Online Appendix 1 Confidence Intervals

I draw 200 samples the size of the estimation sample, with replacement. I estimate the model on each sample, retaining the estimated coefficients. I report the 0.025 and the 0.075 quantiles of each estimate as the lower and upper bounds of the 95% confidence interval. Although I would prefer larger bootstrap replication samples, I am limited by computational constraints. When I compute the confidence intervals in the main specification using 1000 bootstrap reps, the results are very similar.

In practice, estimates on several of the samples do not converge because one of the quantile estimation steps that relies on the Stata quantile regression algorithm does not converge, particularly at the lowest quantiles. I report the number of bootstrap replications that converge by quantile for the main specifications in the bottom rows of Tables OA4 and OA5. I exclude the estimates from the replication samples that do not converge from the calculation of confidence intervals. Specifically, the first quantile regression step does not converge at the 0.50 or 0.90 quantiles in my application, so I cannot obtain point estimates. However, I can obtain estimates from the vast majority of bootstrap replication samples, so I report the mean of the 95% confidence intervals obtained from the replication samples as the point estimate. Tables OA4 and OA5 report the point estimates as well as the mean of the confidence intervals and demonstrate that both are very similar when they are both available. These tables also show the elasticities that are derived from the point estimates and the mean elasticities derived from the replication samples, which are also similar.

Table OA1: Nonparametric Bootstrap vs. Weighted Bootstrap

			Quantile								
2004 Sample			10	20	30	40	50	60	70	80	90	
Dependent Variable: Ln(Expenditure)											
A1. CQIV											**Tobit IV**
N= 29,161	Elasticity	-1.40	-0.76	-1.16	-1.46	-1.49	-1.41	-1.38	-1.41	-1.40	-1.42
	lower bound	-1.49	-1.49	-1.46	-1.49	-1.50	-1.49	-1.45	-1.46	-1.44	-1.49
	upper bound	-1.32	-0.02	-0.86	-1.43	-1.48	-1.33	-1.30	-1.35	-1.35	-1.36
***A2.* CQIV, weighted bootstrap**											**Tobit**
N= 29,161	Elasticity	-1.40	-0.73	-1.20	-1.47	-1.49	-1.41	-1.39	-1.41	-1.40	-1.43
	lower bound	-1.49	-1.49	-1.46	-1.49	-1.50	-1.49	-1.46	-1.45	-1.44	-1.49
	upper bound	-1.31	0.04	-0.94	-1.45	-1.48	-1.33	-1.31	-1.36	-1.35	-1.37

Chernozhukov et al. (2014) provide a proof of the consistency of a weighted bootstrap procedure, as opposed to a nonparamteric bootstrap procedure, because "it has practical advantages over nonparametric bootstrap to deal with discrete regressors with small cell sizes and the proof of its consistence is no overly complex." I estimate results using a weighted bootstrap procedure for comparison. In each bootstrap replication sample, I draw a new set of weights $(e_1, ...e_n)$ from a standard exponential random variable, and I recompute the CQIV estimator in the weighted sample. I report the results in Table OA1. As shown in specification A2, results computed using the weighted bootstrap are very similar results computed the nonparametric bootstrap. The reported point estimates differ because they reflect they reflect the means of the 95% confidence intervals, as discussed above.

Online Appendix 2 Sample Selection

Although selection into the firm that I study could affect external validity, the firm has employees in every region of the United States, and it is large enough that idiosyncratic medical usage should not be a problem. With over 800,000 people covered by the plans offered by this firm, this firm is large, even among other large firms in the Medstat data. Furthermore, all of the component Medstat databases are available for this firm for 2003 and 2004, so I can check for internal consistency by comparing results across both cross-sections. Beginning in the 2003 data, the Medstat data include fields that make the determination of marginal price and continuous enrollment very accurate.

Within the firm, the main selection criterion that I apply is a continuous enrollment restriction. In my main results, which use the 2004 and 2003 data as separate cross-sections, I require that the family is enrolled from January 1 to December 31 of the given year. Selection due to the continuous enrollment restriction eliminates over 30% of the original sample in each year. Analysis of other firms in the Medstat data suggests that the rate of turnover at this firm is comparable to the rate of turnover at other large firms. Since my outcome of interest is year-end expenditure, and family members play a role in the determination of the instrument, I only include individuals in my main sample if their entire families, with the exception of newborns, are enrolled for the entire plan year. I discuss summary statistics on the sample before and after the continuous enrollment restriction in Online Appendix 3, and I report results that relax the continuous enrollment restriction in Online Appendix 11. I retain families with newborns on the grounds that child birth is an important medical expense.

Through selection based on the detailed fields in the Medstat data, I can be confident that my selected sample consists of accurate records. Since families are important to my analysis, I perform all selection steps at the family level. To avoid measurement error, I eliminate families that switch plans, families that have changes in observable covariates over the course of the year, and families that have demographic information that is inconsistent between enrollment and claims information. I also eliminate families that have unresolved payment adjustments. Statistics on each step of the sample selection are available in a supplemental data appendix.[1] Taken together, these steps eliminate less than seven percent of individuals from the continuously enrolled sample.

In this clean sample, approximately 25% of employees with other insured family members are insured in families of four or more. The 2004 main estimation sample includes 29,161 employees insured in families of four or more. Although the stoploss induces some intra-family interactions in marginal price in families of three, I restrict the estimation sample to families of four or more so that deductible interactions are also possible.

To better control for unobservables, I limit my estimation sample to the employee in each family, and I use other family members only in the determination of the instrument. In some specifications, I test robustness to including other family members in the estimation sample. Restricting the sample to employees or employees and spouses sacrifices power because it does not take the price responsiveness of all family members into account, but, arguably, it provides the best control for unobservables on the grounds that employees at the same firm have some common characteristics that they do not necessarily share with the spouses and children of their co-workers. Moreover, restricting the sample to employees eliminates the need to address

[1] `http://www.econ.yale.edu/~ak669/Data_Appendix_07_08_13.pdf`

possible correlations in price responsiveness among family members.

Online Appendix 3 Summary Statistics

In the 2004 sample, mean year-end medical expenditure by the beneficiary and the insurer is \$1,414 in the sample of employees only. As mentioned above, in my full sample, almost 40% of people consume zero care in the entire year, and people in the top 25% of the expenditure distribution are responsible for 93% of expenditures.

The first panel of Table OA2 summarizes the expenditure distribution across bins that follow a logarithmic scale. The first column shows summary statistics for the main analysis sample of employees. As shown, excluding individuals with zero expenditure, the distribution of positive expenditure among employees is very skewed, with 31.3% of individuals in the expenditure range between \$100 and \$1,000, and smaller percentages of individuals in the bins above and below this range.[2] The next three columns show summary statistics for spouses, children and other dependents, and the full sample. The statistics show that children generally have lower expenditures than employees and their spouses. Columns 8 and 9 show characteristics of the sample before the continuous enrollment restriction. As expected, year-end expenditures are smaller in this sample because these individuals have less than the full year to incur expenditures. In columns 10 and 11, I compare my sample to the nationally-representative 2004 Medical Expenditure Panel Survey (MEPS), first to all individuals without injuries under age 65, then to a MEPS sample intended to be comparable to my main estimation sample: a single respondent without injuries in each family of four or more, aged 18–64. Even though my sample is not intended to be nationally representative, the number of people with zero expenditures in my sample is very similar to the number of people with zero expenditures in the MEPS. However, average expenditures in the MEPS are lower, perhaps because the MEPS includes people in much less generous plans than the employer-sponsored plans that I study, as well as the uninsured.

Panel B in Table OA2 depicts the distribution of the endogenous variable, the marginal price for the next dollar of care at the end of the year. I calculate the marginal price to reflect the spending of the individual and his family members. If the individual has not consumed any care and the family deductible has not been met, the marginal price takes on a value of one because the individual still needs to meet the deductible. In the employee sample, 57.2% of beneficiaries face a marginal price of one, 39.0% of employees face the coinsurance rate of 0.2, and 3.8% of employees have met the stoploss and face a marginal price of zero. This price variation should be large enough to be meaningful.

The distribution of the instrument, "family injury," shows that 10.9% of employees have at least one family member who is injured in the course of the year. Since injured employees are excluded from the sample, all of the injuries included in the determination of the instrument in the employee sample are to spouses and other dependents. In the full sample, injuries to employees are included in the determination of the instrument, and the same injury can be reflected as a "family injury" for more than one person. Overall, 10.2% of individuals in the full sample have an injury in the family. I use the same criteria to determine injuries in the MEPS samples in columns 10 and 11, and I find a similar family injury rate.

[2] The first rows of Tables OA7 and OA8 report the unconditional deciles of expenditure.

Table OA2: 2004 Summary Statistics

Variable	Families of Four or More						Families of Two	Before Continuous Enrollment Selection, Families of Four or More		MEPS 2004	
	Employees	Spouses	Child/Other	Everyone	Employees		Employees	Everyone	Employees	Everyone	Ref.
	All	All	All	All	NO Fam. Injury	Fam. Injury	All	All	All	All	All
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)
A. Year-end Expenditure ($)											
0	35.5	31.0	43.8	39.5	36.1	31.2	27.1	47.8	41.0	35.9	38.7
.01 to 100.00	11.0	10.6	13.1	12.2	10.9	11.5	8.5	11.1	10.3	14.6	12.5
100.01 to 1,000	31.3	31.9	32.3	32.0	31.3	31.7	33.8	27.4	28.6	36.5	34.8
1,000.01 to 10,000	19.1	22.0	10.0	14.4	18.8	21.6	25.5	12.0	17.3	12.2	13.2
10,000.01 to 100,000	3.0	4.5	0.8	2.0	2.9	3.9	5.0	1.7	2.8	0.8	0.7
100,000.01 and up	0.0	0.1	0.0	0.0	0.0	0.1	0.1	0.0	0.0	0.0	0.0
Mean	1,414	1,982	610	1,052	1,275	1,689	2,615	893	1,314	631	655
B. Year-end Price											
0	3.8	4.7	2.2	3.0	3.4	7.4	5.2	2.6	3.5	.	.
0.2	39.0	45.4	26.8	33.1	38.1	46.4	45.4	27.4	34.1	.	.
1	57.2	49.9	71.1	63.9	58.5	46.2	49.4	70.1	62.4	.	.
Mean	0.65	0.59	0.76	0.71	0.69	0.57	0.57	0.76	0.69	.	.
C. Family Injury											
0 (NO Family Injury)	89.1	89.6	90.1	89.8	100.0	0.0	96.0	91.3	91.0	88.7	88.2
1 (Family Injury)	10.9	10.4	9.9	10.2	0.0	100.0	4.0	8.7	9.0	11.3	11.8
D. Family Size											
2	0.0	0.0	0.0	0.0	0.0	0.0	100.0	0.0	0.0	22.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	19.3	0.0
4	66.9	64.9	56.0	60.1	67.8	59.0	0.0	57.7	64.7	24.2	58.2
5	24.4	25.7	29.3	27.5	24.0	28.3	0.0	28.0	25.2	13.8	27.6
6	6.5	7.1	10.3	8.8	6.2	9.0	0.0	9.7	7.3	5.4	9.3
7	1.6	1.7	3.2	2.5	1.5	2.9	0.0	3.1	2.0	2.0	2.9
8 to 11	0.5	0.6	1.3	1.0	0.5	0.7	0.0	1.5	0.8	1.7	2.1
E. Relation to Employee											
Employee	100.0	0.0	0.0	22.7	100.0	100.0	100.0	22.5	100.0	39.6	100.0
Spouse	0.0	100.0	0.0	18.9	0.0	0.0	0.0	18.2	0.0	18.9	0.0
Child/Other	0.0	0.0	100.0	58.4	0.0	0.0	0.0	59.3	0.0	41.5	0.0
F. Male											
0 (Female)	42.6	61.0	49.2	50.0	42.7	42.4	61.5	49.9	44.9	50.3	51.0
1 (Male)	57.4	39.0	50.8	50.0	57.3	57.6	38.5	50.1	55.1	49.7	49.0
G. Year of Birth											
1934 to 1943	0.1	0.3	0.0	0.1	0.1	0.2	7.6	0.1	0.2	4.0	1.2
1944 to 1953	3.9	4.5	0.0	1.8	4.0	3.3	33.2	2.0	4.6	13.6	10.6
1954 to 1963	30.9	30.7	0.0	12.8	30.8	31.7	28.7	13.1	32.1	17.3	36.3
1964 to 1973	51.8	47.3	0.0	20.7	51.7	53.2	17.4	18.9	48.6	15.7	35.9
1974 to 1983	13.2	17.1	1.3	7.0	13.4	11.7	12.9	8.0	14.5	15.5	15.8
1984 to 1993	0.0	0.1	47.8	27.9	0.0	0.0	0.2	29.4	0.0	16.8	0.2
1994 to 1998	0.0	0.0	27.5	16.0	0.0	0.0	0.0	15.2	0.0	8.0	0.0
1999 to 2004	0.0	0.0	23.5	13.7	0.0	0.0	0.0	13.2	0.0	9.3	0.0
H. Employee Class											
Salary Non-union	29.8	33.1	29.3	30.2	29.7	30.7	10.3	25.4	25.2	.	.
Hourly Non-union	70.2	66.9	70.7	69.8	70.3	69.3	89.7	74.6	74.8	.	.
I. US Census Region											
New England	1.4	1.5	1.4	1.4	1.5	1.2	1.6	1.5	1.5	18.48	18.9
Middle Atlantic	1.6	1.7	1.5	1.6	1.6	1.6	1.7	1.8	1.8		
East North Central	15.7	15.9	15.6	15.7	15.6	16.6	14.2	15.3	15.3	22.19	23.2
West North Central	11.8	12.1	12.0	12.0	11.9	11.5	10.7	11.5	11.4		
South Atlantic	19.0	18.3	19.0	18.9	19.2	17.1	23.7	20.1	20.2		
East South Central	11.6	11.9	11.2	11.4	11.5	12.1	13.7	11.3	11.4	35.99	34.0
West South Central	28.3	28.4	28.3	28.3	28.3	27.9	25.0	26.8	26.7		
Mountain	7.5	7.2	7.8	7.6	7.4	8.2	6.4	8.4	8.2	23.33	23.9
Pacific	3.2	3.0	3.2	3.2	3.1	3.8	2.9	3.4	3.4		
J. Plan by Individual Deductible											
350	60.0	59.2	60.3	60.0	59.4	64.5	66.2	58.9	58.8	.	.
500	17.0	17.7	16.6	16.9	17.0	16.9	15.0	16.4	16.6	.	.
750	6.3	6.3	6.2	6.2	6.5	5.1	5.4	6.4	6.4	.	.
1000	16.7	16.8	16.9	16.9	17.1	13.4	13.4	18.3	18.2	.	.
K. Outpatient Visits											
0	35.5	31.0	43.9	39.6	36.0	31.1	27.0	47.8	40.9	30.2	32.7
1	16.5	15.2	18.1	17.2	16.6	16.1	12.5	16.0	15.8	17.8	16.5
2	10.8	10.5	11.3	11.0	10.6	11.9	9.5	9.7	9.9	12.1	11.4
3 to 4	13.5	14.0	12.2	12.8	13.4	13.8	14.2	10.8	12.1	14.2	13.8
5 to 10	15.7	17.9	10.6	13.2	15.5	16.7	21.5	10.7	14.1	15.0	15.7
11+	8.0	11.2	3.8	6.2	7.8	10.4	15.3	5.0	7.2	10.7	10.1
L. Inpatient Visits											
0	95.1	91.7	96.9	95.5	95.1	94.5	95.0	96.2	95.3	94.6	92.7
1	4.5	7.3	2.8	4.1	4.5	5.1	4.3	3.4	4.3	4.5	6.4
2+	0.4	0.9	0.3	0.4	0.4	0.5	0.7	0.4	0.4	0.9	0.9
Sample Size	*29,161*	*24,261*	*74,868*	*128,290*	*25,994*	*3,167*	*54,889*	*209,555*	*47,179*	*29,972*	*3,258*

Cells report column % by variable unless otherwise noted.
All statistics from 2004 MEPS are weighted, and the "employee" refers to the reference person.
In 2004 MEPS, Everyone includes the total population below age 65, and Ref. includes the reference person aged 18-64 from families of 4+.
In 2004 MEPS, Year-end Expenditure is calculated as the sum of total expenditure on inpatient care, outpatient hospital care, and office based visits.
In 2004 MEPS, Outpatient Visits is calculated as the sum of hospital outpatient visits and office based outpatient visits.

People with the injuries included in my instrument are excluded from all estimation samples, but I report statistics on injured people in Table OA3. This table includes all categories

Table OA3: Individuals with Injuries

2004 Injured Individuals	ICD-9 (1)	Family Injury Instru-ment? (2)	Total Count of Injured (3)	Mean Year-End Expend. (4)	Mean Expend. on Injury (5)	Coeff-icient (6)	Lower bound (7)	Upper bound (8)
Fractures	800-829	1	2,601	$5,173	$1,914	28	-85	141
Dislocation	830-839	0	663	$6,890	$1,293	387	-17	790
Sprains and Strains of Joints and Adjacent Muscles	840-849	0	4,347	$3,546	$520	242	-10	493
Intracranial Injuries, Excluding Skull Fractures	850-859	0	331	$9,873	$1,934	1108	-1019	3235
Internal Injury of Thorax, Abdomen, and Pelvis	860-869	1	80	$31,353	$7,080	-383	-586	-180
Open Wounds	870-899	0	3,274	$2,696	$603	123	5	240
Injury to Blood Vessels	900-904	1	20	$5,197	$794	-41	-933	851
Late Effects of Injuries, Poisonings, Toxic Effects, and Other External	905-909	1	27	$30,403	$205	93	-999	1186
Superficial Injuries	910-919	0	1,276	$2,448	$200	225	14	436
Contusion with Intact Skin Surface	920-924	0	2,626	$3,553	$304	236	82	391
Crushing Injuries	925-929	0	59	$2,296	$555	1558	53	3063
Foreign Body Injuries	930-939	1	536	$2,591	$400	83	-106	272
Burns	940-949	1	238	$3,146	$977	-86	-366	194
Injuries to Nerves and Spinal Cord	950-957	1	65	$14,322	$601	-227	-581	128
Complications of Trauma	958-959	0	3,194	$4,526	$251	244	101	387
Poisoning by Drugs, Medicinal and Biological Substances	960-979	1	172	$8,540	$1,759	-384	-580	-189
Toxic Effects of Substances Chiefly Nonmedicinal and Other External	980-995	0	1,013	$4,466	$401	267	41	493
Complications of Surgical and Medical Care, Not Elsewhere Classified	996-999	1	531	$27,675	$3,594	94	-160	348
All Injuries			15,548	$3,759	$1,317			
No Injury			116,820	$909	$0			
Everyone			132,368	$1,243	$155			

Individiuals with injuries included in the "family injury instrument" excluded from estimation samples.
Categories of injuries shown need not be mutually exclusive.

of injuries with corresponding ICD-9 codes in the first column. The second column identifies the injuries included in the instrument through the selection process discussed in Section Online Appendix 15. If a person has any claim for an injury with an ICD-9 code in one of the listed categories, he is included in the count in the third column. Sprains and strains of joints and adjacent muscles are the most prominent. In column 4, I report the mean year-end total expenditures for the injured people to demonstrate that their spending should be large enough to have a meaningful effect on the price that their family members face. Total expenditures could capture injury-related spending that is induced by the injury but is categorized under different diagnosis codes. In column 5 I only report spending for which the primary diagnosis code is in a given injury category. Internal injuries of thorax, abdomen, and pelvis appear to be the most expensive. I exploit variation in spending across injuries in Online Appendix 12.

