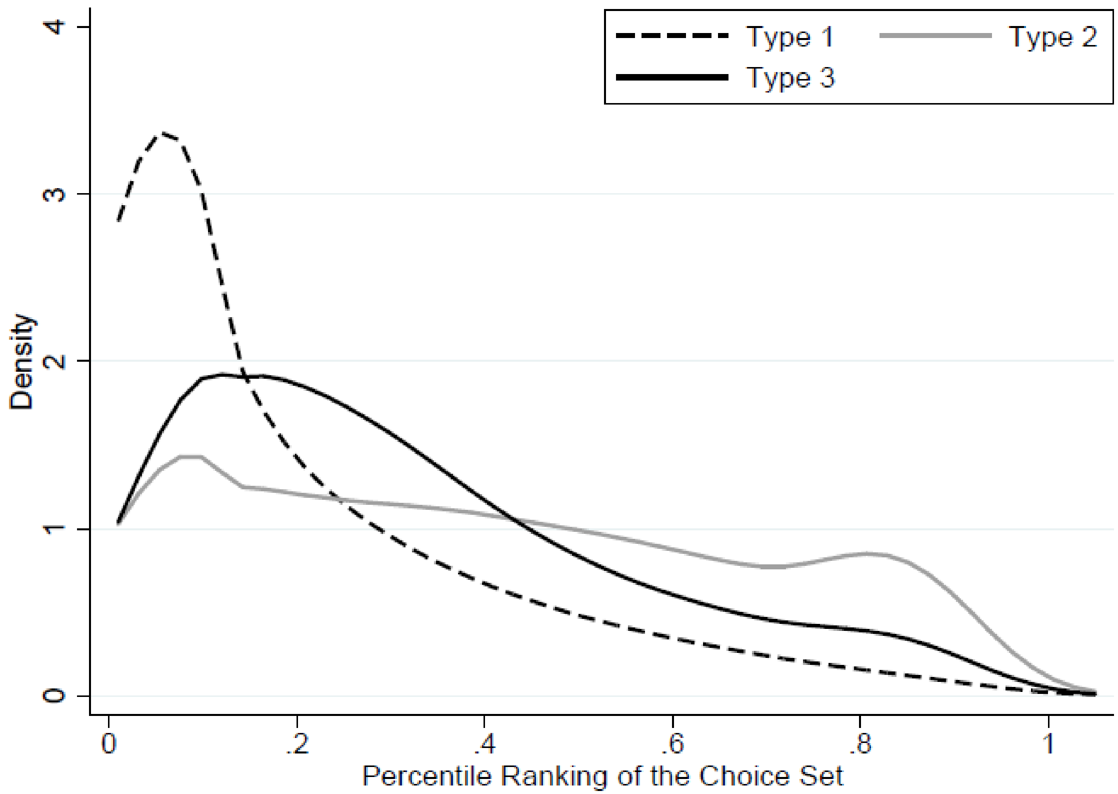
- How do the type-specific parameter differences translate into behavioral differences?
- To begin to address this:
  - We assign each person in the data to their highest posterior probability type
  - Then we compare the types in terms of the characteristics of the PDPs they chose
- Some key type differences are plotted in the next three graphs

Can Consumers Find the Lowest Premium Plans?



