NBER WORKING PAPER SERIES

CENSORED QUANTILE INSTRUMENTAL VARIABLE ESTIMATES OF THE PRICE ELASTICITY OF EXPENDITURE ON MEDICAL CARE

Amanda E. Kowalski

Working Paper 15085
http://www.nber.org/papers/w15085

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
June 2009

I am grateful to Amy Finkelstein, Jon Gruber, and Jerry Hausman for their guidance. I would like to thank the following individuals for thoughtful comments: Michael Anderson, Joshua Angrist, David Autor, John Beshears, Soren Blomquist, Tonja Bowen-Bishop, David Card, Amitabh Chandra, David Cutler, Ian Dew-Becker, Deepa Dhume, Peter Diamond, Joseph Doyle, Esther Duflo, Jesse Edgerton, Matthew Eichner, Brigham Frandsen, John Friedman, Michael Greenstone, Raymond Guiteras, Matthew Harding, Naomi Hausman, Panle Jia, Lisa Kahn, Jonathan Kolstad, Whitney Newey, Joseph Newhouse, Douglas Norton, Christopher Nosko, Matthew Notowidigdo, James Poterba, David Powell, Andrew Samwick, David Seif, Hui Shan, Mark Showalter, Erin Strumpf, and Heidi Williams. Seminar participants at Chicago Booth, Cornell, Duke, Kellogg, MIT, Notre Dame, RAND, Stanford SIEPR, Temple Fox, Texas A&M, Wharton, Wisconsin, Yale, and Yale SOM provided helpful feedback. I gratefully acknowledge National Institute on Aging, Grant Number T32-AG00186. The views expressed herein are those of the author(s) and do not necessarily reflect the views of the National Bureau of Economic Research.

Censored Quantile Instrumental Variable Estimates of the Price Elasticity of Expenditure on Medical Care
Amanda E. Kowalski
NBER Working Paper No. 15085
June 2009
JEL No. I1

ABSTRACT

The extent to which consumers respond to marginal prices for medical care is important for policy. Using recent data and a new censored quantile instrumental variable (CQIV) estimator, I estimate the price elasticity of expenditure on medical care. The CQIV estimator allows the estimates to vary across the skewed expenditure distribution, it allows for censoring at zero expenditure nonparametrically, and it allows for the insurance-induced endogenous relationship between price and expenditure. For identification, I rely on cost sharing provisions that generate marginal price differences between individuals who have injured family members and individuals who do not. I estimate the price elasticity of expenditure on medical care to be stable at -2.3 across the .65 to .95 conditional quantiles of the expenditure distribution. These quantile estimates are an order of magnitude larger than previous mean estimates. I consider several explanations for why price responsiveness is larger than previous estimates would suggest.

Amanda E. Kowalski
National Bureau of Economic Research
1050 Massachusetts Avenue
Cambridge, MA 02138
and NBER
kowalski@nber.org

1 Introduction

The most recent wave of cost control initiatives in medical care depends on consumer responsiveness to price. The Medicare Modernization Act of 2003 included provisions to encourage price responsiveness by establishing tax-advantaged health savings accounts as an incentive for individuals who enroll in high deductible health insurance plans. Relative to traditional plans, high deductible health insurance plans require consumers to face a higher marginal price for each dollar of care that they receive. However, the effects of consumer prices on medical care utilization are not well understood.

Researchers have studied the price elasticity of expenditure on medical care for decades, but three limitations persist: a lack of estimates that allow the price elasticity to vary across the distribution of expenditure, a difficulty in handling censoring of expenditures at zero, and a need for identification strategies to overcome the insurance-induced endogenous relationship between expenditure and price. Estimates based on the RAND Health Insurance Experiment of the 1970's, still widely considered to be the standard in the literature, address the identification issue by randomizing consumers into health insurance plans with varying generosities. Although the RAND estimates address censoring using traditional methods, there is a large and enduring controversy over the appropriateness of the parametric assumptions that these methods require. (See Newhouse et al. (1980), Duan et al. (1983), Mullahy (1998), and Buntin and Zaslavsky (2004).) Perhaps even more important than censoring are the issues that arise because medical spending is so skewed. To my knowledge, none of the existing literature allows for heterogeneity in the price elasticity of expenditure across the expenditure distribution. In my estimation sample, drawn from a large recent data set of employer-sponsored health insurance claims, just 25% of individuals account for 94.5% of expenditures. It seems reasonable, then, that individuals with drastically different levels of expenditure could respond differently to price changes.