Panels D through J of Table OA2 summarize the distribution of covariates. Family size varies from four to eleven, with 66.9% of employees in families of four. The full sample is gender balanced, but 57.4% of employees are male. All employees are between the ages of 20 and 65 in 2004. The distribution of "year of birth" is bimodal because the sample includes parents and their children. Panel H shows that 29.8% of the employees are salaried, and the remaining employees are hourly. One of the limitations of the Medstat data is that it does not include any income measures, but the salaried versus hourly classification serves as a crude proxy. I also investigate potential income effects in other ways, discussed below. The distribution of the sample by Census region demonstrates that the firm has a very national reach. The largest concentration of employees is in the West South Central Census region, where 28.3% of the sample resides.

Panel J depicts the distribution of employees and families across the four plans. Each plan has a unique individual deductible, which I use as the plan identifier. The average deductible is \$510, which very close to the average deductible of \$473 among employer-sponsored PPO plans reported in the 2006 Kaiser Annual Survey of Employer Health Benefits. The most generous plan, which has a \$350 deductible, is the most popular, enrolling 60% of the employee sample. Since this plan is the most popular, and since the low deductible makes the people in this plan the most likely to experience a price change for a fixed amount of spending, it is likely that the behavior of the people in this plan has a substantial influence on my results.

Panels K and L give the distribution of outpatient and inpatient visits.[3] In the employee estimation sample, the average number of outpatient visits is 3.4, but 35.5% of the sample has zero visits, and 8.0% of the sample has eleven or more visits. Inpatient visits are more rare - the average number of inpatient visits is 0.05, 4.5% of the sample has a single visit, and only 0.4% of the sample has two or more visits. Across all visits in the data, the average outpatient visit costs \$244, and the average inpatient visit costs \$6,704. Patients with no inpatient visits spend \$603 on average, and patients with inpatient visits spend \$10,641 on average. Given the small number of inpatient visits, we expect that patients have more discretion over outpatient visits, but given the high amount of inpatient spending, we expect that changes in inpatient visits have a large effect on expenditure. I examine results with visits as the dependent variable in Online Appendix 16.

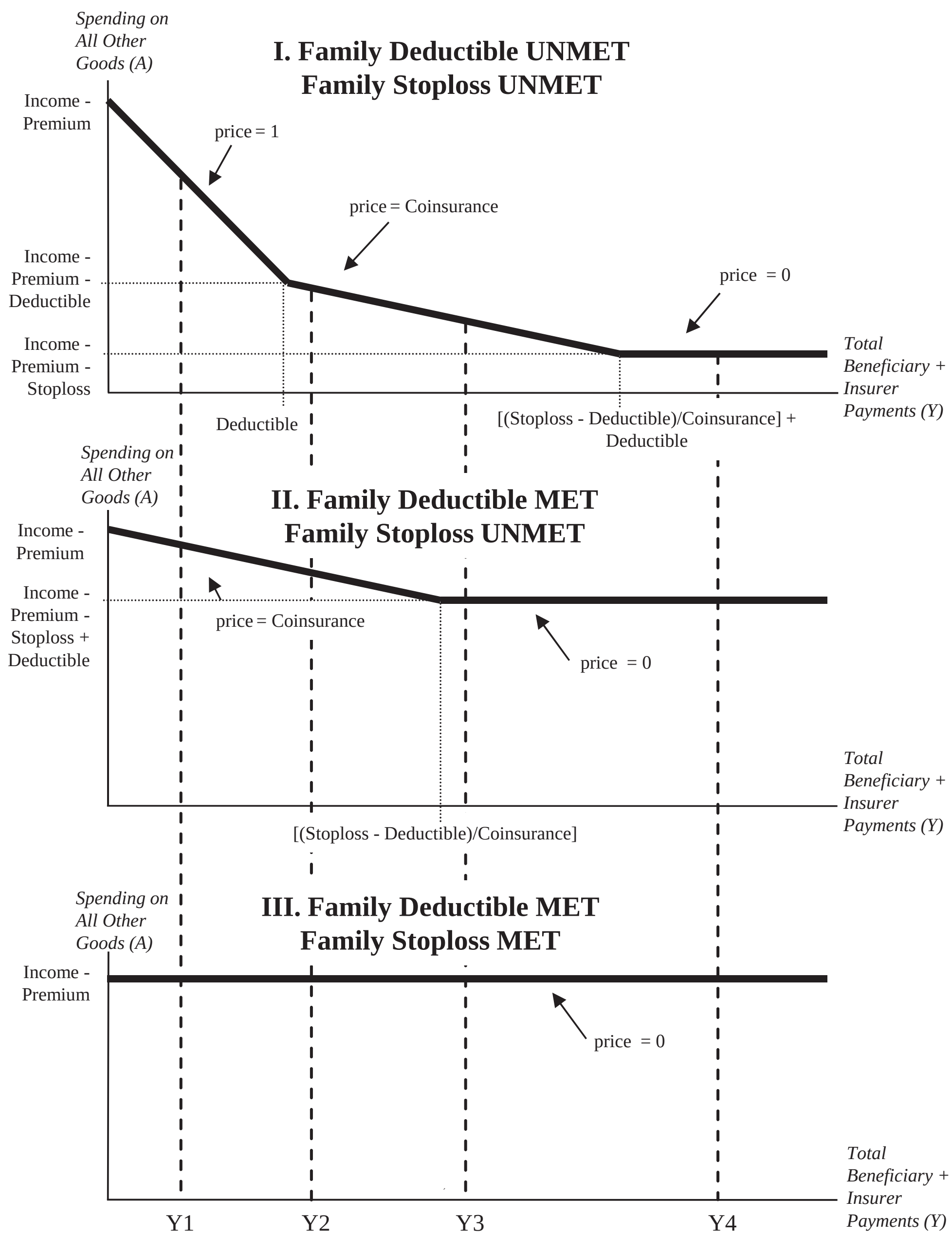
Online Appendix 4 Budget Sets Under Insurance

For an individual that is not insured as a member of a family, the top subfigure of Figure OA1 depicts how the deductible, coinsurance rate, and stoploss induces a nonlinear budget set for the individual, who faces a tradeoff between total beneficiary plus insurer spending on medical care, Y, and spending on all other goods, A. The individual faces three distinct marginal prices, the slopes of each segment.

If a consumer is insured as a member of a family, the general cost sharing structure is the same, but an additional family-level deductible and stoploss enable one family member's spending to affect another family member's marginal price. For example, in my empirical application, one plan has an individual deductible of \$500, and it also has a family deductible that is \$1,500, three times the individual deductible. Each family member must meet the individual deductible unless total family spending toward individual deductibles exceeds the family deductible. Since the family deductible is three times the individual deductible, if a family has fewer than four members, all family members must meet the individual deductible. In that case, only subfigure I of Figure OA1 applies, and it applies separately for each family member. In a family with exactly four members, when the first, second, and third family members go to the doctor, they each face the individual deductible of \$500, as if they were insured as individuals. However, when the fourth family member goes to the doctor, if the family deductible of \$1,500 has been met through the fulfillment of three individual \$500 deductibles, he makes his first payment at the coinsurance rate, as shown in subfigure II of Figure OA1. In families with more than four members, the family deductible is fixed at \$1,500,

[3]In the Medstat data, inpatient claims are organized into visits, but outpatient claims are not. To calculate outpatient visits, I calculate the number of unique days for which a patient has outpatient claims.

Figure OA1: Potential Budget Sets for Individuals



and it can be met by any combination of payments toward individual $500 deductibles.[4] A similar interaction occurs at the level of the stoploss such that an individual whose family has met the family stoploss faces the budget set shown by subfigure III of Figure OA1. Given

[4]In the case of more than four family members, intermediate cases between subfigures I and II are possible, in which the budget sets are shifted to the left because the individual has some spending at the given rate before the family deductible or stoploss is met. For simplicity of exposition, I do not show all of those cases here. However, the empirical implementation is completely general.

the family-level cost sharing parameters, some individuals will face lower marginal prices than their own medical spending would dictate.[5]

Online Appendix 5 Simple Model

In a given year, an individual in a family solves the following problem:

$$\max_{Y} \sum_{b} \pi_b U(Y, A_b|W)$$
$$s.t.\ p_b Y \leq K_b\ \forall b$$

In this simple model, agents derive utility from total agent plus insurer spending on medical care, which is given by Y (reflecting an underlying demand for health), and total spending on all other goods, A_b. W denotes a vector of individual characteristics. The individual faces three potential budget constraints, indexed by b, which include budget constraints I, II, and III from Figure OA1. The probability that the individual places on facing each budget constraint is given by π_b. Following the nonlinear budget set methodology reviewed in Hausman (1985), p_b is the price on the linear segment of budget set b that is relevant for a given Y, and K_b is the relevant "virtual income" (the y-intercept when each segment is extended to the vertical axis, representing the marginal income that the individual trades off against medical care on that segment). In this model, individuals are forward-looking maximizers of expected utility. At the start of the year, each individual forms an expectation over what his and his family members' year-end expenditures will be based on expenditures in the previous year and expected behavior and medical utilization in the coming year.

We can represent the solution to the individual's problem as a demand curve:

$$Y = \phi(p_I, p_{II}, p_{III}, K_I, K_{II}, K_{III}, \pi_I, \pi_{II}, \pi_{III}, W).$$

As shown, the demand curve will be a function, ϕ, of the marginal price and virtual income associated with the utility maximizing segment on each potential budget set, as well as the probability of each potential budget set and covariates. To make the problem more tractable, I approximate the individual's demand curve with the following demand curve:

$$Y = \psi(P|K, W),$$

which is a general function ψ of P, the realized year-end price, income, K, and covariates. This is the demand curve that forms the basis for estimation in Section 2.1. It assumes forward-looking behavior such that the realized year-end price is the utility-maximizing price. Models of forward-looking behavior have a long tradition in the literature on the demand for health care under insurance, including an influential paper by Ellis (1986), which develops a dynamic model in which agents respond to expected year-end price. Here, I am assuming that the expected year-end-price is the realized year-end price.[6]

[5]Eichner (1997, 1998) models a simpler family deductible structure in which people in family plans simply have a single family deductible, but that structure is not appropriate in the plans that I study.

[6]This assumption is true at the end of the year, after all uncertainty is resolved.

We can use the model to understand the impact of a family injury on demand, the impact of the within-year timing of the injury on demand, and the impact of a transitory vs. permanent family injury on demand. We begin with a simple example, in which there are no unexpected family injuries and the individual forms an accurate prediction of each π_b. Suppose that the individual expects to spend a small amount, given by $Y1$ in Figure OA1. Depending on whether his family members meet the family deductible or stoploss before he begins spending, he could face budget set I, II, or II, as depicted. Based on his predictions of his family members' spending, he assigns a probability to each of the three subfigures at the start of the year, and he maximizes his expected utility over all three budget sets.[7]

Consider the impact of an unexpected injury on demand. Suppose that the agent would have expected only two of his family members to meet their individual deductibles, but then, on January 1, a third family member has an injury that is large enough that it will push him over the individual deductible, thus pushing the family over the family deductible. Now, the agent adjusts his medical consumption relative to the case in which his family member was not injured such that he places a higher probability on budget sets II and III relative to I. He adjusts the number of visits that he makes to the doctor or his spending per scheduled visit accordingly (both have the same impact in terms of the model, but we test how much response comes through the visit margin in Online Appendix 16), taking into the account the expected price of the visit and the likely effect that the visit itself will have on his year-end marginal price.

The model that we have discussed so far, and the one that informs the main specifications, is a static model, but we can extend it to understand the effect of a family injury that happens in January relative to a family injury that happens in December on the agent's demand. Suppose that there are multiple periods indexed by t (days or months) within the year. As time progresses, uncertainty is resolved, and the agent changes the probabilities π_{bt} that he assigns to the budget sets in each subfigure. The model does not place any restrictions on the timing of spending - it need only occur before the plan year resets on December 31. However, we assume that all spending cannot occur exactly on December 31, for example, because appointments might not be available, because some types of spending require several appointments, or because some types of spending are incurred after a specific acute event in the middle of the year. If the injury happens on January 1, the agent, his injured family member, and his other family members have the entire year to respond to the injury. Thus, the agent's expected year-end price will be lower *and* his expected year-end spending will be higher if the injury occurs toward the beginning of the year. Suppose instead that the injury occurs on December 1. If the injury itself requires several follow-up visits, and it takes time for the agent's family members to schedule appointments to react to the injury, it could be that even though the injury would have been large enough to change the individual's year-end price if it had occurred on January 1, it will not be large enough to change the year-end price given its occurrence late in the year. Thus, the first stage effect of the injury on the marginal price will be smaller, and the reduced form effect of the injury on the agent's expenditure

[7] If a single agent maximizes utility for the entire family, the optimization process will be similar to the process for each separate individual, because there is an incentive to spread spending across several family members so that the likelihood that that family deductible is met increases. (If only one family member responds, even if he responds a great deal, the family deductible might not be met because the cost sharing rules require spending toward at least three individual deductibles before the family deductible can be met.)

will be smaller than they would have been if the injury occurred sooner. If the attenuation of the first stage effect is the same as the attenuation of the reduced form effect, then the instrumental variable estimate - the ratio of the reduced form effect to the first stage effect - will be unchanged. I present evidence consistent with the model that exploits the within-year timing of injuries in Online Appendix 15.

We can also extend the model to incorporate behavior over multiple years to understand the effect of transitory vs. permanent shocks to family members. Suppose that the arrival rate of some health events is uniform across years, but treatment can be delayed until subsequent years with minimal health cost. If an agent expects that his family members might have higher spending in the subsequent year, resulting in a lower price for his own care, he has an incentive to delay treatment.[8] Suppose a member of the agent's family has a transitory shock - he suffers a burn, but the issues related to the burn will be resolved within the calendar year, so his expected spending next year is the same as it was this year. Such a transitory shock might encourage family members to shift delayed expenditures from the previous calendar year into the current calendar year because they expect their marginal prices to be higher next year. If that is the case, my estimate of how expenditure responds to changes in year-end marginal price will be biased away from zero, relative to what the true response would be to a permanent decrease in year-end price. In contrast, suppose a member of the agent's family has a permanent shock - he develops a new condition this year that is likely to result in higher medical expenditures over the long term. The agent could also shift planned expenditures forward from the subsequent year, but he has less incentive to do so than he does if the shock is transitory because next year's budget set will likely be very similar to this year's.

Online Appendix 6 Empirical Cost Sharing

Figure OA2 shows the empirical cost sharing schedule. The sample in this figure only includes 2004 employees in families of two in the $500 deductible plan so that we can focus on variation around a simple cost sharing relationship. The empirical cost sharing relationship would be more complicated if we included families or people in multiple plans on the same graph. Figure OA2 shows that beneficiary expenses follow the in-network schedule with a high degree of accuracy, indicating that out-of-network expenses are very rare.

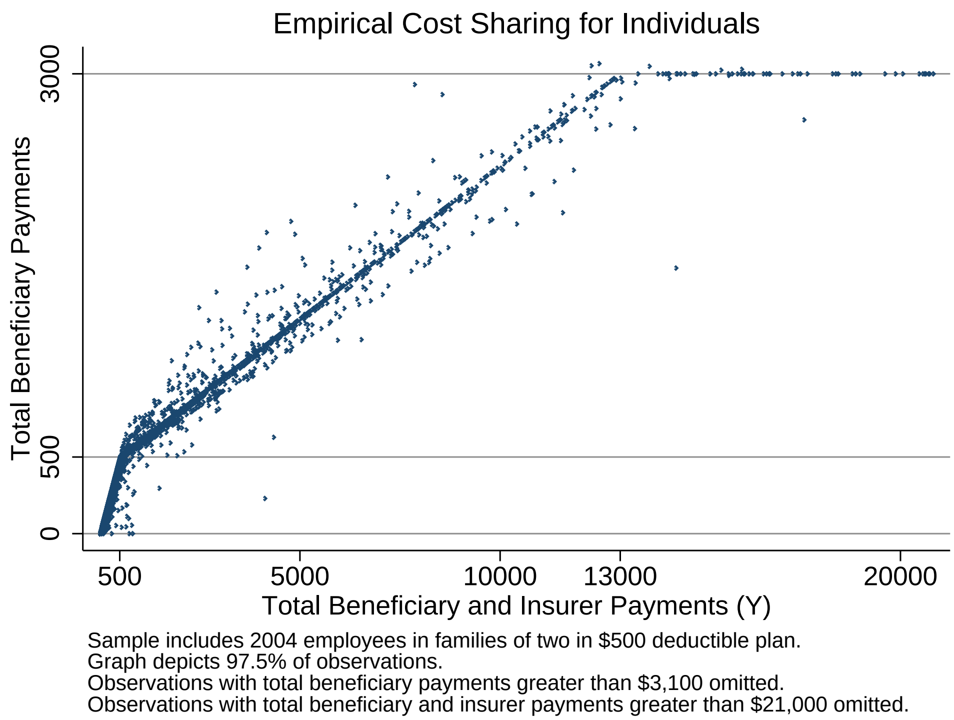
Online Appendix 7 Transforming Coefficients into Arc Elasticities

The CQIV coefficients are not elasticities because year-end price is specified in levels - price cannot be specified in logarithmic form because it can take on a value of zero. Tables OA4 and OA5 report the CQIV coefficients at the conditional deciles for the logarithmic and levels specifications of expenditure, respectively.

To interpret the coefficients, suppose that we have one group of people with a price of 1 (full insurance) and another group with a price of 0 (no insurance), holding the characteristics

[8] Cabral (2011) shows evidence of such delaying of dental care treatment. Although dental care is not included in the major medical expenditure that I study, treatment delay could be possible.