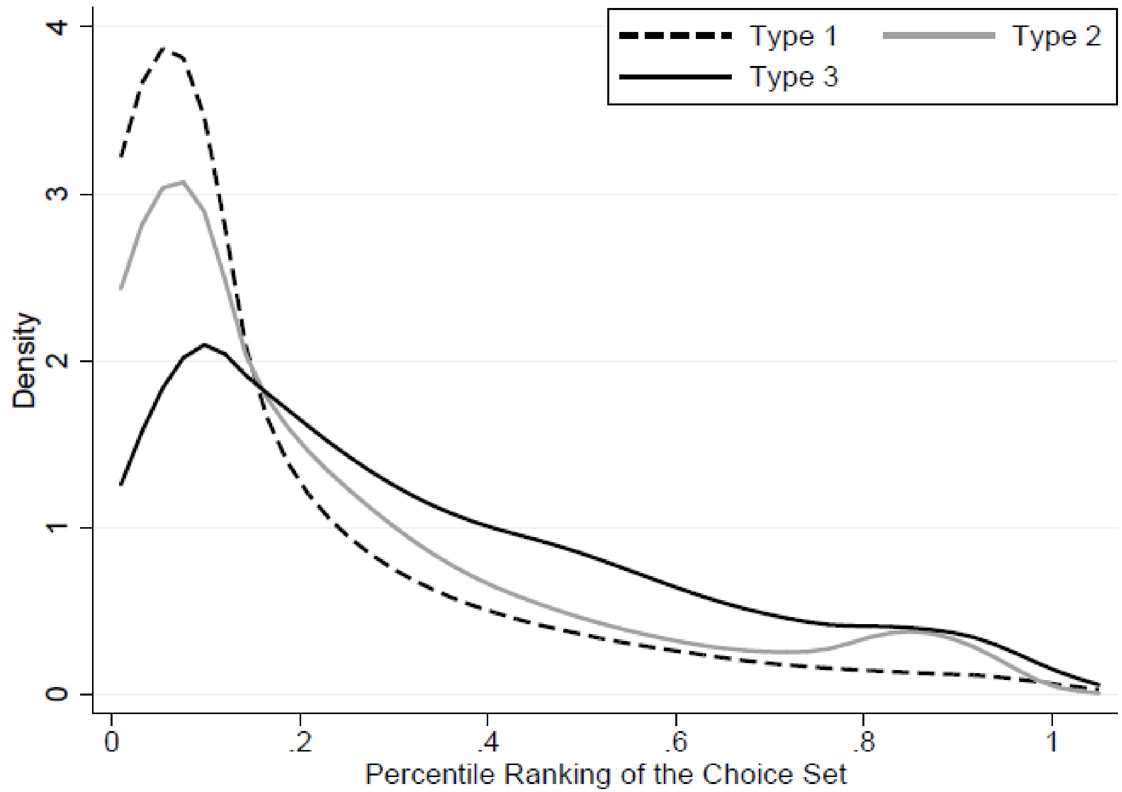
Type 2s are very good at finding one of the lowest premium plans that is available to them.
Types 1s seem to avoid the lowest premium plans (modal choice is about 7th lowest).
The modal choice of the Type 3s is about 23rd lowest (right near the middle!!).

Can Consumers Find the Lowest E(OOP) Plans?



The large majority of type 1s pick one of the lowest E(OOP) plans available to them. For type 2s the distribution is quite flat, as they don't put much weight on OOP. Type 3s are in the middle.

Can Consumers Find the Lowest Total Cost Plan? (P + OOP)



For total cost we get a clear ranking of the distributions, with Type 1 > Type 2 > Type 3.
By focussing on choosing the lowest price plans, the Type 2s seem to do almost as well as the Type 1s. The modal outcome for type 3s is to pick the 7th lowest cost plan available to them.

Additional Correlates of Type Assignments

Coefficients from an Ordered Logit Model of Types 1, 2 and 3

	(1)	(2)	(3)
understands OOP costs vary across plans	-0.34***	-0.36***	
gets help making insurance decisions	-0.44***	-0.44***	-0.37***
searched for CMS info: 1-800-Medicare	-0.18	-0.16	-0.27*
searched for CMS info: internet	-0.26**	0.19	0.19
high school graduate	0.01	-0.02	-0.02
college graduate	0.04	0.26	0.34**
college graduate * internet search for CMS info		-0.51*	-0.56**
income>$25,000	0.39***	0.50***	0.42***
income>$25,000 * internet search for CMS info		-0.48*	-0.39
married	0.16	0.16	0.17*
has living children	-0.16	-0.17	-0.26
nonwhite	0.28	0.28	0.37**
female	-0.02	-0.02	-.002
Alzheimer's disease and related dementia	0.46**	0.44**	0.49***
age	-0.02**	-0.02**	-0.02***
sample size	3,777	3,777	5,200
pseudo R^2	0.015	0.018	0.017

***,** and * denote statistical significance at the 1%, 5% and 10% levels using heteroskedasticity-robust standard errors.