In this paper, I produce new estimates of the price elasticity of expenditure on medical care that address heterogeneity across the expenditure distribution, censoring, and identification. I use a new censored quantile instrumental variable (CQIV) estimator, developed specifically for this application by Chernozhukov, Fernandez-Val, and Kowalski (2008). The CQIV estimator is particularly well-suited to address the limitations of the literature. First, the CQIV estimator allows me to obtain estimates of the price elasticity of expenditure on medical care that vary across the expenditure distribution. Relative to mean estimators, quantile estimators such as CQIV are more robust to values in the tails of the distribution, which is particularly advantageous given the skewness in the distribution of medical expenditures.

Second, the CQIV estimator allows me to handle censoring without any distributional assumptions. Econometrically, expenditures are censored at zero since they cannot be negative. In my estimation sample, approximately 40% of individuals consume zero medical care each year, making censoring an important econometric issue. The parametric assumptions required by traditional censored mean estimators could affect the estimates in ways that are not straightforward. In contrast, the CQIV estimator handles censoring nonparametrically in the tradition of Powell (1986).

Third, the CQIV estimator allows me to address endogeneity with an instrumental variable identification strategy. In traditional health insurance policies, the price of an additional dollar of care is a function of expenditure. Thus, observed relationships between price and expenditure will be biased if they do not account for this endogeneity.

The intuition behind my instrumental variable identification strategy is that because of the cost-sharing provisions that govern family health insurance policies, some individuals face lower prices for their own medical care when a family member gets injured. This identification strategy builds on that of Eichner (1997, 1998). As formalized below, the maintained assumption is that one family member's injury can only affect another family member's expenditure through its effect on his marginal price. Although this assumption cannot be tested directly, I take several steps to increase its plausibility. Furthermore, in an indirect test, I find strong evidence in favor of the identification assumption: in families for which cost sharing interactions *cannot* occur, one family member's injury does *not* appear to be related to another family member's medical expenditure.

My main results show that the price elasticity of expenditure on medical care is -2.3 across the .65 to .95 quantiles of the expenditure distribution, with a point-wise 95% confidence interval at the .80 quantile of -2.5 to -2.0. Although I allow the price elasticity estimate to vary with expenditure, I find a fairly stable elasticity across the estimated quantiles. This estimate is an order of magnitude larger than the RAND estimate of the mean elasticity of -0.2. Quantile estimates are not directly comparable to mean estimates, but I consider several pieces of evidence that suggest that the price elasticity of expenditure on medical care is larger than previous estimates would suggest, particularly across the upper quantiles of the expenditure distribution. Notably, the underlying variation that I use for identification is so pronounced that I can illustrate it in simple figures. Furthermore, estimates based on traditional estimators in my data are also much larger than those in the literature. I examine several sources of heterogeneous treatment effects, but in each setting, the variation in the estimates is small relative to the magnitude of the main estimates.

This paper proceeds as follows. In Section 2, I provide background information on the cost sharing provisions of traditional health insurance plans and formalize my identification strategy. In Section 3, I describe the data. In Section 4, I present results based on CQIV and other estimators. In Section 5, I two present robustness tests that use additional data to supplement the main estimation sample. In Section 6, I examine sources of heterogeneity in the main estimates. I conclude and discuss directions for future research in Section 7.