Figure OA2: Empirical Cost Sharing for Individuals



controlled for by the regressors fixed. The CQIV coefficient at the median in Table OA4 implies that median expenditure will be 522% higher for the group that faces a price of 0 than it is for the group that faces a price of 1. The coefficient at the 0.90 quantile implies that the 0.90 conditional quantile of expenditure will be 431% higher. The second set of coefficients in the table is calculated without the corner calculation. As expected, the lowest conditional quantiles of expenditure are more likely to be below zero, so the corner calculation attenuates the coefficients at the lowest conditional quantiles the most and has almost no impact on the highest conditional quantiles. If we were to rely on the coefficients without the corner calculation, we would estimate larger price responsiveness at all quantiles, but such responsiveness would not be feasible because negative expenditure is not possible.

The above claims about behavior in response to a 100 percentage point change in marginal price are misleading because such a change is not well defined at marginal prices other than 0 and 1, and this measure is not unit free. To convert this coefficient into an elasticity, I use two arc elasticity formulas. Arc elasticities are generally preferred to point elasticities for large changes. My preferred arc elasticity formula is the following midpoint formula at each conditional quantile:

$$\eta = \frac{(Y_a - Y_b)/|Y_a + Y_b|}{(a-b)/(a+b)},$$

where Y_x is the conditional quantile of expenditure at price x. Allen and Lerner (1934)

Table OA4: Coefficients and Elasticities

Dependent variable: Ln(Expenditure). Price specified as a single variable.

2004 Sample	Censored Quantile IV 10	20	30	40	50	60	70	80	90	Tobit IV
Coefficients										
Year-end Price	-4.54	-3.86	-3.05	-4.17	-5.22	-5.47	-4.25	-4.59	-4.31	-5.25
point estimate	-2.81	-0.78	-3.55	-4.25	.	-5.39	-4.28	-4.61	.	-5.50
lower bound	-6.50	-7.69	-4.60	-4.77	-7.11	-7.39	-5.22	-5.48	-4.94	-6.74
upper bound	-2.57	-0.03	-1.49	-3.56	-3.33	-3.55	-3.28	-3.70	-3.68	-3.76
Year-end Price (without corner)	-5.17	-3.94	-6.99	-9.72	-8.42	-5.47	-4.25	-4.59	-4.31	-5.26
point estimate	-6.52	-1.62	-8.30	-9.92	.	-5.39	-4.28	-4.61	.	-5.51
lower bound	-6.76	-7.82	-10.51	-11.13	-10.18	-7.39	-5.22	-5.48	-4.94	-6.76
upper bound	-3.58	-0.07	-3.48	-8.32	-6.66	-3.55	-3.28	-3.70	-3.68	-3.76
Elasticities (Midpoint)										
Elasticity	-1.40	-0.76	-1.16	-1.46	-1.49	-1.41	-1.38	-1.41	-1.40	-1.42
point estimate	-1.48	-0.81	-1.29	-1.48	.	-1.46	-1.41	-1.43	.	-1.46
lower bound	-1.49	-1.49	-1.46	-1.49	-1.50	-1.49	-1.45	-1.46	-1.44	-1.49
upper bound	-1.32	-0.02	-0.86	-1.43	-1.48	-1.33	-1.30	-1.35	-1.35	-1.36
Elasticity (without corner)	-1.41	-0.77	-1.41	-1.50	-1.49	-1.41	-1.38	-1.41	-1.40	-1.42
point estimate	-1.48	-0.86	-1.50	-1.50	.	-1.46	-1.41	-1.43	.	-1.46
lower bound	-1.49	-1.49	-1.50	-1.50	-1.50	-1.49	-1.45	-1.46	-1.44	-1.49
upper bound	-1.34	-0.04	-1.33	-1.50	-1.49	-1.33	-1.30	-1.35	-1.35	-1.36
Stoploss Elasticity	-0.47	-0.33	-0.56	-0.74	-0.68	-0.48	-0.40	-0.43	-0.41	-0.47
point estimate	-0.57	-0.16	-0.68	-0.76	.	-0.49	-0.40	-0.43	.	-0.50
lower bound	-0.59	-0.65	-0.78	-0.80	-0.77	-0.63	-0.48	-0.50	-0.46	-0.59
upper bound	-0.34	0.00	-0.33	-0.68	-0.58	-0.34	-0.32	-0.35	-0.35	-0.36
Stoploss Elast. (without corner)	0.70	0.50	0.84	1.11	1.01	0.73	0.60	0.64	0.61	0.71
point estimate	-0.57	-0.16	-0.68	-0.76	.	-0.49	-0.40	-0.43	.	-0.50
lower bound	0.52	0.01	0.50	1.02	0.87	0.51	0.48	0.53	0.53	0.54
upper bound	0.88	0.98	1.17	1.21	1.15	0.94	0.72	0.75	0.69	0.88
Rand Range Elasticity	-1.57	-0.87	-1.58	-1.71	-1.70	-1.57	-1.51	-1.56	-1.54	-1.58
point estimate	-1.68	-0.88	-1.70	-1.71	.	-1.64	-1.55	-1.58	.	-1.64
lower bound	-1.68	-1.70	-1.71	-1.71	-1.71	-1.69	-1.63	-1.64	-1.61	-1.68
upper bound	-1.46	-0.04	-1.44	-1.70	-1.68	-1.45	-1.40	-1.47	-1.47	-1.48
Rand range Elasticity (without corner)	-1.56	-0.86	-1.25	-1.65	-1.70	-1.57	-1.51	-1.56	-1.54	-1.58
point estimate	-1.68	-0.84	-1.37	-1.67	.	-1.64	-1.55	-1.58	.	-1.64
lower bound	-1.68	-1.70	-1.65	-1.69	-1.71	-1.69	-1.63	-1.64	-1.61	-1.68
upper bound	-1.44	-0.02	-0.85	-1.60	-1.68	-1.45	-1.40	-1.47	-1.47	-1.48
Logarithmic Elasticities										
Elasticity	-2.53	-1.95	-2.25	-3.35	-3.96	-2.72	-2.11	-2.28	-2.14	-2.62
point estimate	-3.22	-0.76	-2.45	-3.46	.	-2.68	-2.13	-2.29	.	-2.74
lower bound	-3.31	-3.88	-3.23	-3.81	-4.70	-3.67	-2.60	-2.72	-2.46	-3.36
upper bound	-1.76	-0.02	-1.26	-2.89	-3.22	-1.77	-1.63	-1.84	-1.83	-1.87
Elasticity (without corner)	-2.57	-1.96	-3.48	-4.83	-4.18	-2.72	-2.11	-2.28	-2.14	-2.62
point estimate	-3.24	-0.81	-4.12	-4.93	.	-2.68	-2.13	-2.29	.	-2.74
lower bound	-3.36	-3.89	-5.22	-5.53	-5.06	-3.67	-2.60	-2.72	-2.46	-3.36
upper bound	-1.78	-0.03	-1.73	-4.14	-3.31	-1.77	-1.63	-1.84	-1.83	-1.87
Rand Range Elasticity	-2.69	-2.05	-2.41	-3.61	-4.22	-2.87	-2.23	-2.40	-2.26	-2.76
point estimate	-3.41	-0.81	-2.54	-3.78	.	-2.83	-2.25	-2.42	.	-2.89
lower bound	-3.51	-4.09	-3.49	-4.18	-4.96	-3.87	-2.74	-2.87	-2.59	-3.55
upper bound	-1.86	-0.02	-1.34	-3.04	-3.47	-1.86	-1.72	-1.94	-1.93	-1.97
Rand range Elasticity (without corner)	-2.71	-2.07	-3.67	-5.10	-4.41	-2.87	-2.23	-2.40	-2.26	-2.76
point estimate	-3.42	-0.85	-4.35	-5.20	.	-2.83	-2.25	-2.42	.	-2.89
lower bound	-3.55	-4.10	-5.51	-5.83	-5.34	-3.87	-2.74	-2.87	-2.59	-3.55
upper bound	-1.88	-0.04	-1.83	-4.36	-3.49	-1.86	-1.72	-1.94	-1.93	-1.97
Boostrap reps converged	94	153	154	142	156	157	145	143	163	200

discussed a version of this formula. I have extended it to include the absolute value so that it will not yield wrong-signed elasticities for negative estimate conditional quantiles of expenditure (prices are always positive in my application). The midpoint formula is a means of calculating a percentage change that depends on both the starting and ending values instead

Table OA5: Coefficients and Elasticities

Dependent variable: Expenditure. Price specified as a single variable.

2004 Sample	Censored Quantile IV 10	20	30	40	50	60	70	80	90	Tobit IV
Coefficients										
Year-end Price	163	-2254	-13856	-18299	-10151	-1373	-3210	-5501	-9607	-3372
point estimate	206	-2471	-15593	-18815	-176	-1427	-3248	.	.	-3369
lower bound	-143	-5338	-17222	-19396	-21373	-2127	-3790	-6474	-10266	-4612
upper bound	468	829	-10489	-17202	1071	-619	-2630	-4527	-8948	-2133
Year-end Price (without corner)	187	-5829	-32647	-43278	-24144	-1664	-3409	-5848	-9715	-7873
point estimate	348	-5902	-36821	-44015	-390	-1621	-3383	.	.	-7869
lower bound	-288	-12785	-40779	-45578	-50056	-2441	-3996	-6693	-10290	-10771
upper bound	662	1127	-24516	-40978	1769	-886	-2822	-5003	-9139	-4975
Elasticities (Midpoint)										
Elasticity	-0.26	0.14	-0.64	-0.67	-0.50	-1.21	-1.35	-1.33	-1.39	-1.50
point estimate	0.57	-0.63	-0.64	-0.64	-0.21	-1.20	-1.38	.	.	-1.50
lower bound	-1.50	-0.66	-0.68	-0.73	-1.47	-1.46	-1.44	-1.42	-1.44	-1.50
upper bound	0.98	0.94	-0.60	-0.62	0.47	-0.97	-1.27	-1.25	-1.34	-1.49
Elasticity (without corner)	-2.89	0.01	-1.46	-1.47	-13.71	-2.60	-3.20	-3.90	-1.87	-226.33
point estimate	1.42	-1.34	-1.45	-1.47	-1.02	-1.42	-1.54	.	.	-59.03
lower bound	-7.38	-1.46	-1.50	-1.51	-28.97	-3.94	-5.10	-6.49	-2.39	-418.32
upper bound	1.59	1.49	-1.41	-1.43	1.56	-1.26	-1.30	-1.31	-1.35	-34.35
Stoploss Elasticity	0.17	0.03	-0.35	-0.38	-0.13	-0.10	-0.11	-0.11	-0.11	-0.23
point estimate	0.17	-0.80	-0.95	-0.94	-0.05	-0.10	-0.11	.	.	-0.22
lower bound	-0.12	-0.35	-0.38	-0.41	-0.40	-0.12	-0.12	-0.12	-0.11	-0.26
upper bound	0.47	0.41	-0.32	-0.35	0.14	-0.09	-0.10	-0.10	-0.11	-0.21
Stoploss Elast. (without corner)	-2.89	55.15	2.67	2.03	3.74	0.15	0.17	0.17	0.16	0.37
point estimate	0.40	-1.12	-1.29	-1.15	-0.16	-0.10	-0.11	.	.	-0.22
lower bound	-6.31	-1.80	1.59	1.50	-0.34	0.13	0.15	0.15	0.16	0.31
upper bound	0.53	112.09	3.75	2.57	7.82	0.17	0.18	0.19	0.17	0.44
Rand Range Elasticity	-2.89	0.01	-1.46	-1.47	-13.71	-2.60	-3.20	-3.90	-1.87	-226.33
point estimate	1.42	-1.34	-1.45	-1.47	-1.02	-1.42	-1.54	.	.	-59.03
lower bound	-7.38	-1.46	-1.50	-1.51	-28.97	-3.94	-5.10	-6.49	-2.39	-418.32
upper bound	1.59	1.49	-1.41	-1.43	1.56	-1.26	-1.30	-1.31	-1.35	-34.35
Rand range Elasticity (without corner)	-0.33	0.12	-0.06	0.00	-0.56	-1.28	-1.39	-1.37	-1.40	-1.70
point estimate	0.51	-0.36	0.00	0.00	-0.21	-1.21	-1.38	.	.	-1.71
lower bound	-1.52	-0.61	-0.12	0.00	-1.59	-1.54	-1.50	-1.48	-1.46	-1.71
upper bound	0.86	0.85	0.00	0.00	0.46	-1.02	-1.27	-1.25	-1.34	-1.69
Logarithmic Elasticities										
Elasticity	-9.26	0.19	0.00	0.00	-0.20	-2.74	-1.54	-1.42	-1.80	0.00
point estimate	0.30	0.00	0.00	0.00	-0.12	-1.40	-2.06	.	.	0.00
lower bound	-18.95	0.00	0.00	0.00	-0.65	-4.86	-2.11	-2.16	-2.23	-0.01
upper bound	0.44	0.38	0.00	0.00	0.25	-0.61	-0.97	-0.68	-1.37	0.00
Elasticity (without corner)	-9.56	0.76	2.01	2.33	0.44	-3.01	-2.04	-1.88	-2.14	0.75
point estimate	-0.28	1.10	1.96	2.15	0.04	-1.60	-2.27	.	.	1.43
lower bound	-20.91	-0.16	1.68	1.78	-1.80	-5.16	-2.41	-2.36	-2.40	-1.04
upper bound	1.79	1.68	2.34	2.87	2.69	-0.87	-1.68	-1.41	-1.89	2.54
Rand Range Elasticity	-2.81	-3.30	-3.59	-2.92	-3.93	-5.17	-5.80	-6.20	-6.61	-5.72
point estimate	.	-3.26	.	.	-4.13	-5.26	-5.81	.	.	-5.85
lower bound	-3.46	-3.76	-5.89	-4.79	-5.17	-5.55	-5.92	-6.30	-6.65	-6.12
upper bound	-2.16	-2.85	-1.30	-1.06	-2.70	-4.79	-5.68	-6.11	-6.56	-5.32
Rand range Elasticity (without corner)	-3.26	-6.33	-7.50	-7.73	-5.29	-5.20	-5.81	-6.22	-6.61	-6.40
point estimate	.	-6.24	-7.61	-7.74	-4.20	-5.27	-5.82	.	.	-6.45
lower bound	-4.34	-6.98	-7.68	-7.77	-7.84	-5.58	-5.94	-6.33	-6.65	-6.69
upper bound	-2.18	-5.68	-7.31	-7.69	-2.73	-4.82	-5.68	-6.11	-6.56	-6.11
Boostrap reps converged	132	172	180	174	128	54	89	88	95	200

of just one or the other. For example, in my application, the main price change is from before to after the deductible: 1 to 0.2. From 1 to 0.2 is an 80% change, but from 0.2 to 1 is a 400% change. The midpoint arc elasticity formula computes the change relative to the middle of the price range (0.6) and arrives at a 133% change in price. (There is a 66% change in price from 0.2 to 0.6 and another 66% change in price from 0.6 to 1, resulting in a 133% change in price over the entire range.)

I also examine arc elasticities following the alternative logarithmic elasticity formula introduced by Gallego-Diaz (1944):

$$\eta_{logarithmic} = \frac{\ln(\frac{Y_a}{Y_b})}{\ln(\frac{a}{b})}.$$

Both arc elasticity formulas are unit free, and symmetrical with respect to prices a and b, so that a price increase and a symmetric price decrease generate the same elasticity. It can be shown that logarithmic arc elasticity gives the elasticity of the isoelastic curve, $\ln Y = \eta_{logarithmic} \ln p$, that connects points (a,Y_a) and (b,Y_b). By an appeal to the mean value theorem, it can be shown that there is at least one point on the curve between these two points at which the point elasticity is exactly the log arc elasticity. See Gallego-Diaz (1944). Both arc elasticity formulas yield a value of 0 when expenditure does not change regardless of the price, and both arc elasticity formulas yield a value of 1 when outlays are equal: $aY_a = bY_b$.

Table OA6: Coefficients and Elasticities

Elasticity - Before to After Deductible													
Expenditure | P=1	1	1	1	1	1	1	1	1	1	1	1	1	0
Expenditure | P=.2	0.2	1	1.29	1.35	3.05	5	10	30	199	1000	10000	125	1
Midpoint	1.00	0.00	-0.19	-0.22	-0.76	-1.00	-1.23	-1.40	-1.49	-1.50	-1.50	-1.48	-1.50
Logarithmic	1.00	0.00	-0.16	-0.19	-0.69	-1.00	-1.43	-2.11	-3.29	-4.29	-5.72	-3.00	.
Stoploss Elasticity - Before to After Stoploss													
Expenditure | P=.2	1	1	1	1	1	1	1	1	1	1	1	.	0
Expenditure | P=0	0	1	1.29	1.35	3.05	5	10	30	199	1000	10000	.	1
Midpoint	1.00	0.00	-0.13	-0.15	-0.51	-0.67	-0.82	-0.94	-0.99	-1.00	-1.00	.	-1.00
Logarithmic	.	.	.	.		.	.	.		.	.		.
Rand Range Elasticity - From .25 to .95													
Expenditure | P=.95	1	1	1	1	1	1	1	1	1	1	1	1.1664	0
Expenditure | P=.25	25/95	1	1.29	1.35	3.05	5	10	30	199	1000	10000	64	1
Midpoint	1.00	0.00	-0.22	-0.26	-0.87	-1.14	-1.40	-1.60	-1.70	-1.71	-1.71	-1.65	-1.71
Logarithmic	1.00	0.00	-0.19	-0.22	-0.84	-1.21	-1.72	-2.55	-3.97	-5.17	-6.90	-3.00	.

However, when outlays are not equal, particularly when the elasticity is greater than one, both formulas yield very different values. Table OA6 shows how the arc elasticities vary with expenditures over three different price changes: the change from before to after the deductible in my application, the change from before to after the stoploss in my application, and the "Rand range" from 0.25 to 0.95 (the range used for the calculation of an elasticity from the Rand Health Insurance Experiment). The table shows that as the difference in expenditure at each price grows large, the midpoint arc elasticity approaches a limiting value, but the logarithmic arc elasticity continues to grow.[9] Given this property, the logarithmic elasticity is probably more informative in my application because the elasticities that I find are almost always greater than one. However, to the extent that readers would like to compare my results to the elasticity derived from the Rand Health Insurance Experiment, the midpoint elasticity formula is more appropriate.

[9] In the price range that I study from 1 to 0.2, the midpoint arc elasticity approaches -1.5. Over the Rand range, the midpoint arc elasticity approaches approximately -1.71. Given this property, the upper bound of the confidence interval is not informative if its value is -1.5.

Elasticities are a unit free measure, but if the underlying relationship is not isoelastic, the range over which they are taken matters. If the underlying relationship *is* isoelastic (it has the form $\ln Y = \eta_{logarithmic} \ln p$, where $\eta_{logarithmic}$ is a constant), then the range over which the logarithmic arc elasticity is taken will *not* matter for the estimate. However, with the midpoint formula, even if the underlying relationship is isoelastic, the range over which the midpoint arc elasticity is taken will matter. For example, in the penultimate column of Table OA6, the underlying relationship is isoelastic with $\eta_{logarithmic} = -3$, however, $\eta = -1.5$ for the range from 1 to 0.2, and $\eta = -1.65$ for the Rand price range.[10] If the underlying relationship is linear ($Y = \gamma p$), where γ is any constant, the range over which the arc elasticities is taken does not matter; as shown in the first column of Table OA6, both arc elasticities yield a value of one.