Annual Overspending by Group ($)

Overspending	Mean	Std. Dev.	10th pct.	90th pct.
Whole Sample	333.79	650.67	21.66	730.75
Alzheimer's or Depression	393.62	1239.60	38.08	848.00
Age > 80	359.53	415.00	37.07	778.17
Type 1	189.10	275.72	0.00	471.96
Type 2	280.27	399.04	0.00	711.87
Type 3	346.32	682.47	41.80	740.80
Random Choice	512.19	738.20	127.20	1281.17
Random within Top 50%	293.08	513.79	70.30	547.01

Welfare Analysis

- What would happen to welfare if Type 2 and Type 3 individuals started to behave like Type 1?
- To answer this question we:
  - Simulate the choices of Type 2/3 individuals using their own type's decision utility and then a counterfactual Type 1 decision utility.
  - Then we evaluate the change in welfare using the Type 1 hedonic utility function.
- We can see the distribution of welfare benefits in the following table.

Welfare Analysis

ANNUAL WELFARE BENEFITS FOR ADOPTING TYPE 1 BEHAVIOR (IN $)

	Type 3 Individuals			Type 2 Individuals		
	Mean Benefit	90th pctile.	95th pctile.	Mean Benefit	90th pctile.	95th pctile.
New Entrants:						
Full Error	221.20	628.23	986.07	320.62	886.56	1345.73
Predicted Error	231.96	791.11	1171.62	249.29	921.05	1395.30
No Error	233.96	792.42	1168.76	249.70	917.49	1388.00
No Error (Ideal)	357.95	849.22	1212.82	376.09	982.74	1442.35
All Years:						
Full Error	162.65	473.91	720.07	225.58	610.57	937.10
Predicted Error	63.42	459.86	771.05	220.36	768.05	1069.22
No Error	68.01	468.66	778.61	220.04	763.69	1067.06
No Error (Ideal)	190.58	567.58	837.62	342.94	812.89	1115.92