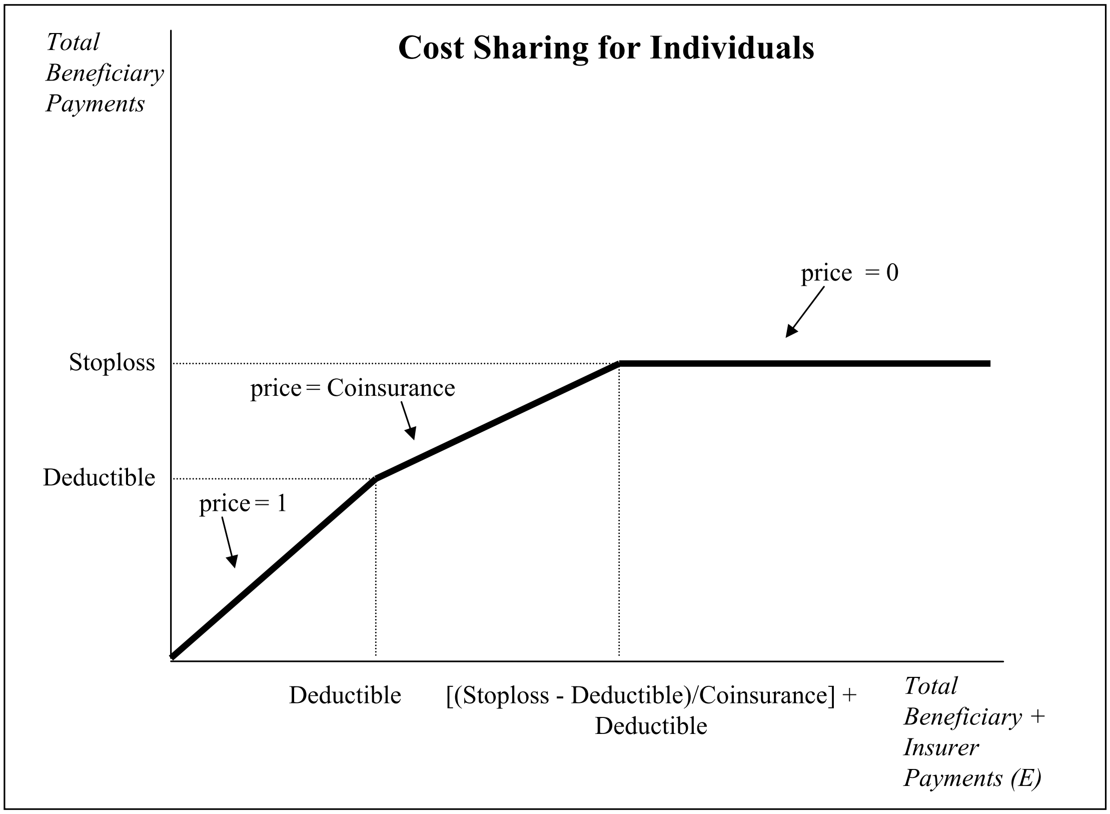
2 Background

2.1 Marginal Pricing for Medical Care

Traditional employer-sponsored health insurance plans have three major cost sharing parameters: a deductible, a coinsurance rate, and a stoploss. The "deductible" is the amount that the consumer must pay before the insurer makes any payments. Before reaching the deductible, the consumer pays one dollar for one dollar of care, so the marginal price is one. After meeting the deductible, the insurer pays a fractional amount for each dollar of care, and the consumer pays the rest. The marginal price that the consumer pays is known as the "coinsurance rate." After the consumer has paid the deductible and a fixed amount in coinsurance, the consumer reaches the "stoploss," and the insurer pays all expenses. For consumers that have met the stoploss, the marginal price is zero. Figure 1 depicts how the deductible, coinsurance rate, and stoploss induce a nonlinear relationship between the total amount paid by the consumer and the total amount paid by the consumer plus the insurer. The consumer faces three distinct marginal prices, the slope of each segment. The intercepts on each axis are exact for a consumer insured as an individual with no family members, but they can move toward the origin for a consumer insured as part of a family.

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. As a concrete example, suppose that a plan has an individual deductible of $500, and it also has a family deductible that is three times the individual deductible ($1,500). 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

Figure 1: Cost Sharing for Individuals



family members must meet the individual deductible. In a family of four, when the first, second, and third family members go to the doctor, they each face the individual deductible of \$500, and then they pay according to the coinsurance rate, 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. In families with more than four members, the family deductible is fixed at \$1,500, and it can be met by any combination of payments toward individual \$500 deductibles. A similar interaction occurs at the level of the stoploss. Given the family-level cost sharing parameters, some individuals will face lower marginal prices than their own medical spending would dictate.

The marginal price variation induced by the family cost sharing parameters suggests a simple way to study price responsiveness: compare expenditures of individuals whose families have and have not met the family deductible. The flaw with this simple identification strategy is that individuals in families that have met the family deductible may be more likely to consume medical care for reasons unrelated to its price, such as contagious illnesses or hereditary diseases. For this reason, instead of comparing individuals according to whether or not their family members have met the family deductible, I compare individuals according to an instrumental variable.