I present results from both elasticity measures with and without the corner calculation in Tables OA4 and OA5. When I specify the dependent variable as the level of expenditure, the corner calculation involves setting the conditional quantile of expenditure for some observations to zero for one or both values of the price. As shown in the last column of Table OA6, when one value of expenditure is zero and the other value is nonzero, the midpoint arc elasticity takes on its limiting value, and the logarithmic arc elasticity cannot be computed because the logarithm of zero is undefined. This is relevant in my application because the conditional quantiles of expenditure are censored at zero in the calculation of corner arc elasticities. In the corner arc elasticities, if both conditional quantiles of expenditure are censored at zero, I set the elasticity equal to zero.

The lower panels of Table OA4 show the midpoint and logarithmic elasticities, with and without the corner calculation, over the same three price changes shown in Table OA6. The elasticities vary dramatically based on the arc elasticity formula and the price change used. The midpoint corner arc elasticities, the first set of elasticities in the table, vary from -0.76 to -1.49 across the conditional quantiles of expenditure. These elasticity estimates convey that if price increases by one percent, expenditure will decrease between 0.76 percent and 1.49 percent.

Table OA7 provides more information on how each of the elasticities in Table OA4 can be interpreted in terms of conditional quantiles in my context. The first row shows the actual quantiles of expenditure in the data. The second row shows the average estimated expenditure at each quantile, conditional on all covariates, including the observed year-end price.[11] Comparing the two rows makes clear that the conditional quantiles of expenditure do not correspond to the actual quantiles of expenditure. The next rows shows the average

[10] The example cannot be extended to the stoploss price change because one of the prices is zero.

[11] To obtain the estimated conditional quantiles of expenditure for the estimates derived from the logarithmic model, I exponentiate the predicted value for each observation, using the point estimates at each quantile. Because the estimator did not converge at the 0.5 or 0.9 quantiles, there are no conditional quantile estimates at those values. After exponentiating, all expenditures are positive, so I do not need to apply the corner calculation. However, before I exponentiate, some values are below the censoring point. For the corner calculations, I first set the conditional quantile of log expenditure equal to the censoring point for people with values below the censoring point, and then I exponentiate.

In the health economics literature, it is common to apply a Duan (1983) "smearing" transformation to estimated medical expenditures derived from a logarithmic model of the conditional mean because the logarithm of the expectation is not equal to the expectation of the logarithm, especially under heteroskedasticity. However, the smearing correction is not merited in my context because the logarithm of the quantile is equal to the quantile of the logarithm.

Table OA7: Actual Quantiles of Expenditure, Conditional Quantiles of Expenditure, And Elasticities

Dependent variable: Ln(Expenditure).

			Quantile									
2004 Employee Sample	Price	Corner	10	20	30	40	50	60	70	80	90	Tobit IV
Average actual quantiles of expenditure, mean shown in last column												
			0	0	0	60	130	261	531	1215	3939	1414
Average actual quantiles of expenditure, conditional quantiles of expenditure using point estimate from model with dependent variable: Ln(Expenditure)												
	Actual	1	94.5	300.1	902.0	1,277.6	.	1,931.0	2,825.2	4,344.8	.	95
	Actual	0	95.1	300.6	902.0	1,277.6	.	1,931.0	2,825.2	4,344.8	.	95
	1	1	0.5	37.5	0.6	0.5	.	23.1	83.3	131.5	.	1
	1	0	0.5	37.5	0.3	0.1	.	23.1	83.3	131.5	.	0
	0.95	1	0.7	40.6	0.7	0.5	.	30.3	103.2	165.6	.	1
	0.95	0	0.7	40.6	0.5	0.2	.	30.3	103.2	165.6	.	1
	0.25	1	65.0	126.3	150.8	207.4	.	1,319.8	2,070.8	4,164.3	.	65
	0.25	0	65.0	126.3	150.8	207.4	.	1,319.8	2,070.8	4,164.3	.	65
	0.2	1	90.0	137.0	228.3	340.5	.	1,728.2	2,565.6	5,242.9	.	90
	0.2	0	90.0	137.0	228.3	340.5	.	1,728.2	2,565.6	5,242.9	.	90
	0	1	331.5	189.4	1,199.4	2,473.4	.	5,080.8	6,044.6	13,173.2	.	331
Elasticities	0	0	331.5	189.4	1,199.4	2,473.4	.	5,080.8	6,044.6	13,173.2	.	331
Elasticity (Midpoint)		1	-1.48	-0.86	-1.49	-1.50	.	-1.46	-1.41	-1.43	.	-1.48
Elasticity (Midpoint)		0	-1.48	-0.86	-1.50	-1.50	.	-1.46	-1.41	-1.43	.	-1.48
Stoploss Elasticity (Midpoint)		1	-0.57	-0.16	-0.68	-0.76	.	-0.49	-0.40	-0.43	.	-0.57
Stoploss Elasticity (Midpoint)		0	-0.57	-0.16	-0.68	-0.76	.	-0.49	-0.40	-0.43	.	-0.57
Rand Range Elasticity (Midpoint)		1	-1.68	-0.88	-1.70	-1.71	.	-1.64	-1.55	-1.58	.	-1.68
Rand Range Elasticity (Midpoint)		0	-1.68	-0.88	-1.70	-1.71	.	-1.64	-1.55	-1.58	.	-1.68
Elasticity (Logarithmic)		1	-3.22	-0.81	-3.69	-4.06	.	-2.68	-2.13	-2.29	.	-3.22
Elasticity (Logarithmic)		0	-3.24	-0.81	-4.12	-4.93	.	-2.68	-2.13	-2.29	.	-3.24
Rand Range Elasticity (Logarithmic)		1	-3.41	-0.85	-3.99	-4.49	.	-2.83	-2.25	-2.42	.	-3.41
Rand Range Elasticity (Logarithmic)		0	-3.42	-0.85	-4.35	-5.20	.	-2.83	-2.25	-2.42	.	-3.42

estimated conditional quantiles of expenditure at prices of 1, 0.95, 0.25, 0.20, and 1, all of the actual prices in my empirical setting as well as the actual prices used in the calculation of the Rand elasticity.

Table OA8: Actual Quantiles of Expenditure, Conditional Quantiles of Expenditure, And Elasticities

Dependent variable: Expenditure.

			Quantile									
2004 Employee Sample	Price	Corner	10	20	30	40	50	60	70	80	90	Tobit IV
Average actual quantiles of expenditure, mean shown in last column												
			0	0	0	60	130	261	531	1215	3939	1414
Average actual quantiles of expenditure, conditional quantiles of expenditure using point estimate from model with dependent variable: Ln(Expenditure)												
	Actual	1	80	173	486	644	562	986	1376	.	.	1320
	Actual	0	57	-3120	-17135	-20155	57	977	1363	.	.	-634
	1	1	202	0	0	0	425	423	204	.	.	0
	1	0	178	-5187	-30031	-35572	-80	410	178	.	.	-3390
	0.95	1	192	0	0	0	434	500	368	.	.	0
	0.95	0	161	-4892	-28190	-33371	-60	491	347	.	.	-2997
	0.25	1	80	26	0	0	565	1625	2715	.	.	2512
	0.25	0	-83	-761	-2416	-2560	212	1625	2715	.	.	2511
	0.2	1	73	128	207	310	574	1706	2884	.	.	2905
	0.2	0	-100	-466	-575	-360	232	1706	2884	.	.	2905
	0	1	44	720	6789	8443	614	2030	3560	.	.	4478
Elasticities	0	0	-170	715	6789	8443	310	2030	3560	.	.	4478
Elasticity (Midpoint)		1	0.70	-1.50	-1.50	-1.50	-0.22	-0.90	-1.30	.	.	-1.50
Elasticity (Midpoint)		0	5.33	-1.25	-1.44	-1.47	-3.08	-0.92	-1.33	.	.	-19.44
Stoploss Elasticity (Midpoint)		1	0.25	-0.70	-0.94	-0.93	-0.03	-0.09	-0.10	.	.	-0.21
Stoploss Elasticity (Midpoint)		0	0.26	-4.74	-1.19	-1.09	-0.14	-0.09	-0.10	.	.	-0.21
Rand Range Elasticity (Midpoint)		1	0.70	-1.71	.	.	-0.22	-0.91	-1.31	.	.	-1.71
Rand Range Elasticity (Midpoint)		0	5.33	-1.25	-1.44	-1.47	-3.08	-0.92	-1.33	.	.	-19.44
Elasticity (Logarithmic)		1	0.63	.	.	.	-0.19	-0.87	-1.65	.	.	.
Elasticity (Logarithmic)		0	.	1.50	2.46	2.85	.	-0.89	-1.73	.	.	.
Rand Range Elasticity (Logarithmic)		1	0.65	.	.	.	-0.20	-0.88	-1.50	.	.	-9.54
Rand Range Elasticity (Logarithmic)		0	.	1.39	1.84	1.92	.	-0.90	-1.54	.	.	.

The corner calculation has a much more pronounced impact on the elasticities from the model of the level of expenditure than it does in the model of the logarithm of expenditure. The rows with a value of 1 for "corner" censor the estimated conditional quantile of expenditure from below at zero and then average over all observations. In the logarithmic specification, the corner calculation has a negligible impact, indicating that most of the estimated conditional quantiles are positive. However, in the corresponding Table OA8, based on estimates from the level specification, the corner calculation has a much larger impact. The differences between the corner elasticities in the logarithmic and level specifications reflect differences in the models more than they reflect real phenomena. In the levels model, if the estimated conditional quantile is less than zero, the corner calculation requires that it is set to zero. In contrast, in the logarithmic model, if the estimated conditional logarithm of expenditure is less than -0.7, the corner calculation requires that it is set to zero; many negative values will not have to be censored. In practice, values less than the censoring point occur much less frequently in the logarithmic model than they do in the levels model, perhaps because the logarithmic transformation reduces the dispersion in the dependent variable. Thus, the elasticities in the levels model are smaller at smaller quantiles because the corner calculation has a larger impact at the lower conditional quantiles of expenditure, which are more likely to be zero. This pattern arises because changes in observed expenditure are less likely when observed expenditure is zero, not because people at the lowest conditional quantiles have lower latent price responsiveness.

The final rows of Table OA7 show all of the elasticities reported in Table OA4, calculated directly from the previous rows.[12] The elasticity of -1.48 at the 0.10 quantile of expenditure conveys that, holding covariates fixed, the estimated 0.10 conditional quantile of expenditure is 50 cents among a group of people with a price of 1, but it increases to $331.50 at a price of 0. The elasticity of -1.43 at the 0.80 conditional quantile conveys that, holding covariates fixed, the estimated 0.80 conditional quantile of expenditure is $131.5 for a group of people with a price of 1, but it increases to $13,173.2 at a price of 0. Even though the difference in the estimated conditional quantiles between a price of 1 and a price of 0 is vastly different between the 0.10 and 0.80 conditional quantiles, the elasticities only differ by 0.05 (-1.48 vs. -1.43). In contrast, the first logarithmic elasticity reported, which is calculated using the exact same expenditures and prices, is 0.93 larger at the 0.10 quantile than it is at the 0.80 quantile (-3.22 vs. -2.29). These calculations demonstrate that the choice of arc elasticity formula can have a large impact on the estimated elasticity. The other elasticities in the table show that the price range over which the arc elasticities are calculated can have a dramatic impact on the reported elasticity in this application. I cannot calculate the sloploss elasticities with the logarithmic formula because one of the prices is zero. The midpoint stoploss elasticities calculated over the stoploss range from 0.2 to 1 are much smaller - from -0.16 to -0.76 across the quantiles, and the elasticities calculated over the Rand range from 0.25 to 0.95 are slightly larger - from -0.88 to -1.71 across the quantiles.

The issues inherent in the arc elasticity measures complicate the comparison between arc elasticity estimates across the quantiles and the comparison between elasticity estimates in my application and the elasticity estimate obtained from the Rand Health Insurance Experiment.

[12] The elasticities in this table are obtained by applying the various arc elasticity formulas to average estimated expenditure at each conditional for ease of exposition. In all other tables, I predict expenditure at each conditional quantile and calculate the elasticity for each individual and then take the average. Comparison of the elasticities in the "point estimate" row of Table OA4 to the elasticities reported here show the average elasticity is very similar to the elasticity at the average .

However, estimates obtained from both elasticity calculations, the logarithmic and the level specifications of the dependent variable, and my price range and the Rand price range are generally an order of magnitude larger than -0.22, the Rand elasticity estimate. As Table OA6 shows, the Rand elasticity estimate of -0.22 implies that spending is 1.29 times higher if agents face a price of 0.25 instead of a price of 0.95. My CQIV estimates obtained from the logarithmic specification of -.76 to -1.49 imply that spending is 3.4 to 199 times higher at a price of 0.2 than it is at a price of 1. The first lesson learned from this comparison are that studies that attempt to compare estimates to the Rand elasticity should take the particular arc elasticity calculation used in Rand in their empirical exercise seriously. The second is that no matter the means of comparison, my elasticity estimates are much larger than the Rand elasticity estimate.

The choice of the arc elasticity calculation is not well-informed by theory. However, to the extent that my elasticity estimates will be compared to the Rand elasticity estimate, in subsequent specifications, I report elasticities obtained using the arc midpoint elasticity formula with the corner calculation. As identification in my setting comes mainly from the price change from 1 to 0.2, I focus on midpoint corner elasticities calculated over this range, and I also discuss midpoint corner elasticities calculated from the price change from 0.2 to 0. To inform comparison with Rand, it could be argued that I need the same price change as well as the same formula. However, the Rand price range would involve out-of-sample predictions in my data, so I prefer the price change from 0.2 to 0, which does occur in my data. As the above tables have shown, if I calculate elasticities in my setting over the counterfactual range of the Rand elasticity, my estimates would be even larger.

Online Appendix 8 Stoploss Elasticities

How do we expect the stoploss elasticity to compare to the elasticity from 1 to 0.2? If the true curve is isoelastic, the elasticities over both ranges should be approximately equal (as discussed in Online Appendix 7, the logarithmic elasticities would be exactly equal). However, we might not expect the true curve to be isoelastic because the people who face the two price changes are different people. The people who face the second price change are higher spenders, for example the people at $Y2$ and $Y3$ in Figure OA1 as opposed to the people at $Y1$. We might expect high spenders to be more or less price responsive relative to lower spenders. On the one hand, high spenders might have less control over their spending because care decisions for high spenders might be mostly driven by doctor decisions. On the other hand, high spenders might exercise more control over their spending because they have more money at stake. In addition to considering the amount of spending, we might also want to consider the magnitude of the price change. Chetty (2012) argues that labor supply elasticities in response to smaller wage changes are smaller because of "optimization frictions" - people change their behavior most in response to the largest incentives. In this context, it is unclear which price change should be the most meaningful. In absolute terms, the price change around the stoploss is much smaller than the price change around the deductible, but it is larger in percentage terms.

Online Appendix 9 Alternative Estimators

For comparison with previous literature, I compare my conditional quantile CQIV estimates to conditional mean estimates. However, conditional quantile estimators and conditional mean estimators are not likely to yield the same point estimates because they do not estimate the same quantities. Quantile estimates and mean estimates are only similar to the extent that the underlying treatment effect is linear and the error distribution is symmetric and homoskedastic. In this application, CQIV estimates are particularly likely to be different from estimates obtained with mean estimators because medical expenditures are skewed and censored. Compared to conditional mean estimates, CQIV estimates are less sensitive to extreme values, and they are not based on parametric assumptions.

One of the most popular censored estimators, the Tobit estimator, developed by Tobin (1958), is based on the parametric assumption that the error term is homoskedastic and normally distributed. The Tobit IV estimator, developed by Newey (1987) provides a good comparison to the CQIV estimator because it incorporates endogeneity. Eichner (1997, 1998) used a version of the Tobit IV estimator. Relative to the Tobit estimator, the Tobit IV estimator requires additional parametric assumptions: a homoskedasticity assumption on the first stage error term and a joint normality distributional assumption on the structural and first stage error terms. In this application, it is unlikely that the Tobit IV assumption of homoskedasticity in the structural equation holds given the discreteness of the endogenous variable, year-end price.

A Hausman (1978) joint test of the Tobit IV normality and homoskedasticity assumptions can be conducted through comparison of the Tobit IV and CQIV estimates at each quantile. Under the null hypothesis, Tobit IV is consistent and efficient, and CQIV is consistent. Although it would be intuitive to compare the Tobit IV estimate, which is an estimate of the mean conditional elasticity, to a CQIV conditional median elasticity estimate, a conditional median estimate is not necessary for the comparison. Since Tobit IV imposes a constant treatment effect across all quantiles, the single Tobit IV coefficient can be compared directly to the CQIV coefficients at each quantile. The first row of Table 3 in the paper shows the CQIV and Tobit IV coefficients before the elasticity transformation. The Tobit IV coefficient is within the range of the CQIV coefficients, and the 95% confidence intervals overlap, indicating that the null hypothesis that the assumptions required by Tobit IV hold is not rejected. As shown in specification A1 of Table OA9, the Tobit IV coefficient implies a larger elasticity than the CQIV coefficients at most quantiles, indicating that the use of the CQIV estimator alone does not explain the large size of my estimates relative to other estimates in the literature.

For comparative purposes, I also estimate instrumental variable adaptations of two other common censored mean estimators with endogeneity: a truncated model and a two-part model. The truncated arc elasticity estimate is -0.93 with a bootstrapped 95% confidence interval of (-1.16,-0.71).[13] The two-part model arc elasticity estimate is -0.86 with a bootstrapped 95% confidence interval of (-1.19,-0.53).[14] As with the Tobit IV estimate, these estimates are

[13]To obtain this estimate, I estimate the control function using OLS, and then I include it as an independent variable using the Stata command *truncreg*. The dependent variable is the logarithm of expenditure, truncated for the observations with zero expenditure. To predict expenditure for use in the arc elasticity calculation, I use the estimated coefficients to predict expenditure.