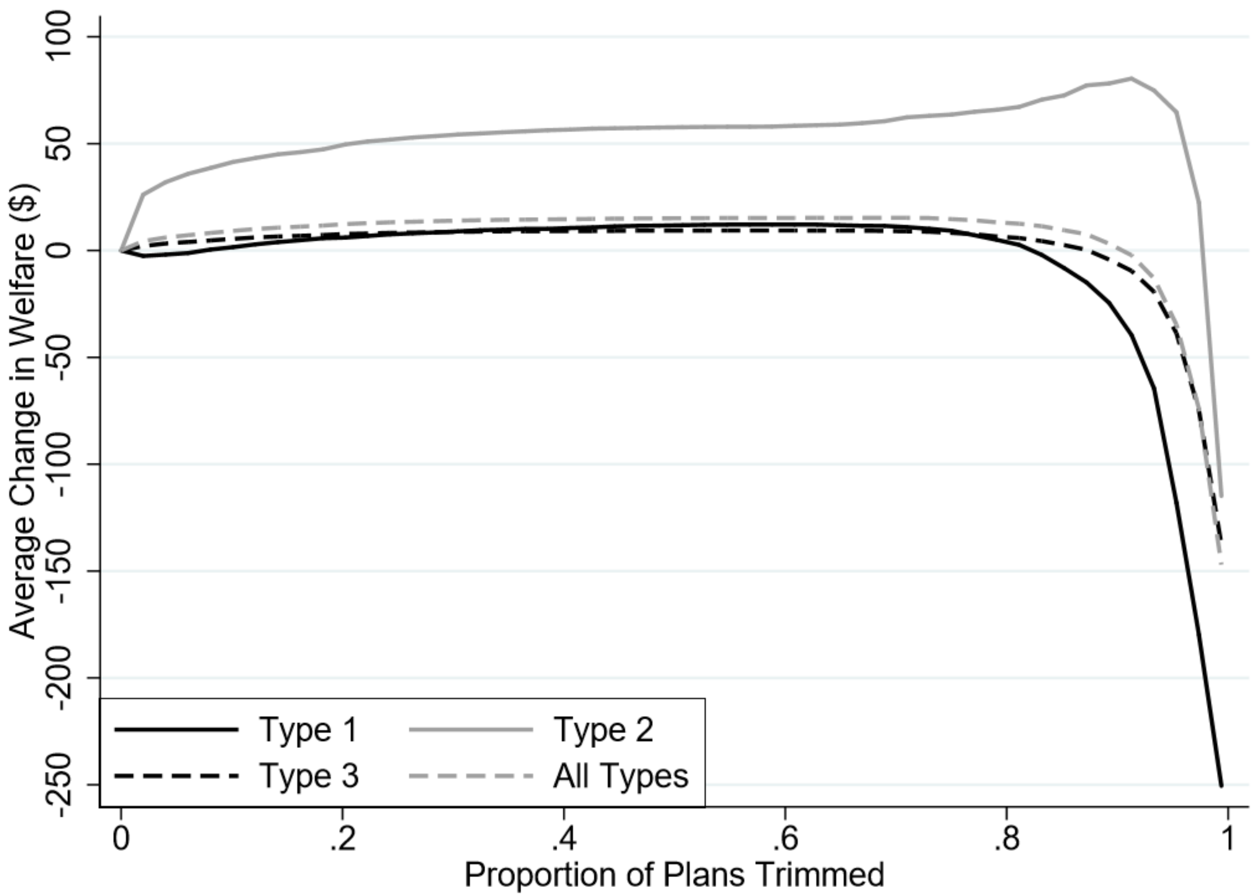
Policy Experiment: Plan Trimming

- Next, we consider whether a simplification of the choice set would improve average welfare.
- This has the potential for improving welfare as confusion/irrationality can lead to consumers picking bad plans that yield sub-optimal hedonic utility.
- We consider two types of plan ordering:
  - ‘Sharp’ ordering: Plans are ordered perfectly by their individual impact on market welfare.
  - ‘Blunt’ ordering: Plans are ordered imperfectly by the percentage of times they are dominated by another plan in the same market.
- We then trim plans from worst to best under either ordering.

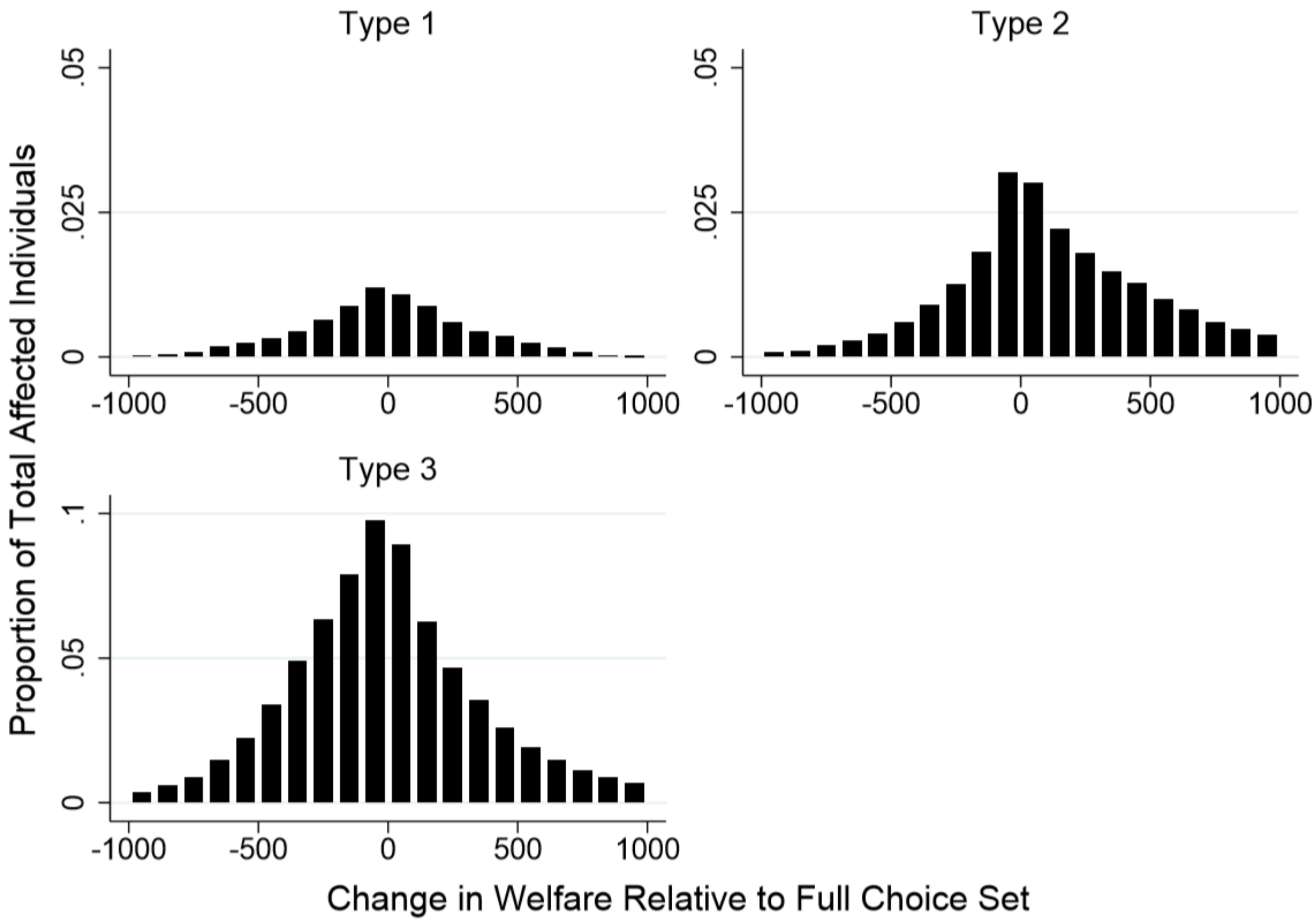
Average Annual Welfare Change for 'Sharp' Plan Trimming ($)