[14]It is unclear how to extend the two part model for endogeneity in both parts, and it is also unclear how to use it to predict expenditure for use in arc elasticity calculations. To obtain a two part model estimate

Table OA9: CQIV vs. CQR, QIV, and QR

2004 Sample			Quantile 10	20	30	40	50	60	70	80	90	
Dependent Variable: Ln(Expenditure)												
A1. CQIV												**Tobit IV**
N=	29,161	Elasticity	-1.40	-0.76	-1.16	-1.46	-1.49	-1.41	-1.38	-1.41	-1.40	-1.42
		lower bound	-1.49	-1.49	-1.46	-1.49	-1.50	-1.49	-1.45	-1.46	-1.44	-1.49
		upper bound	-1.32	-0.02	-0.86	-1.43	-1.48	-1.33	-1.30	-1.35	-1.35	-1.36
A2. CQR (no endogeneity)												**Tobit**
N=	29,161	Elasticity	-1.47	-1.48	-1.50	-1.50	-1.50	-1.47	-1.42	-1.41	-1.39	-1.50
		lower bound	-1.49	-1.49	-1.50	-1.50	-1.50	-1.50	-1.42	-1.41	-1.40	-1.50
		upper bound	-1.46	-1.48	-1.49	-1.50	-1.50	-1.45	-1.41	-1.40	-1.39	-1.50
Dependent Variable: Expenditure												
B1. CQIV												**Tobit IV**
N=	29,161	Elasticity	-0.26	0.14	-0.64	-0.67	-0.50	-1.21	-1.35	-1.33	-1.39	-1.50
		lower bound	-1.50	-0.66	-0.68	-0.73	-1.47	-1.46	-1.44	-1.42	-1.44	-1.50
		upper bound	0.98	0.94	-0.60	-0.62	0.47	-0.97	-1.27	-1.25	-1.34	-1.49
B2. CQR (no endogeneity)												**Tobit**
N=	29,161	Elasticity	-1.15	-1.28	-1.43	-1.48	-1.43	-1.40	-1.36	-1.30	-1.25	-1.50
		lower bound	-1.50	-1.49	-1.50	-1.50	-1.50	-1.45	-1.40	-1.39	-1.41	-1.50
		upper bound	-0.80	-1.06	-1.37	-1.47	-1.36	-1.34	-1.31	-1.22	-1.09	-1.50
B3. QIV (no censoring)												**IV**
N=	29,161	Elasticity	-1.31	-0.75	-1.50	-1.39	-1.49	-1.34	-0.50	-1.26	-0.87	-1.20
		lower bound	-1.50	-1.50	-1.50	-1.50	-1.50	-1.46	-1.40	-1.42	-1.44	-1.49
		upper bound	-1.11	0.00	-1.50	-1.28	-1.49	-1.22	0.41	-1.10	-0.29	-0.90
B4. Quantile regression (no censoring or endogeneity)												**OLS**
N=	29,161	Elasticity	-1.25	-1.18	-0.99	-0.99	-1.34	-0.42	-0.83	-0.56	-0.90	-1.42
		lower bound	-1.50	-1.50	-1.50	-1.50	-1.50	-1.44	-1.38	-1.38	-1.41	-1.43
		upper bound	-1.00	-0.85	-0.48	-0.49	-1.17	0.60	-0.29	0.27	-0.39	-1.41

generally much larger than those in the literature. However, Eichner (1998) reports a Tobit IV elasticity of -0.8 (the Eichner (1997) elasticity estimates vary from -0.22 to -0.32).[15]

Given the insurance-induced mechanical relationship between price and expenditure, we expect elasticity estimates that do not account for endogeneity to be biased upward. Cutting in the other direction, we expect elasticity estimates that do not account for censoring to be biased downward. Table OA9 reports a variety of estimators that account and do not account for endogeneity and censoring. The first panel specifies the logarithm of expenditure as the dependent variable, and the second panel specifies the level of expenditure as the dependent variable. Because of endogeneity, we expect that the CQR estimates should be larger than the CQIV estimates and that the Tobit estimates should be larger than the Tobit IV estimates. The results look slightly larger in the logarithmic specifications, but the endogeneity bias is apparent in the comparison of the CQR coefficients to the CQIV coefficients in the levels specifications. Because of censoring, we expect that the QIV estimates and the IV estimates should be biased toward zero relative to the CQIV and Tobit IV specifications. We can

with endogeneity, I estimate the control function using OLS, and then I include it as an independent variable in the first part probit regression as well as in the second part OLS log expenditure regression estimated only on the observations with positive expenditure. To predict expenditure for use in the arc elasticity calculation, I use the coefficients from the second part to predict expenditure for all observations. I then multiply this prediction by the predicted probability of being nonzero from the first part.

[15] The Eichner (1997, 1998) elasticities are point elasticities, so caution must be applied in comparing them to the arc elasticities elsewhere reported in this paper.

only estimate these specifications with the level of expenditure as the dependent variable, unless we specify a specific value for the logarithm of expenditure at zero in the logarithmic specifications. Indeed, we find that the IV elasticity is closer to zero than the Tobit IV elasticity. The bias in the QR and OLS estimates relative to the CQIV and Tobit IV estimates depends on the relative influence of endogeneity and censoring. At some quantiles, the QIV estimate is biased toward zero relative to the CQIV estimate, but it is biased away from zero at others. It appears that endogeneity induces more of a bias in the QR estimates relative to the CQIV estimates and censoring induces more of a bias in the OLS estimates relative to the Tobit IV estimates. Regardless of the estimator used, the elasticities that I find are large relative to those in the literature.

As another method of comparison between the CQIV estimates and estimates in the literature, I use a back-of-the-envelope calculation to transform the CQIV estimates into a single estimate. To do this, I take the average of the CQIV estimates from the 0.10 quantile to the 0.90 quantile, for an average elasticity estimate of -1.32. This estimate is still much larger than the Rand elasticity estimate of -0.22.

Online Appendix 10 Heterogeneity in Treatment Effects

Although one advantage of the CQIV estimator is that it allows the estimates to vary with unobserved heterogeneity, we might also be interested in how price elasticities vary along observed dimensions. In this section, I estimate models that restrict the sample on the basis of observable characteristics. I present models by gender, salaried/hourly status, and plan.

In Tables OA10 and OA11, specifications B and C, I present the main logarithmic and levels specifications, restricted to males and females, respectively. Average spending for females is \$1,867, which is much higher than average spending for males of \$1,067. Although the elasticities are generally in the same range in the logarithmic and levels models, women appear to be slightly more price responsive than men, perhaps because they spend more.

Specifications D and E report separate results for salaried and hourly employees. Although we do not observe income, salaried/hourly status could serve as a proxy for income in our main results, where we expect hourly employees to have lower incomes. Hourly employees spend more on medical care than salaried employees, with average spending of \$1,485, as opposed to \$1,245 for hourly employees. The results show that hourly employees are more price responsive than salaried employees, perhaps because of their lower income and higher spending. However, the results are much less precisely estimated for salaried employees because of the smaller sample size, which biases the reported elasticity in the positive direction because the reported elasticity is the average of the upper and lower bounds of the 95% confidence interval, and the lower bound cannot get more negative than -1.5.

Specifications F through I show separate specifications for each of the four offered plans, starting with the most generous. We might expect people in the least generous \$1,000 deductible plan to be more price responsive throughout the distribution because they always face higher prices. Since the most generous plan has the largest enrollment, the confidence intervals are the tightest. In the logarithmic specification, the estimates are very similar across plans. In the levels specification, people in the least generous (\$1,000 deductible) plan appear more price responsive at low conditional quantiles and less price responsive at high conditional quantiles.

Online Appendix 11 Robustness to Sample Selection

In this section, I examine the robustness of my results to the main sample restrictions. First I examine the impact of restricting the estimation sample to the employee in each family. Specifications J and K of Tables OA10 and OA11 present coefficients estimated on samples that include spouses and all other family members, respectively. The patterns in the estimates across the quantiles are very similar to those in the employee sample, but the estimates are slightly more precise given the larger sample sizes, even with reported bootstrapped confidence intervals that account for intra-family correlations.[16]

[16]I account for intra-family correlations using a block bootstrap by family. I draw replication samples that have the same number of families as the main sample, drawing all observations in a family as a single block.

Table OA10: Additional Specifications

Dependent Variable: Ln(Expenditure) unless noted otherwise.

			Censored Quantile IV								
			10	20	30	40	50	60	70	80	90	Tobit IV
2004 Sample											
A. Employee											
N= 29,161	Elasticity	-1.40	-0.76	-1.16	-1.46	-1.49	-1.41	-1.38	-1.41	-1.40	-1.42
	lower bound	-1.49	-1.49	-1.46	-1.49	-1.50	-1.49	-1.45	-1.46	-1.44	-1.49
	upper bound	-1.32	-0.02	-0.86	-1.43	-1.48	-1.33	-1.30	-1.35	-1.35	-1.36
B. Male Employee											
N= 16,730	Elasticity	-0.41	-0.50	-0.90	-1.32	-1.47	-1.41	-1.33	-1.33	-1.38	-1.38
	lower bound	-1.49	-1.49	-1.42	-1.47	-1.50	-1.50	-1.44	-1.44	-1.45	-1.49
	upper bound	0.67	0.48	-0.37	-1.16	-1.44	-1.33	-1.21	-1.22	-1.31	-1.27
C. Female Employee											
N= 12,431	Elasticity	-1.48	-1.49	-1.49	-1.49	-1.50	-1.27	-1.37	-1.40	-1.35	-1.45
	lower bound	-1.49	-1.49	-1.50	-1.50	-1.50	-1.50	-1.50	-1.49	-1.47	-1.50
	upper bound	-1.47	-1.49	-1.49	-1.48	-1.49	-1.05	-1.25	-1.32	-1.22	-1.39
D. Salaried Employee											
N= 8,703	Elasticity	-0.41	-0.46	-0.71	-1.01	-1.49	-0.19	-0.93	-1.17	-1.31	-0.16
	lower bound	-1.48	-1.49	-1.43	-1.48	-1.50	-1.47	-1.46	-1.47	-1.48	-1.48
	upper bound	0.65	0.58	0.02	-0.54	-1.48	1.09	-0.40	-0.87	-1.14	1.17
E. Hourly Employee											
N= 20,458	Elasticity	-1.12	-0.69	-1.29	-1.48	-1.49	-1.48	-1.42	-1.41	-1.38	-1.47
	lower bound	-1.49	-1.49	-1.50	-1.50	-1.50	-1.50	-1.47	-1.46	-1.45	-1.49
	upper bound	-0.75	0.11	-1.08	-1.46	-1.48	-1.47	-1.36	-1.37	-1.32	-1.44
F. $350 Deductible Employee											
N= 17,483	Elasticity	-0.84	-0.71	-1.24	-1.46	-1.49	-1.49	-1.34	-1.38	-1.37	-1.33
	lower bound	-1.48	-1.49	-1.48	-1.49	-1.50	-1.50	-1.45	-1.45	-1.44	-1.48
	upper bound	-0.20	0.06	-0.99	-1.42	-1.48	-1.48	-1.22	-1.31	-1.30	-1.18
G. $500 Deductible Employee											
N= 4,952	Elasticity	-0.46	-0.46	-0.56	-0.97	-1.50	-0.31	-0.69	-1.15	-0.86	-1.16
	lower bound	-1.49	-1.49	-1.50	-1.50	-1.50	-1.49	-1.48	-1.48	-1.47	-1.50
	upper bound	0.58	0.58	0.37	-0.43	-1.50	0.86	0.10	-0.81	-0.25	-0.82
H. $750 Deductible Employee											
N= 1,844	Elasticity	-0.97	-0.52	-0.69	-1.21	-1.45	-0.06	-0.02	-0.01	-0.02	-0.02
	lower bound	-1.49	-1.50	-1.50	-1.50	-1.50	-1.50	-1.50	-1.50	-1.49	-1.50
	upper bound	-0.46	0.45	0.12	-0.92	-1.40	1.37	1.46	1.47	1.45	1.46
I. $1,000 Deductible Employee											
N= 4,882	Elasticity	-1.45	-0.52	-0.60	-0.58	-1.50	-1.37	-1.29	-1.13	-1.30	-1.49
	lower bound	-1.49	-1.50	-1.50	-1.50	-1.50	-1.50	-1.50	-1.50	-1.50	-1.50
	upper bound	-1.42	0.45	0.30	0.34	-1.50	-1.24	-1.07	-0.77	-1.10	-1.49
J. Employee and Spouse											
N= 53,422	Elasticity	-1.44	-1.27	-1.44	-1.48	-1.50	-1.47	-1.42	-1.41	-1.42	-1.48
	lower bound	-1.49	-1.49	-1.50	-1.49	-1.50	-1.50	-1.46	-1.45	-1.45	-1.49
	upper bound	-1.40	-1.04	-1.39	-1.47	-1.50	-1.45	-1.38	-1.37	-1.38	-1.46
K. Everyone											
N= 128,290	Elasticity	-1.48	-1.49	-1.45	-1.48	-1.50	-1.43	-1.34	-1.35	-1.36	-1.49
	lower bound	-1.48	-1.49	-1.50	-1.49	-1.50	-1.46	-1.39	-1.39	-1.39	-1.49
	upper bound	-1.47	-1.48	-1.40	-1.48	-1.50	-1.40	-1.28	-1.31	-1.33	-1.48
L. Employee including injured											
N= 29,764	Elasticity	-1.29	-0.94	-1.30	-1.48	-1.49	-1.39	-1.40	-1.42	-1.41	-1.44
	lower bound	-1.49	-1.38	-1.47	-1.49	-1.50	-1.50	-1.46	-1.46	-1.45	-1.49
	upper bound	-1.09	-0.49	-1.12	-1.46	-1.48	-1.28	-1.33	-1.38	-1.36	-1.40

Table OA10: Additional Specifications Continued

M. Employee - Do not require employees or their family members continuously enrolled											
N= 47,179	Elasticity	-1.38	-1.01	-1.41	-1.49	-1.50	-1.50	-1.45	-1.44	-1.42	-1.49
	lower bound	-1.49	-1.38	-1.48	-1.49	-1.50	-1.50	-1.47	-1.46	-1.44	-1.50
	upper bound	-1.26	-0.65	-1.34	-1.48	-1.49	-1.50	-1.43	-1.41	-1.39	-1.49
N. Employee - Only require employees, and NOT their family members, continuously enrolled											
N= 36,754	Elasticity	-1.36	-0.74	-1.20	-1.47	-1.49	-1.40	-1.36	-1.38	-1.39	-1.44
	lower bound	-1.49	-1.49	-1.47	-1.49	-1.50	-1.49	-1.44	-1.44	-1.44	-1.49
	upper bound	-1.23	0.02	-0.93	-1.44	-1.49	-1.31	-1.29	-1.32	-1.34	-1.40
O. Employee, 2003 Sample											
N= 30,077	Elasticity	-0.56	-1.10	-1.43	-1.48	-1.49	-1.36	-1.35	-1.32	-1.35	-1.42
	lower bound	-1.48	-1.43	-1.49	-1.49	-1.50	-1.49	-1.46	-1.43	-1.43	-1.49
	upper bound	0.36	-0.77	-1.37	-1.46	-1.49	-1.23	-1.23	-1.22	-1.28	-1.36
P. Instrument: All Injury Categories											
N= 26,790	Elasticity	-0.74	-1.34	-1.43	-1.48	-1.49	-1.50	-1.39	-1.40	-1.40	-1.49
	lower bound	-1.48	-1.49	-1.50	-1.50	-1.50	-1.50	-1.45	-1.44	-1.44	-1.50
	upper bound	0.00	-1.19	-1.36	-1.46	-1.49	-1.50	-1.34	-1.37	-1.37	-1.49
Q. Instrument: Injuries to Children Only, Employees with Spouse Injuries Excluded from Sample											
N= 28,547	Elasticity	-1.39	-0.75	-1.16	-1.46	-1.49	-1.39	-1.39	-1.41	-1.40	-1.41
	lower bound	-1.48	-1.49	-1.46	-1.49	-1.50	-1.50	-1.46	-1.46	-1.45	-1.49
	upper bound	-1.30	-0.01	-0.87	-1.43	-1.49	-1.28	-1.31	-1.37	-1.35	-1.33
R. Instrument: Injuries to Spouses Only, Employees with Child Injuries Excluded from Sample											
N= 26,690	Elasticity	-1.12	-0.57	-1.01	-0.99	-1.42	-0.97	-0.97	-0.99	-1.01	-1.20
	lower bound	-1.49	-1.49	-1.50	-1.50	-1.50	-1.49	-1.42	-1.43	-1.46	-1.50
	upper bound	-0.75	0.35	-0.52	-0.49	-1.35	-0.45	-0.52	-0.56	-0.57	-0.90
S. Instrument: Simulated Spending on Injury											
N= 29,161	Elasticity	-1.03	-0.73	-1.01	-1.38	-1.49	-1.39	-1.33	-1.38	-1.39	-1.36
	lower bound	-1.48	-1.49	-1.44	-1.49	-1.50	-1.49	-1.44	-1.45	-1.44	-1.48
	upper bound	-0.58	0.03	-0.58	-1.28	-1.47	-1.30	-1.22	-1.32	-1.34	-1.23
T. Instrument: Dummies for Nine Injury Types											
N= 29,161	Elasticity	-1.47	-1.45	-1.48	-1.49	-1.50	-1.45	-1.41	-1.43	-1.43	-1.49
	lower bound	-1.49	-1.49	-1.50	-1.50	-1.50	-1.50	-1.46	-1.46	-1.45	-1.50
	upper bound	-1.46	-1.40	-1.47	-1.49	-1.50	-1.40	-1.37	-1.40	-1.40	-1.48
U. Instrument: Separate Dummies for Injuries in First and Second Half of Year											
N= 29,161	Elasticity	-1.27	-0.72	-1.13	-1.46	-1.49	-1.42	-1.38	-1.40	-1.40	-1.42
	lower bound	-1.48	-1.49	-1.45	-1.49	-1.50	-1.49	-1.46	-1.46	-1.44	-1.49
	upper bound	-1.06	0.05	-0.80	-1.44	-1.48	-1.35	-1.30	-1.35	-1.35	-1.36
V. Ln(Outpatient Expenditure)											
N= 29,161	Elasticity	-1.43	-1.48	-1.46	-1.47	-1.49	-1.37	-1.31	-1.34	-1.37	-1.42
	lower bound	-1.48	-1.49	-1.50	-1.50	-1.50	-1.48	-1.44	-1.43	-1.44	-1.48
	upper bound	-1.38	-1.47	-1.43	-1.45	-1.49	-1.26	-1.18	-1.24	-1.30	-1.35

		Censored Quantile IV								
		91	92	93	94	95	96	97	98	99
W. Baseline (Higher Estimated Quantiles)										
N= 29,161	Elasticity	-1.39	-1.39	-1.38	-1.37	-1.36	-1.37	-1.35	-1.35	-1.36
	lower bound	-1.44	-1.43	-1.43	-1.42	-1.42	-1.41	-1.42	-1.42	-1.46
	upper bound	-1.34	-1.35	-1.33	-1.32	-1.30	-1.32	-1.28	-1.28	-1.25

Lower and upper bounds for specifications J and K account for intra-family correlations.

Table OA11: Additional Specifications

Dependent Variable: Expenditure unless noted otherwise.

			Censored Quantile IV								
			10	20	30	40	50	60	70	80	90	Tobit IV
2004 Sample											
A. Employee											
N= 29,161	Elasticity		-0.26	0.14	-0.64	-0.67	-0.50	-1.21	-1.35	-1.33	-1.39	-1.50
		lower bound	-1.50	-0.66	-0.68	-0.73	-1.47	-1.46	-1.44	-1.42	-1.44	-1.50
		upper bound	0.98	0.94	-0.60	-0.62	0.47	-0.97	-1.27	-1.25	-1.34	-1.49
B. Male Employee											
N= 16,730	Elasticity		-0.42	0.04	-0.51	-0.56	-0.80	-1.00	-1.37	-1.38	-1.38	-1.45
		lower bound	-1.50	-0.43	-0.59	-0.68	-1.42	-1.43	-1.45	-1.44	-1.44	-1.50
		upper bound	0.67	0.51	-0.43	-0.44	-0.18	-0.56	-1.29	-1.33	-1.33	-1.40
C. Female Employee											
N= 12,431	Elasticity		-0.80	-1.01	-1.15	-1.16	-1.13	-1.16	-1.21	-1.30	-1.33	-1.46
		lower bound	-0.84	-1.25	-1.50	-1.50	-1.50	-1.41	-1.41	-1.42	-1.42	-1.50
		upper bound	-0.75	-0.77	-0.81	-0.82	-0.75	-0.91	-1.01	-1.17	-1.24	-1.41
D. Salaried Employee											
N= 8,703	Elasticity		-0.37	-0.06	-0.56	-0.67	-0.44	-1.17	-1.27	-1.33	-1.36	-1.07
		lower bound	-1.50	-0.57	-0.59	-0.79	-1.47	-1.45	-1.44	-1.44	-1.44	-1.50
		upper bound	0.76	0.46	-0.53	-0.55	0.58	-0.88	-1.11	-1.22	-1.28	-0.64
E. Hourly Employee											
N= 20,458	Elasticity		-0.37	-0.15	-0.66	-0.77	-0.63	-1.29	-1.32	-1.35	-1.37	-1.47
		lower bound	-1.50	-0.65	-0.69	-0.89	-1.39	-1.42	-1.41	-1.42	-1.44	-1.50
		upper bound	0.76	0.36	-0.64	-0.65	0.12	-1.17	-1.22	-1.29	-1.30	-1.43
F. $350 Deductible Employee											
N= 17,483	Elasticity		-0.19	0.09	-0.75	-0.76	-0.63	-0.44	-1.25	-1.30	-1.40	-1.46
		lower bound	-1.33	-0.76	-0.76	-0.79	-1.28	-1.46	-1.40	-1.41	-1.46	-1.50
		upper bound	0.95	0.93	-0.73	-0.74	0.03	0.57	-1.09	-1.20	-1.33	-1.41
G. $500 Deductible Employee											
N= 4,952	Elasticity		-0.21	-0.47	-0.89	-1.00	-1.11	-1.36	-1.33	-1.33	-1.35	-0.34
		lower bound	-1.50	-1.26	-1.25	-1.47	-1.48	-1.46	-1.45	-1.42	-1.45	-1.50
		upper bound	1.08	0.33	-0.52	-0.54	-0.73	-1.27	-1.21	-1.25	-1.26	0.82
H. $750 Deductible Employee											
N= 1,844	Elasticity		-0.31	-0.76	-0.50	-0.45	-0.31	-1.01	-1.35	-1.21	-0.72	-0.29
		lower bound	-1.48	-1.49	-1.48	-1.50	-1.48	-1.45	-1.44	-1.45	-1.44	-1.50
		upper bound	0.86	-0.03	0.48	0.59	0.86	-0.57	-1.26	-0.96	0.00	0.92
I. $1,000 Deductible Employee											
N= 4,882	Elasticity		-1.38	-0.50	-0.72	-0.69	-1.48	-0.86	-0.67	-0.60	-0.55	-1.50
		lower bound	-1.50	-1.50	-1.50	-1.50	-1.50	-1.37	-1.13	-1.07	-0.93	-1.50
		upper bound	-1.27	0.49	0.06	0.12	-1.46	-0.35	-0.21	-0.12	-0.18	-1.50
J. Employee and Spouse											
N= 53,422	Elasticity		-0.29	-0.65	-0.85	-0.96	-0.49	-1.09	-1.16	-1.10	-1.33	-1.50
		lower bound	-1.50	-0.72	-1.05	-1.21	-1.46	-1.40	-1.43	-1.45	-1.43	-1.50
		upper bound	0.93	-0.58	-0.65	-0.71	0.48	-0.77	-0.89	-0.76	-1.22	-1.50
K. Everyone											
N= 128,290	Elasticity		-0.33	-0.50	-1.01	-0.95	-0.36	-0.94	-1.17	-1.27	-1.28	-1.45
		lower bound	-1.50	-1.50	-1.50	-1.37	-1.47	-1.46	-1.42	-1.40	-1.44	-1.50
		upper bound	0.84	0.49	-0.52	-0.52	0.76	-0.41	-0.93	-1.15	-1.12	-1.40
L. Employee including injured											
N= 29,764	Elasticity		-0.27	-0.65	-0.67	-0.87	-0.63	-0.48	-1.31	-1.34	-1.37	-1.50
		lower bound	-1.50	-0.69	-0.69	-1.12	-1.49	-1.47	-1.42	-1.41	-1.44	-1.50
		upper bound	0.97	-0.61	-0.64	-0.63	0.24	0.52	-1.20	-1.26	-1.30	-1.49

Table OA11: Additional Specifications Continued

M. Employee - Do not require employees or their family members continuously enrolled											
N= 47,179	Elasticity	-0.39	-0.18	-0.57	-0.85	-0.90	-1.30	-1.34	-1.35	-1.37	-1.50
	lower bound	-1.50	-0.60	-0.61	-1.14	-1.21	-1.48	-1.43	-1.44	-1.45	-1.50
	upper bound	0.71	0.23	-0.53	-0.56	-0.58	-1.12	-1.24	-1.27	-1.29	-1.50
N. Employee - Only require employees, and NOT their family members, continuously enrolled											
N= 36,754	Elasticity	-0.34	-0.60	-0.62	-0.80	-0.52	-0.40	-1.30	-1.34	-1.30	-1.49
	lower bound	-1.50	-0.67	-0.64	-0.98	-1.49	-1.44	-1.43	-1.42	-1.45	-1.50
	upper bound	0.81	-0.53	-0.60	-0.63	0.45	0.64	-1.18	-1.26	-1.15	-1.49
O. Employee, 2003 Sample											
N= 30,077	Elasticity	-0.29	-0.60	-0.67	-0.72	-0.50	-0.41	-1.28	-1.34	-1.37	-1.50
	lower bound	-0.72	-0.62	-0.74	-0.84	-1.48	-1.47	-1.40	-1.41	-1.43	-1.50
	upper bound	0.14	-0.59	-0.60	-0.61	0.48	0.65	-1.17	-1.27	-1.31	-1.49
P. Instrument: All Injury Categories											
N= 26,790	Elasticity	-0.54	0.22	-0.60	-0.69	-0.50	-1.36	-1.35	-1.36	-1.38	-1.50
	lower bound	-1.50	-0.59	-0.62	-0.79	-1.17	-1.46	-1.41	-1.40	-1.43	-1.50
	upper bound	0.42	1.02	-0.58	-0.59	0.17	-1.26	-1.29	-1.32	-1.34	-1.50
Q. Instrument: Injuries to Children Only, Employees with Spouse Injuries Excluded from Sample											
N= 28,547	Elasticity	-1.29	-0.77	-1.24	-1.47	-1.36	-0.73	-0.39	-0.38	-0.34	-1.49
	lower bound	-1.50	-1.50	-1.50	-1.50	-1.50	-1.12	-0.49	-0.44	-0.39	-1.50
	upper bound	-1.07	-0.04	-0.98	-1.44	-1.22	-0.34	-0.30	-0.32	-0.28	-1.49
R. Instrument: Injuries to Spouses Only, Employees with Child Injuries Excluded from Sample											
N= 26,690	Elasticity	-0.24	-0.61	-0.64	-0.95	-0.45	-0.25	-1.33	-1.27	-1.38	-0.85
	lower bound	-1.50	-0.63	-0.67	-1.32	-1.50	-1.42	-1.41	-1.34	-1.44	-1.50
	upper bound	1.02	-0.59	-0.61	-0.58	0.60	0.93	-1.25	-1.19	-1.32	-0.19
S. Instrument: Simulated Spending on Injury											
N= 29,161	Elasticity	-0.25	0.13	-0.88	-0.64	-0.53	-0.55	-1.31	-1.32	-1.37	-1.43
	lower bound	-1.50	-0.63	-1.17	-0.67	-1.48	-1.46	-1.41	-1.41	-1.44	-1.50
	upper bound	1.00	0.89	-0.59	-0.62	0.43	0.35	-1.21	-1.24	-1.30	-1.35
T. Instrument: Dummies for Nine Injury Types											
N= 29,161	Elasticity	-0.38	-0.15	-0.88	-1.06	-0.90	-1.24	-1.29	-1.38	-1.39	-1.50
	lower bound	-1.50	-1.05	-1.13	-1.47	-1.48	-1.46	-1.41	-1.41	-1.44	-1.50
	upper bound	0.74	0.74	-0.63	-0.64	-0.31	-1.01	-1.17	-1.34	-1.34	-1.50
U. Instrument: Separate Dummies for Injuries in First and Second Half of Year											
N= 29,161	Elasticity	-0.28	0.15	-0.62	-0.72	-0.62	-1.20	-1.32	-1.36	-1.36	-1.50
	lower bound	-1.50	-0.66	-0.64	-0.82	-1.48	-1.45	-1.42	-1.42	-1.43	-1.50
	upper bound	0.94	0.97	-0.61	-0.63	0.25	-0.95	-1.21	-1.30	-1.29	-1.49

Lower and upper bounds for specifications J and K account for intra-family correlations.

Specifications L, M, and N explore alternative sample selection criteria in the employee sample. The main specification, A excludes injured employees on the grounds that the exclusion restriction is less likely to be satisfied for them - another family member's injury could be related to their injury through a mechanism other than price if both family members were injured at the same time, leading to a larger estimated elasticity. However, the cost of this sample selection is that it might reduce external validity of the findings, or it might bias the findings through selection on the dependent variable. Specification L leaves the 603 employees with own injuries in the estimation sample. Since the instrument is defined as an injury of another family member, not all of these employees have a value of one for the instrument - only 100 have another family injury. Estimates are similar or only slightly larger in the specifications that include injured employees, suggesting that sample selection to exclude injured employees does not have a large impact on the results.

Specifications M and N examine robustness to the major sample restriction - the requirement that everyone be continuously enrolled. I impose this restriction because the dependent variable measures year-end spending. Year-end spending will be artificially low if agents are not in the sample for the entire year, and we will not capture all responses to family injuries if agents are not in the sample for the entire year. However, as long as individuals are not differentially continuously enrolled based on family injury status, the results should be the same as they are in the main specification. In specification M, I do not require employees or their family members to be continuously enrolled, and in specification N, I only require employees, and not their family members, to be continuously enrolled. The results are generally similar.

Finally, I examine robustness to estimating the results using 2004 data instead of 2003 data. The elasticities presented in specification O show remarkably similar patterns. The similarity of the estimates between 2003 and 2004 provides some evidence of robustness, and it suggests that price responsiveness did not change between 2003 and 2004.

Online Appendix 12 Robustness to Specification of the Instrument

In this section, I examine the robustness of my results to alternative specifications of the instrumental variable. As discussed in Online Appendix 15, I restrict the categories of injuries in my instrument to those for which employee spending is similar before the injuries. If employees in families with some categories of injuries spend more even before the injuries occur, estimates from specifications that use those categories will result in elasticity estimates away from zero. In specification P of Tables OA10 and OA11, I specify the instrument to include *all* injury categories. The results are larger than those in the main specification at most quantiles, but only slightly, suggesting that some injury categories that are not included in my main specification might be endogenous, but that any bias that they would generate is small. In all other specifications, I consider only the injury categories included in the main specification.

Next, I explore alternative specifications of the instrument to investigate a specific channel for violation of the exclusion restriction. There is potential cause for concern if family injuries affect family income and family income affects expenditure. I cannot control for income directly because I do not observe it. However, I can informally investigate the role of income effects by estimating separate specifications based on injuries to spouses and injuries to children. If there are large income effects due to the injury of a wage earner, we might expect an employee's response to a spouse's injury to be different than an employee's response to a child's injury. In specification Q, I only include child injuries in the instrument and drop employees with other family injuries; in specification R, I only include spouse injuries in the instrument and drop employees with other family injuries. The logarithmic specification with just child injuries gives almost the exact same point estimates as the main specification, which is not surprising given that 4/5 of the injuries in the sample are to children. The specification with just spouse injuries, which is not as well identified, also yields point estimates that are the similar in magnitude but slightly smaller, suggesting that variation in the estimates due to child vs. spouse injuries is not large relative to the main elasticity estimates.

In the next three specifications, I incorporate more variation into the specification of the instrument by exploiting spending on injuries, injury categories, and the timing of injuries. In the main specification, the instrument is a dummy variable that indicates whether another family member had an injury. I do not specify the instrument as the amount of spending on the family injury because such a specification could violate the exclusion restriction if family members see the same expensive doctors or consume care in the same expensive geographic area. However, by restricting our instrument to a dummy variable, I lose some meaningful variation. In specification S, I incorporate the amount of spending on the injury using simulation techniques pioneered by Currie and Gruber (1996a,b). First, within each of the seven injury categories, I calculate mean spending over all individuals with that injury. Then, I assign this mean to the family members of the injured, removing only their family member's spending from the mean. In this way, I am able to use variation in the magnitude of spending on the injury, excluding variation that could be correlated within families. Although it would have been possible to create a plan-specific measure of average injury spending, in practice, some of the cell sizes are very small. As compared to the main specification, the results in specification S are slightly smaller, but in a similar range. The smaller estimates seem to be an artifact of the wider confidence intervals. I might have expected the confidence intervals to be tighter because of the additional variation in the instrument, but it seems that this specification sacrifices power because of the greater demands that it puts on the data.

Specification T incorporates more variation into the instrument by specifying it as a vector of dummy variables for each of the nine injury categories included in the instrument, listed in Table OA3. The results are very similar to the main results. Finally, in specification U, I exploit additional variation in injury timing in the specification of the instrument. I generate two instruments by interacting the family injury dummy with an indicator for the half of the year when the injury occurred. Again, the results are very similar to those from the main specification. The next section considers the implications of within-year injury timing for the estimates and demonstrates that the estimates should indeed be similar regardless of injury timing.

Online Appendix 13 Implications of Within-Year Injury Timing

In this section, I examine the implications of injury timing for my instrumental variable estimates. To do so, I examine separate CQIV specifications in which I only include injuries in the first half of the year or only include injuries in the second half of the year in the determination of the instrument, excluding individuals with family injuries in the other half of the year from the sample. In the first panel of Tables OA12 and OA13, I present estimates with the dependent variable in logarithmic and level form. The estimates are very similar, regardless of the timing of the family injury used. This finding is consistent with the model discussed in Online Appendix 5: after an injury occurs, it takes time for the family member's marginal price, as well as his spending, to respond. On net, if the family member's marginal price responds at roughly the same rate as his spending responds, the IV estimate, which reflects the spending response (the reduced form) divided by the price response (the first stage), will be roughly the same regardless of the injury timing.

In the bottom panel of Tables OA12 and OA13, I examine the spending response to the injury (the reduced form), and the price response to the injury (the first stage) separately. To

Table OA12: Timing of Family Injury

Dependent Variable: Ln(Expenditure).

2004 Sample		Censored Quantile IV 10	20	30	40	50	60	70	80	90		
Injury in first half of the year											**Tobit IV**	
N= 27,104	Elasticity	-0.74	-0.74	-1.06	-1.45	-1.49	-1.41	-1.38	-1.43	-1.40	-1.43	
	lower bound	-1.49	-1.49	-1.47	-1.49	-1.50	-1.50	-1.47	-1.47	-1.45	-1.50	
	upper bound	0.00	0.02	-0.66	-1.40	-1.48	-1.33	-1.30	-1.38	-1.36	-1.37	
Injury in second half of the year											**Tobit IV**	
N= 27,030	Elasticity	-1.24	-0.58	-1.05	-1.44	-1.49	-1.34	-1.33	-1.32	-1.37	-1.31	
	lower bound	-1.49	-1.49	-1.48	-1.49	-1.50	-1.50	-1.45	-1.46	-1.46	-1.49	
	upper bound	-1.00	0.33	-0.61	-1.39	-1.48	-1.17	-1.21	-1.18	-1.29	-1.13	
		Censored Quantile Regression										**first stage**
Injury in first half of the year - reduced form											**Tobit**	**OLS**
N= 27,104	Family Injury	.	0.20	0.32	0.52	0.44	0.40	0.35	0.35	0.29	0.65	-0.10
	lower bound	.	-0.20	-0.10	0.23	0.22	0.23	0.21	0.19	0.13	0.36	-0.12
	upper bound	.	0.60	0.74	0.82	0.66	0.58	0.49	0.51	0.45	0.94	-0.08
Injury in second half of the year - reduced form											**Tobit**	**OLS**
N= 27,030	Family Injury	.	0.17	0.43	0.37	0.28	0.25	0.23	0.31	0.28	0.50	-0.10
	lower bound	.	0.00	0.00	0.11	0.10	0.09	0.05	0.14	0.12	0.23	-0.12
	upper bound	.	0.34	0.86	0.63	0.45	0.41	0.41	0.48	0.44	0.77	-0.08

In first stage results, dependent variable is year-end price.

Table OA13: Timing of Family Injury

Dependent Variable: Expenditure.