Plans Trimmed	1%	5%	10%	25%	50%	75%	90%
Type 1 individuals:							
Full Error	-4.22	-4.27	-4.38	-6.52	-18.89	-61.24	-153.77
Predicted Error	-2.53	-1.37	1.49	7.78	11.78	9.21	-29.24
No Error	-2.54	0.48	4.23	10.08	13.72	9.59	-29.89
Type 2 individuals:							
Full Error	26.37	26.40	26.39	25.98	20.32	1.08	-39.92
Predicted Error	22.53	33.92	41.30	52.47	57.67	63.62	80.01
No Error	24.55	34.84	41.63	50.47	56.31	64.49	77.32
Type 3 individuals:							
Full Error	-0.69	-0.67	-0.69	-1.19	-5.19	-21.78	-66.52
Predicted Error	1.09	3.69	5.44	8.28	9.32	8.27	-5.79
No Error	1.81	4.63	6.61	9.50	10.57	9.68	-4.41
All individuals:							
Full Error	2.05	2.05	2.03	1.38	-3.63	-23.05	-72.06
Predicted Error	3.21	6.67	9.17	13.30	15.10	14.70	1.70
No Error	3.97	7.66	10.36	14.22	16.09	15.91	2.39

Average Annual Welfare Change for 'Sharp' Plan Trimming ($)



Distribution of Welfare Benefits with 5% 'Sharp' Plan Trimming



Average Annual Welfare Change for 'Blunt' Plan Trimming ($)

Plans Trimmed	1%	5%	10%	25%	50%	75%	90%
Type 1 individuals:							
Full Error	-0.09	-0.62	-2.45	-13.31	-50.19	-172.28	-356.89
Predicted Error	0.08	0.41	0.97	1.45	-11.79	-94.78	-182.98
No Error	0.09	0.43	1.00	1.58	-11.46	-95.71	-183.07
Type 2 individuals:							
Full Error	-0.02	-0.28	-1.29	-9.77	-38.55	-87.06	-205.25
Predicted Error	0.01	0.09	0.76	4.20	2.15	10.20	-23.98
No Error	0.01	0.09	0.74	4.09	2.11	10.21	-15.73
Type 3 individuals:							
Full Error	-0.03	-0.25	-1.68	-10.40	-31.63	-77.28	-199.61
Predicted Error	0.03	0.18	-0.35	-4.81	-15.46	-46.29	-106.37
No Error	0.04	0.19	-0.31	-4.65	-14.88	-45.28	-104.69
All individuals:							
Full Error	-0.03	-0.29	-1.71	-10.61	-34.24	-87.69	-215.64
Predicted Error	0.04	0.19	-0.09	-3.16	-13.07	-44.57	-104.46
No Error	0.04	0.20	-0.06	-3.04	-12.60	-43.85	-102.16

Conclusion

- We have proposed new methods to model behaviour and conduct welfare analysis in complex choice environments.
- Applying these methods to the Medicare Part D market, we find:
  – Average welfare losses from sub-optimal choices are somewhat small.
  – Consumers with dementia and depression have larger losses.
  – Policies that simplify choice sets offer small average benefits, helping some people but harming others.