2004 Sample		Censored Quantile IV 10	20	30	40	50	60	70	80	90			
Injury in first half of the year											**Tobit IV**		
N= 27,104	Elasticity	-0.26	-0.58	-0.62	-0.78	-0.66	-1.29	-1.34	-0.80	-1.39	-1.49		
	lower bound	-1.50	-0.63	-0.63	-0.93	-1.47	-1.48	-1.41	-1.43	-1.45	-1.50		
	upper bound	0.98	-0.53	-0.62	-0.62	0.16	-1.09	-1.26	-0.17	-1.33	-1.49		
Injury in second half of the year											**Tobit IV**		
N= 27,030	Elasticity	-0.28	-0.64	-0.65	-0.83	-0.55	-1.23	-1.32	-1.31	-1.39	-1.24		
	lower bound	-1.50	-0.67	-0.69	-1.04	-1.50	-1.44	-1.41	-1.41	-1.45	-1.50		
	upper bound	0.93	-0.60	-0.62	-0.62	0.39	-1.03	-1.22	-1.21	-1.32	-0.98		
		Censored Quantile Regression											**first stage**
Injury in first half of the year - reduced form											**Tobit**	**OLS**	**OLS**
N= 27,104	Family Injury	.	26	16	46	71	109	209	472	1,175	907	526	-0.10
	lower bound	.	0	-7	13	27	48	80	190	325	433	187	-0.12
	upper bound	.	51	40	78	115	170	339	754	2,025	1,380	865	-0.08
Injury in second half of the year - reduced form											**Tobit**	**OLS**	**OLS**
N= 27,030	Family Injury	.	28	27	29	48	75	154	394	924	603	347	-0.10
	lower bound	.	0	2	-7	13	5	31	153	234	268	85	-0.12
	upper bound	.	57	53	66	83	144	277	635	1,613	939	609	-0.08

In first stage results, dependent variable is year-end price.

examine the reduced form, I estimate CQR, Tobit, and OLS specifications.[17] These results, especially the Tobit results, show that the expenditure response is larger for injuries that occur in the first half of the year than it is for injuries that occur in the second half of the year. The first stage OLS results also show that the *price* response is larger for injuries that occur in the first half of the year. The combination of these findings explains why the CQIV results are generally not sensitive to injury timing.

[17] I only estimate an OLS specification of the reduced form for the levels model because it is unclear how to model individuals with zero expenditure in an OLS model of the logarithm of expenditure.

Online Appendix 14 Analysis of Data on Smaller Families

Using data beyond my estimation sample, I examine the spending behavior of employees in families with fewer than four individuals (the main results are based on employees in families of four or more). People in smaller families have the potential to serve the basis for an indirect test of the exclusion restriction, but the test has several potential flaws. The results do not yield support for the exclusion restriction, but I report them here in the interest of transparency.

The exclusion restriction requires that one family member's injury can only affect another family member's expenditure through its effect on his marginal price. If program rules dictate that cost sharing interactions *cannot* occur, one family member's injury should *not* be related to another family member's medical expenditure. At the firm that I study, in policies for families of two, one family member's spending has no mechanical effect on another family member's marginal price as shown in Table 1 in the paper. Therefore, any effects of one family member's injury on another family member's spending presumably operate through another channel. Although the exclusion restriction is not an econometrically testable restriction in the main sample of families of four or more, evidence that there is no effect of one family member's injury on another family member's spending in a family of two would support the validity of the exclusion restriction in the main sample.

To formalize this test, I estimate the following specification with censored quantile regression:

$$
\begin{aligned}
Y &= \max(Y^*, C) = T((Y_i)^*) \\
Y^* &= W'\theta(U) + \xi(U)Z
\end{aligned}
$$

where the regressors are defined above. This specification differs from the main specification only in that it examines the reduced form effect of the family injury on Y directly. A traditional instrumental variable specification would not be informative here because the first stage should not exist in families of two.[18]

The main limitation of this test is that if employees in families of two are different from employees in families of four or more, the effect of one member's injury on another member's spending could be nonzero in families of two even if there is no violation of the exclusion restriction in families of four or more. Indeed employees in families of two have different observable characteristics than employees in families of four or more, so it is unclear if effects of one family member's injury on another family member's spending should be comparable across samples. I investigate several sample restrictions to make the samples of employees in families of both sizes similar in terms of observable characteristics, but unobservable characteristics could still differ.

Column 7 of Table OA2 in the paper presents summary statistics on employees in all families of two. Comparison of year of birth with column 1 shows that this population is much older than the population in families of four - it consists of many older "empty nesters."

[18] If we think of the instrumental variable specification as the reduced form estimate divided by the first stage estimate, the instrumental variable estimate will not exist when the first stage estimate is zero.

Furthermore, only 27% of employees in families of two consume zero care, as opposed to 36% in families of four or more. Employees in families of two have much higher average expenditures on medical care than their counterparts in families of four or more ($2,615 vs. $1,414). Given that employees in families of two consume more medical care, we might be more likely to observe spurious effects of other family injuries on spending in the families of two than we are in the main sample. We should keep this caveat in mind when interpreting the results.

I first report results from the sample of employees in all families of two without any sample restrictions. If the test is valid and the exclusion restriction holds, we expect to see no effect of another family member's injury on the employee spending in families of two, but the CQR results in specification A of Tables OA14 and OA15 show that employees with a family injury spend from 0 to 41% more or from $0 to $736 more than employees without family injuries. In larger families, we expect to see a larger impact of one family member's injury on another family member's expenditure. As shown in specification G of Tables OA14 and OA15, the coefficients in the family specification suggest that employees with an injured spouse or child spend 15 to 29% more or from $20 to $813 more on their own medical care than employees without family injuries. Although the Tobit coefficient is slightly larger in the sample of families of four or more, suggesting that there is more of an effect of one family member's injury on the employee's expenditure in larger families, the Tobit coefficient in families of two is not a precisely estimated zero. It is possible that employees in families of two are different from employees in families of four or more, invalidating the test.

One reason why families of two might be different from larger families is that identification in the smaller families comes mainly from injuries to spouses, and identification in the larger families comes from injuries to children as well as spouses. In specification B, I restrict the sample to employee-dependent families of two, and in specification H, I restrict the main sample so that it does not include employees with spouse injuries; both samples are identified from injuries to non-spouse dependents. In specifications C and I, I restrict the samples so that they are identified from injuries to spouses. Again, the results are somewhat larger in the specifications in families of four, but the results are nonzero in families of two.

Yet another concern with the comparison by family size is that employees in families of two are much older than the employees in larger families. In specifications D and J, I restrict both samples to include employees aged 40 and under. Even though the unrestricted sample of employees in families of two is much larger than the sample of employees in families of four or more, it only includes 6,951 employees under age 40, but the main sample includes 18,972 employees under age 40. Similarly, in specifications E and K, I restrict both samples to include employees over age 40. As shown, in families of four or more, the effect of a family member's injury on the employee's expenditure is larger if the employee is older. The mechanism for this comparison is unclear, but it suggests that we are more likely to find relationships in the older sample of families of two than we are in the younger main sample based on age. Unfortunately, the younger sample of families of two is so small that the results are imprecise.

For completeness, I also present results estimated on families of three. As discussed in the paper, there are some cost sharing interactions in families of three that occur through the stoploss, so the results in families of three should show some effect of one family member's injury on the employee's expenditure. This is indeed the case, as shown in specification F.

Although the results would be compelling if the specification restricted to families of two yielded a precisely estimated zero, there are several reasons why it would not in the absence of a violation of the exclusion restriction. For example, if employees do not understand the interaction between the individual and family deductibles, then they might think that their own price will go down after their family member has an injury regardless of family size. Especially since many members of the sample of families of two appear to be empty-nesters, they might behave as if another family member's injury will affect their spending, as it did when they had children on their policy in the past. In sum, this analysis does not provide support for the exclusion restriction, but there are many reasons to question its usefulness.

Table OA14: Family Injuries in Smaller and Larger Families

Dependent variable: Ln(Expenditure)

2004 Sample			Censored Quantile Regression 10	20	30	40	50	60	70	80	90	Tobit
Families of Two												
A. Employees in All Familes of Two												
N=	54,889	Family Injury	0.00	0.41	0.31	0.36	0.23	0.24	0.20	0.21	0.16	0.48
		lower bound	0.00	0.00	0.12	0.19	0.11	0.10	0.07	0.06	0.01	0.29
		upper bound	0.00	0.82	0.51	0.53	0.36	0.39	0.32	0.37	0.30	0.67
B. Employees in Employee-Child Families of Two												
N=	17,338	Family Injury	0.00	0.59	-0.11	0.15	0.25	0.13	0.20	0.14	0.08	0.35
		lower bound	0.00	-0.22	-0.52	-0.14	-0.03	-0.22	-0.04	-0.14	-0.19	-0.06
		upper bound	0.00	1.40	0.29	0.45	0.54	0.47	0.43	0.42	0.35	0.77
C. Employees in Employee-Spouse Families of Two												
N=	37,551	Family Injury	0.34	0.40	0.46	0.33	0.27	0.27	0.19	0.17	0.14	0.49
		lower bound	-0.17	0.00	0.18	0.15	0.11	0.09	0.02	0.01	-0.03	0.28
		upper bound	0.85	0.79	0.73	0.50	0.43	0.45	0.36	0.32	0.32	0.70
D. Employees 40 and under in Employee-Spouse Families of Two												
N=	6,951	Family Injury	.	2.41	0.30	0.29	0.27	0.22	0.21	0.22	0.15	0.50
		lower bound	.	0.00	-0.05	0.00	-0.08	-0.04	-0.03	-0.09	-0.09	0.10
		upper bound	.	4.81	0.65	0.58	0.62	0.48	0.46	0.53	0.39	0.90
E. Employees over 40 in Employee-Spouse Families of Two												
N=	30,600	Family Injury	0.00	0.58	0.37	0.30	0.25	0.25	0.18	0.16	0.09	0.44
		lower bound	0.00	0.06	0.13	0.15	0.11	0.10	0.02	0.02	-0.10	0.24
		upper bound	0.00	1.11	0.60	0.45	0.40	0.40	0.34	0.30	0.28	0.64
Families of Three												
F. Employees in All Families of Three												
N=	25,482	Family Injury	0.00	2.15	0.38	0.34	0.31	0.20	0.13	0.07	0.04	0.64
		lower bound	0.00	0.00	0.00	0.12	0.14	0.07	0.00	-0.10	-0.13	0.39
		upper bound	0.00	4.31	0.77	0.55	0.48	0.33	0.25	0.24	0.21	0.90
Families of Four or More												
G. Employees in Families of Four or More												
N=	29,161	Family Injury	.	0.15	0.37	0.46	0.36	0.31	0.30	0.31	0.29	0.57
		lower bound	.	0.00	0.00	0.25	0.22	0.17	0.18	0.20	0.16	0.36
		upper bound	.	0.30	0.73	0.66	0.49	0.45	0.41	0.43	0.42	0.78
H. Employees in Families of Four or More - Excluding Employees with Spouse Injuries												
N=	28,547	Family Injury	.	0.00	0.42	0.44	0.36	0.35	0.33	0.35	0.32	0.59
		lower bound	.	0.00	0.00	0.18	0.17	0.20	0.20	0.22	0.18	0.36
		upper bound	.	0.00	0.84	0.69	0.56	0.50	0.46	0.48	0.46	0.82
I. Employees in Families of Four or More - Excluding Employees with Child Injuries												
N=	26,690	Family Injury	.	0.62	0.30	0.44	0.28	0.27	0.27	0.23	0.10	0.54
		lower bound	.	-0.01	-0.18	0.06	-0.03	0.04	0.07	0.01	-0.18	0.15
		upper bound	.	1.25	0.78	0.82	0.60	0.49	0.47	0.45	0.39	0.93
J. Employees 40 and under in Families of Four or More												
N=	18,972	Family Injury	.	-0.03	0.20	0.30	0.40	0.32	0.23	0.29	0.22	0.54
		lower bound	.	-0.11	-0.05	0.09	0.23	0.19	0.09	0.13	0.06	0.31
		upper bound	.	0.06	0.46	0.51	0.58	0.46	0.36	0.45	0.38	0.76
K. Employees over 40 in Families of Four or More												
N=	10,189	Family Injury	.	0.00	0.37	0.59	0.40	0.38	0.44	0.45	0.36	0.61
		lower bound	.	0.00	0.00	0.15	0.15	0.11	0.24	0.21	0.15	0.30
		upper bound	.	0.00	0.74	1.02	0.64	0.64	0.65	0.69	0.58	0.91

Family of Two controls: male dummy, plan (saturated), census region (saturated), salary dummy (vs. hourly), count family born 1944 to 1953, count family born 1954 to 1963, count family born 1964 to 1973, count family born 1974 to 1983, count family born 1984 to 1993.

Family of Four controls: family of two controls, spouse on policy dummy, YOB of oldest dependent, YOB of youngest dependent, family size (saturated with 8-11 as one group), count family born 1994 to 1998, count family born 1999, count family born 2000, count family born 2001, count family born 2002, count family born 2003, count family born 2004.

Table OA15: Family Injuries in Smaller and Larger Families

Dependent variable: Expenditure

		Censored Quantile Regression									
2004 Sample		10	20	30	40	50	60	70	80	90	Tobit
Families of Two											
A. Employees in All Familes of Two											
N= 54,889	Family Injury	0	5	19	29	121	70	288	83	736	909
	lower bound	0	-39	-51	-75	-46	-235	-247	-1,326	-539	445
	upper bound	0	48	88	134	288	375	823	1,491	2,011	1,373
B. Employees in Employee-Child Families of Two											
N= 17,338	Family Injury	0	25	1	12	39	78	151	401	490	373
	lower bound	0	0	-37	-61	-25	-60	-39	-134	-538	-224
	upper bound	0	51	39	86	102	216	341	935	1,519	971
C. Employees in Employee-Spouse Families of Two											
N= 37,551	Family Injury	.	16	47	67	130	181	272	919	342	1,014
	lower bound	0	-41	-13	-25	-2	-303	-253	-155	-1,778	419
	upper bound	46	73	106	159	262	666	798	1,993	2,461	1,608
D. Employees 40 and under in Employee-Spouse Families of Two											
N= 6,951	Family Injury	.	35	39	59	54	88	107	449	511	502
	lower bound	.	0	-2	0	2	-27	-25	-42	-280	8
	upper bound	.	70	81	118	105	203	238	940	1,302	997
E. Employees over 40 in Employee-Spouse Families of Two											
N= 30,600	Family Injury	0	16	38	61	50	84	739	729	237	935
	lower bound	0	-30	-16	-23	-256	-542	33	-574	-1,907	385
	upper bound	0	62	93	145	356	709	1,445	2,031	2,382	1,484
Families of Three											
F. Employees in All Families of Three											
N= 25,482	Family Injury	-29	30	41	79	78	103	122	173	314	524
	lower bound	-58	0	12	23	28	36	4	-78	-340	171
	upper bound	0	61	70	135	128	170	240	424	969	878
Families of Four or More											
G. Employees in Families of Four or More											
N= 29,161	Family Injury	.	23	20	39	54	94	172	379	813	727
	lower bound	.	0	2	14	21	35	70	91	-148	422
	upper bound	.	47	38	63	88	154	274	667	1,775	1,032
H. Employees in Families of Four or More - Excluding Employees with Spouse Injuries											
N= 28,547	Family Injury	.	10	18	38	56	102	189	468	1,269	839
	lower bound	.	0	0	11	17	52	83	214	633	483
	upper bound	.	20	35	64	95	152	296	722	1,905	1,195
I. Employees in Families of Four or More - Excluding Employees with Child Injuries											
N= 26,690	Family Injury	.	59	25	33	52	101	144	430	582	476
	lower bound		0	-7	-18	-13	-9	-29	82	-486	33
	upper bound		118	56	83	117	211	317	779	1,650	920
J. Employees 40 and under in Families of Four or More											
N= 18,972	Family Injury	.	26	18	92	51	85	138	290	602	488
	lower bound		0	-4	5	12	19	48	80	35	261
	upper bound		53	40	180	89	152	228	500	1,168	715
K. Employees over 40 in Families of Four or More											
N= 10,189	Family Injury	.	24	22	41	83	158	336	708	1,737	1,197
	lower bound		0	-8	8	24	61	158	221	478	536
	upper bound		48	51	74	143	256	515	1,194	2,995	1,859

Family of Two controls: male dummy, plan (saturated), census region (saturated), salary dummy (vs. hourly), count family born 1944 to 1953, count family born 1954 to 1963, count family born 1964 to 1973, count family born 1974 to 1983, count family born 1984 to 1993.

Family of Four controls: family of two controls, spouse on policy dummy, YOB of oldest dependent, YOB of youngest dependent, family size (saturated with 8-11 as one group), count family born 1994 to 1998, count family born 1999, count family born 2000, count family born 2001, count family born 2002, count family born 2003, count family born 2004.

Online Appendix 15 Mechanisms: Timing of Injury Response

A key feature of my identification strategy is that the response to injuries should occur after they happen but not before. Accordingly, I restrict my instrument to categories of injuries for which employees that have them in their families do not spend more on their own medical care before the injuries occur. I first describe the process of selecting the injury categories and show that spending does not respond during the portion of the year before the injuries occur. Next, using longitudinal data, I show that spending also does not appear to respond in the textityear before the injuries occur.

My approach is motivated by that of Card et al. (2009), who select a subset of diagnoses that appear to have similar rates of appearance in emergency rooms on weekdays and weekends. Using only this subset of diagnosis categories, they measure whether the start of Medicare eligibility at age 65 results in lower mortality. In my approach, I select a set of injury categories for which employees in families with injuries do not appear to spend more than employees in similar families without injuries in the part of the year before the injury has occurred.

To select the categories of injuries listed in column 2 of Table OA3, I use data on spending patterns *within* the year. I create a dataset such that each employee with a family injury in a given category has a single observation for spending before the month of the family injury. For example, if the injury is in March, the observation reports cumulative spending through February. To control for seasonality in medical spending, I organize the data so that each employee without any family injury can serve as a control for any employee with a family injury, regardless of injury month. Therefore, each employee without any family injuries has an observation for cumulative spending before each month. For example, a given employee without any family injuries has one observation for January spending, another observation for January through February spending, another observation for January through March spending, etc. For each injury category, I then run a regression in which I predict spending with a dummy variable for the eventual injury, clustering by employee since employees without any family injuries have multiple observations. I repeat the regression for all categories of family injuries shown in Table OA3. The sample size varies slightly across specifications because I eliminate the employees that themselves will have the injury as I do in the baseline elasticity estimates, but mean spending is always approximately \$700 before the injury.

I report the coefficient on the dummy variable for the eventual injury by category in column 6 of Table OA3, and I report the associated upper and lower bounds of the 95% confidence intervals in columns 7-8. The injuries included in the instrument are those for which the value of the coefficient is less than 100, indicating that employees with upcoming family injuries in the given category spend no more than \$100 more than employees with no upcoming family injuries in that category in the period before the injury. Note that spending appears to be lower, sometimes much lower, in families with upcoming injuries in some categories. However, as the primary concern with my identification strategy is that employees in families with injuries could be higher spenders than employees in families without injuries, leading to elasticity estimates that are too large, I include all categories with negative coefficients in the instrument.[19]

Having selected the injury categories for the instrument, I test that once I have included all of these categories, the result still holds that spending does not respond to an injury before it occurs. Further, I also examine spending during the part of the year after the injury occurs to examine the mechanism behind my main results. Spending after the injury occurs should be higher for families with injuries.

To implement the test, I create a new dataset. For each employee with a family injury, there is one observation for the period before the first injury and one observation for the period after the injury, excluding the injury month itself. For example, an employee with a family injury in March has an observation for spending from January through February and another observation for spending from April through December. Each employee without a family injury has an observation for the part of the year before as well as after each month. For example, an employee without a family injury has one observation for January spending and one observation for March through December spending; that same employee also has one observation for January through February spending and one observation for April through December spending, etc.

I run a regression in which I compare families with injuries to families without injuries, after the injury relative to before the injury, controlling for all of the covariates in the main specification, as well as a full set of injury month fixed effects. I report the results using the new selected injuries as a single category in the top panel of Table OA16. In column 1, the coefficient on "Family Injury," defined as in the main CQIV specifications, shows that individuals with family injuries spend about 23 cents more in the period before the injury relative to people without injuries. We can be 95% confident that they spend between 80 dollars less and 80 dollars more than similar families without injuries in the period before the injury, relative to an average of approximately 700 dollars. The coefficient on "After," a dummy variable for the period after the injury, indicates that there are statistically significant seasonal patterns in medical spending. The coefficient on "Family Injury x After" indicates that after controlling for seasonality by comparing employees with family injuries to employees without family injuries, employees with family injuries spend approximately 400 dollars more after the injury occurs. Taken together, these results suggest that people with family injuries as defined by my instrument do not spend more before the injuries occur, but they do respond to incentives by spending more after the injury occurs.

The next two columns show that a similar pattern holds for dependent variables that indicate any expenditure or any outpatient visit. Although people with injuries in their families do not tend to have statistically more inpatient visits before the injury occurs, they also do not tend to have statistically more inpatient visits after the injury occurs, as shown in the fourth column. This finding is consistent with the evidence in Online Appendix 16 that shows that employees generally respond to injuries on the outpatient visit margin. Overall, the results of this test lend support to the main identification assumption.

I perform another exercise using the longitudinal data, which also lends some support to the main identification assumption. To perform the test, I create a dataset that includes

[19] I do not select the new categories of injuries based on the confidence intervals because the size of the confidence intervals has a strong relationship to the total count of injured individuals; if I used the confidence intervals in my selection criteria, I would be much more likely to reject the null hypothesis of no difference for the categories with higher counts, retaining only those categories that offer little power.

Table OA16: Within-Year Test

Dependent Variable:	Expenditure	Expenditure >0	Outpatient Visit >0	Inpatient Visit >0
Family Injury x After	363.08	0.048	0.047	0.007
lower bound	166.62	0.025	0.024	-0.001
upper bound	559.54	0.071	0.071	0.015
Family Injury	0.23	0.011	0.009	0.000
lower bound	-79.33	-0.007	-0.008	-0.005
upper bound	79.78	0.028	0.026	0.006
After	37.27	0.028	0.008	-0.001
lower bound	11.62	0.024	0.004	-0.002
upper bound	62.92	0.033	0.012	0.000

all employees whose full families are enrolled in 2003 and 2004, and I drop individuals with family injuries in 2003 and with injuries themselves in either year, resulting in a dataset with 17,092 observations (requiring longitudinal data severely limits the sample, just as requiring continuous enrollment limits the main sample). I run a regression to predict whether having a family injury in 2004 predicts spending in the previous year of 2003. For this test, I regress spending in 2003 on a dummy variable for having a family injury in 2004, controlling for all of the 2003 and 2004 values of the covariates. The coefficient on "New Family Injury, 2004 Only," a dummy variable that indicates having a family injury in 2004 but not 2003, suggests that individuals who will have a family injury in 2004 spend around \$109 less in 2003 than similar individuals who will not have a family injury in 2004, with a 95% confidence interval of -\$505 to \$286. This result is not statistically significant, likely because restricting the sample to employees for whom longitudinal data are available severely restricts the sample size from 29,886 employees in the main 2003 estimation sample to 17,092. However, the coefficient is small relative to mean 2003 spending in the sample of approximately \$1,000. Overall, the results of this test are consistent with the main identification assumption.

Online Appendix 16 Mechanisms: Visits and Outpatient Spending

By estimating models that specify the dependent variable in terms of visits instead of expenditure, we can examine one potential mechanism through which agents respond to prices: by deciding whether to go to the doctor. Table OA17 reports results that specify the dependent variable in Equation 2a as the logarithm or number of total visits, outpatient visits, and inpatient visits. The first row in each specification shows the coefficients. The CQIV corner coefficients across all quantiles suggest that as the marginal price goes from one to zero, patients reduce total visits by up to 2.83 percent, or up to 18.78 visits. If we transform these coefficients into elasticities, we can compare the price elasticity of visits to the price elasticity of expenditure from previous specifications. Holding covariates constant, the median elasticity in specification A1 suggests that if we have two groups of individuals and one group faces prices that are 1% higher, the median number of visits in that group will be 1.02% lower. Tobit IV arc elasticities have a similar order of magnitude, as do arc elasticities calculated from a Poisson regression.[20] As compared to the expenditure elasticity, the visit elasticity is of roughly the same order of magnitude but slightly smaller. Also, the visit elasticity is larger at higher conditional quantiles, just as the expenditure elasticity in the levels specification is larger at higher conditional quantiles. The comparison between the visit elasticity and the expenditure elasticity suggests that a large part of the expenditure elasticity occurs on the visit margin, and a smaller part occurs on the intensive margin of spending within a given visit.

Specifications B and C in Table OA17 show that most of the visit elasticity comes from outpatient visits, and there is a very small elasticity of inpatient visits. The inpatient visit results are much less reliable because few bootstrap replications converge at all quantiles except the 0.9 quantile. Unreported elasticity estimates from a CQIV specification with outpatient expenditure as the dependent variable are very similar to the elasticity estimates in the main specification.

Quantile estimators are less sensitive to extreme values than mean estimators. However, to demonstrate that individuals with the highest expenditures are not driving the results, I estimate the main specification at the very highest conditional quantiles. Even at very highest conditional quantiles, where we expect more inpatient expenditures conditional on observed characteristics, the unreported estimated elasticities remain fairly stable.

Online Appendix 17 Comparison to Rand

My estimates are an order of magnitude larger than those commonly cited from the Rand experiment. There could be a multitude of reasons for this discrepancy, including a possible change in the underlying expenditure elasticity over the decades between the Rand study and my study or a difference in behavior between people in experimental plans and people in actual plans. Here, I examine differences in methodology between my estimates and the Rand estimates.

To induce subjects to participate in the Rand experiment, researchers had to guarantee that participants would be subject to very low out-of-pocket costs, so all plans in the experiment had a yearly stoploss of $1,000 or less in 1974-1982 dollars. Furthermore, each year, all families were given lump sum payments that equaled or exceeded their out-of-pocket payments. The experimenters randomized families into plans with initial marginal prices of 0%, 25%, 50%, and 95%, but after family spending reached the stoploss, marginal price was zero for the rest of the year, regardless of plan. In practice, the stoploss was binding for a large fraction (roughly 20%) of participants. Approximately 35% of individuals in the least generous plan exceeded the stoploss, as did approximately 70% of individuals who had any inpatient care. To put these rates in a broader context, less than 4% of individuals met the stoploss in my non-experimental data.

Rand researchers recognized that the stoploss affected their ability to calculate the price

[20] Poisson regression, also known as quasi-MLE Poisson regression if the researcher only believes that he has specified the conditional mean correctly, is a popular regression for count data models. The raw coefficients are not simple to interpret, but the arc elasticities should be directly comparable to the other arc elasticities.

Table OA17: Visits

2004 Employee Sample		Censored Quantile IV 10	20	30	40	50	60	70	80	90	Tobit IV	Poisson
A1. Dependent Variable: Ln(All Visits)												
N= 29,161	Year-end price	-0.01	-0.51	-1.11	-1.26	-2.03	-2.25	-2.52	-2.74	-2.83	-1.94	
	lower bound	-1.24	-1.27	-1.60	-1.72	-3.23	-2.99	-2.60	-3.55	-3.50	-2.29	
	upper bound	1.22	0.24	-0.61	-0.80	-0.82	-1.52	-2.43	-1.93	-2.16	-1.59	
	Elasticity	-0.10	-0.20	-0.58	-0.76	-1.02	-1.03	-1.15	-1.15	-1.19	-1.01	
	lower bound	-0.69	-0.54	-0.83	-1.07	-1.29	-1.25	-1.17	-1.33	-1.33	-1.18	
	upper bound	0.49	0.14	-0.33	-0.44	-0.75	-0.81	-1.12	-0.97	-1.05	-0.85	
B1. Dependent Variable: Ln(Outpatient Visits)												
N= 29,161	Year-end price	-0.21	-0.46	-0.90	-1.11	-2.05	-2.24	-2.46	-2.68	-2.71	-1.94	
	lower bound	-1.48	-1.15	-1.43	-1.61	-3.19	-2.86	-2.48	-3.47	-3.39	-2.28	
	upper bound	1.05	0.24	-0.38	-0.61	-0.90	-1.63	-2.43	-1.90	-2.02	-1.59	
	Elasticity	-0.11	-0.15	-0.53	-0.69	-1.05	-1.04	-1.13	-1.14	-1.16	-1.01	
	lower bound	-0.71	-0.45	-0.83	-1.05	-1.29	-1.22	-1.14	-1.32	-1.31	-1.17	
	upper bound	0.48	0.14	-0.22	-0.33	-0.81	-0.86	-1.12	-0.96	-1.00	-0.85	
C1. Dependent Variable: Ln(Inpatient Visits)												
N= 29,161	Year-end price	0.00	.	.	0.00	0.00	0.00	0.00	-0.20	-1.34	-0.14	
	lower bound	0.00	.	.	0.00	0.00	0.00	0.00	-0.39	-2.31	-0.25	
	upper bound	0.00	.	.	0.00	0.00	0.00	0.00	0.00	-0.37	-0.04	
	Elasticity	0.00	.	.	0.00	0.00	0.00	0.00	-0.03	-0.10	0.00	
	lower bound	0.00	.	.	0.00	0.00	0.00	0.00	-0.05	-0.18	0.00	
	upper bound	0.00	.	.	0.00	0.00	0.00	0.00	0.00	-0.03	0.00	
A2. Dependent Variable: All Visits												
N= 29,161	Year-end price	0.05	-3.50	-6.63	-9.23	-7.20	-3.85	-8.34	-11.31	-18.78	-4.99	-1.23
	lower bound	-1.40	-5.39	-8.34	-11.39	-14.30	-6.65	-8.75	-13.35	-22.27	-6.39	-1.33
	upper bound	1.51	-1.61	-4.92	-7.07	-0.11	-1.06	-7.93	-9.27	-15.29	-3.58	-1.13
	Elasticity	-0.56	-0.65	-0.75	-1.02	-0.79	-1.00	-1.16	-1.15	-1.19	-1.49	-1.11
	lower bound	-1.50	-0.68	-0.88	-1.39	-1.49	-1.36	-1.20	-1.31	-1.32	-1.50	-1.26
	upper bound	0.37	-0.62	-0.63	-0.64	-0.08	-0.64	-1.12	-0.99	-1.06	-1.49	-0.95
B2. Dependent Variable: Outpatient Visits												
N= 29,161	Year-end price	-0.35	-2.90	-5.68	-8.36	-6.30	-7.04	-8.09	-11.55	-18.59	-4.90	-1.22
	lower bound	-2.11	-4.66	-7.44	-10.53	-12.75	-12.28	-8.75	-13.35	-22.46	-6.28	-1.32
	upper bound	1.40	-1.14	-3.91	-6.19	0.15	-1.80	-7.43	-9.74	-14.72	-3.52	-1.12
	Elasticity	-0.56	-0.64	-0.85	-1.04	-0.70	-1.12	-1.15	-1.18	-1.19	-1.49	-1.10
	lower bound	-1.50	-0.74	-1.07	-1.44	-1.49	-1.50	-1.20	-1.31	-1.32	-1.50	-1.26
	upper bound	0.38	-0.55	-0.63	-0.64	0.08	-0.74	-1.10	-1.05	-1.06	-1.49	-0.94
C2. Dependent Variable: Inpatient Visits												
N= 29,161	Year-end price	.	0.00	.	0.00	0.00	0.00	0.00	-0.19	-0.49	-0.14	-0.07
	lower bound	.	0.00	.	0.00	0.00	0.00	0.00	-0.38	-0.60	-0.24	-0.12
	upper bound	.	0.00	.	0.00	0.00	0.00	0.00	0.00	-0.37	-0.04	-0.03
	Elasticity	.	0.00	.	0.00	-0.26	-0.50	0.00	-0.75	-0.77	-0.01	-1.50
	lower bound	.	0.00	.	0.00	-0.52	-0.50	0.00	-1.50	-1.50	-0.01	-1.50
	upper bound	.	0.00	.	0.00	0.00	-0.50	0.00	0.00	-0.03	0.00	-1.49

elasticity of expenditure on medical care based on the experimentally randomized prices:

> "In order to compare our results with those in the literature, however, we must extrapolate to another part of the response surface, namely, the response to coinsurance variation when there is no maximum dollar expenditure. Although any such extrapolation is hazardous (and of little practical relevance given the considerable departure from optimality of such an insurance policy), we have undertaken such an extrapolation rather than forego entirely any comparison with the literature." (Manning et al., 1987, p. 267)

Manning et al. (1987) cited three sources of estimates of the price elasticity of expenditure on medical care in the Rand data, the most prominent of which was based on a simulation by Keeler and Rolph (1988) and not on the Manning et al. (1987) four-part model. Keeler and Rolph (1988) recognized that a comparison of year-end expenditures based on the experimentally induced coinsurance rates across plans could be misleading because behavior was influenced by stoplosses. They therefore used the experimental data to simulate year-end-expenditures in hypothetical plans without stoplosses, and they based their elasticity estimates on this simulated behavior. To conduct the simulation, they assumed myopic responses to marginal price and examined the frequency of visits for all participants in the period for which their families still had over $400 remaining before meeting the stoploss. Notably, they included people in families that far exceeded the stoploss in the simulation. Based on calibrated parametric assumptions on the frequency of visits by type and the cost per visit by type, they forecasted year-end expenditures, and they compared forecasted expenditures across coinsurance plans relative to the free plan to attain their elasticity estimates using the following midpoint arc elasticity formula:

$$\eta = \frac{(Y_a - Y_b)/(Y_a + Y_b)}{(a - b)/(a + b)},$$

where p denotes the coinsurance rate and Y denotes simulated expenditures. The often-cited Rand elasticity estimate of -0.22 comes from a comparison of predicted expenditures across plans with 95% and 25% coinsurance rates with $a = 25$, $b = 95$ $Y_a = 71$ and $Y_a = 55$. Similar calculations based on the predictions from the four-part model and the experimental means yield estimates of -0.14 and -0.17, respectively. The 95% to 25% price change that forms the basis for this arc elasticity should be roughly comparable to the price change on which I base my arc elasticities - from 100% before the deductible to the 20% coinsurance rate. One key methodological difference, however, is that I use within-plan price variation instead of across-plan price variation.

A key difference between the Rand methodology and my methodology comes from the underlying treatment of myopia vs. foresight. While I assume forward looking behavior, the Keeler and Rolph (1988) methodology assumes complete myopia. Recent research by Aron-Dine et al. (2012) examines myopic vs. forward-looking responses to price and rejects the null of myopic behavior. If consumers are forward-looking, it is problematic to assume that the initial statutory marginal price ever governs behavior of participants who expect to meet the stoploss, even in the period before the stoploss is met. Including these participants in the elasticity calculation should bias estimates of price responsiveness downward because variation across plans will be less pronounced among participants who expect to meet the stoploss and thus do not respond at all to the statutory marginal price. Furthermore, participants with the highest coinsurance rates are more likely than participants with the lowest coinsurance rates to meet the stoploss, and thus they are more likely to behave as if medical care is free, which further attenuates elasticity estimates toward zero. The relative treatment of myopia and foresight is one potential explanation for the difference between my estimates and the Rand estimates.

References

Allen, R. G. D. and A. P. Lerner (1934). The concept of arc elasticity of demand. *The Review of Economic Studies 1*(3), pp. 226–230.

Aron-Dine, A., L. Einav, A. Finkelstein, and M. R. Cullen (2012, February). Moral hazard in health insurance: How important is forward looking behavior? Working Paper 17802, National Bureau of Economic Research.

Cabral, M. (2011). Claim timing and ex post insurance selection: Evidence from dental "insurance".

Card, D., C. Dobkin, and N. Maestas (2009). Does medicare save lives? *The Quarterly Journal of Economics 124*(2), 597–636.

Chernozhukov, V., I. Fernández-Val, and A. E. Kowalski (forthcoming 2014). Quantile regression with censoring and endogeneity. *The Journal of Econometrics*.

Chetty, R. (2012). Bounds on elasticities with optimization frictions: A synthesis of micro and macro evidence on labor supply. *Econometrica 80*(3), 969–1018.

Currie, J. and J. Gruber (1996a, May). Health Insurance Eligibility, Utilization of Medical Care, and Child Health. *The Quarterly Journal of Economics 111*(2), 431–66.

Currie, J. and J. Gruber (1996b, December). Saving Babies: The Efficacy and Cost of Recent Changes in the Medicaid Eligibility of Pregnant Women. *Journal of Political Economy 104*(6), 1263–96.

Duan, N. (1983). Smearing estimate: A nonparametric retransformation method. *Journal of the American Statistical Association 78*(383), pp. 605–610.

Eichner, M. J. (1997). *Medical Expenditures and Major Risk Health Insurance*. Ph. D. thesis, Massachusetts Institute of Technology.

Eichner, M. J. (1998). The demand for medical care: What people pay does matter. *The American Economic Review 88*(2), pp. 117–121.

Ellis, R. P. (1986). Rational behavior in the presence of coverage ceilings and deductibles. *The RAND Journal of Economics 17*(2), pp. 158–175.

Gallego-Diaz, J. (1944). A note on the arc elasticity of demand. *The Review of Economic Studies 12*(2), 114–115.

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica 46*(6), pp. 1251–1271.

Hausman, J. A. (1985). The econometrics of nonlinear budget sets. *Econometrica 53*(6), pp. 1255–1282.

Keeler, E. B. and J. E. Rolph (1988). The demand for episodes of treatment in the health insurance experiment. *Journal of Health Economics 7*(4), 337–367.

Manning, W. G., J. P. Newhouse, N. Duan, E. B. Keeler, and A. Leibowitz (1987). Health insurance and the demand for medical care: Evidence from a randomized experiment. *The American Economic Review 77*(3), pp. 251–277.

Medical Expenditure Panel Survey (2004). Agency for Healthcare Research and Quality.

Newey, W. K. (1987). Efficient estimation of limited dependent variable models with endogenous explanatory variables. *Journal of Econometrics 36*(3), 231–250.

Tobin, J. (1958). Estimation of relationships for limited dependent variables. *Econometrica 26*(1), pp. 24–36.