2.2 Identification Strategy

To identify the effect of marginal price on an individual's medical care expenditure, I use an instrumental variable – whether or not a family member has an injury. The first stage effect of a family member's injury on the individual's marginal price is possible in families of four or more because of the family deductible and family stoploss described above. When one family member receives treatment for an injury, the family is more likely to meet the family deductible than it otherwise would have been, and any individual in the family is more likely to face a lower marginal price than his own spending would dictate. Empirically, I find that one family member's injury does indeed affect another family member's marginal price.

Given the first stage, the key to the identification strategy is an exclusion restriction: one family member's injury cannot affect another family member's medical spending outside of its effect on his marginal price. Strictly speaking, direct violations of the exclusion restriction are not possible. Since the outcome that I study is the medical spending of an individual in a family, and not the medical spending of the entire family, expenditure for the treatment of one family member's injury is not included in the outcome variable. Furthermore, since one family member's injury does have a direct effect on his own medical expenditure, and the injury itself likely influences his decision to consume follow-up medical care and care for secondary illnesses, I use injured family members only to construct the instrument, and I do not include them in the estimation sample. If two or more family members are injured, all injured family members are excluded from the estimation sample. As discussed in Section 5, family injuries have limited persistence across years, so sample selection issues due to the exclusion of injured parties should not be a cause for concern.

Other potential violations of the exclusion restriction involve indirect effects of one family member's injury on another family member's medical spending that occur through a mechanism other than the marginal price. I include only specific injury categories in the determination of the instrument to preclude any mechanisms that involve physical contagion. The complete set of injury categories included in the determination of the instrument are intracranial injuries, superficial injuries (injuries to the skin), crushing injuries, foreign body injuries, burns, and complications of trauma and injuries to the nerves and spinal cord. These injury categories should be severe and unexpected enough that treatment for an injury in these categories should not be related to an underlying family-level propensity to seek treatment, which could lead to a violation of the exclusion restriction. Indirect tests for violations of the exclusion restriction based on injuries in families with no possible cost-sharing interactions, presented in Section 5, lend support to my identification strategy.

To further avoid violations of the exclusion restriction, and also to avoid measurement error, I determine the instrument only on the basis of whether an individual was treated for an injury, and not on the basis of the spending associated with the treatment. If the instrument included a measure of injury spending, the instrument could be related to another family member's medical spending through a family-level propensity to go to expensive doctors, thus violating the exclusion restriction. Since my instrument is only based on the treatment margin, a family-level propensity to go to expensive doctors will not violate the exclusion restriction. However, such a propensity could raise concerns if the price elasticity of expenditure on medical care is not homogenous in the population.

In any instrumental variable setting, if the treatment effect of interest is not homogenous in the population, the estimated effect is a "local average treatment effect," which is the average effect on "compliers" who would not have received the treatment absent the intervention of the instrument. In this setting, compliers are people who have a family injury which causes them to face a lower price than they would have absent the injury. Although it is not possible to identify compliers because doing so would involve the observation of a counterfactual state in which a family member did not get injured, Angrist, Imbens, and Rubin (1996) propose a formal methodology to examine the average characteristics of compliers in a setting with a binary treatment and a binary instrument. The multivalued treatment in my application precludes the use of the Angrist et al. (1996) methodology, but I can still informally describe the compliers as the population for which the first stage is likely to be the strongest. For example, the first stage will likely be strongest among people who go to more expensive doctors, because the higher the expense, the higher the likelihood of meeting the family deductible. In addition, the first stage will likely be strongest among accident-prone families, because having an injury in the family is a necessary prerequisite to being a complier. The first stage is also likely to be strongest among large families because large families have more people to contribute to the fixed family deductible.

Furthermore, the first stage is likely to be strongest among individuals that have a family injury that occurs early in the year. One limitation of my approach is that it is static in the sense that I do not explicitly model intra-year timing issues. However, the instrumental variables framework automatically allows for some dynamics. Consider the case of an injury that occurs on December 31. Such an injury many initially seem problematic for my strategy because it is unlikely that family expenditure can respond to such an injury before the end of the year. However, it is also unlikely that the family member's marginal price can respond to such an injury before the end of the year. Thus, the minimal expenditure effect will be scaled by the minimal price effect, therefore taking the intra-year timing of the injury into account. Stated more explicitly, my approach only requires that there is time for the expenditure to respond if there time for marginal price to respond.

3 Data

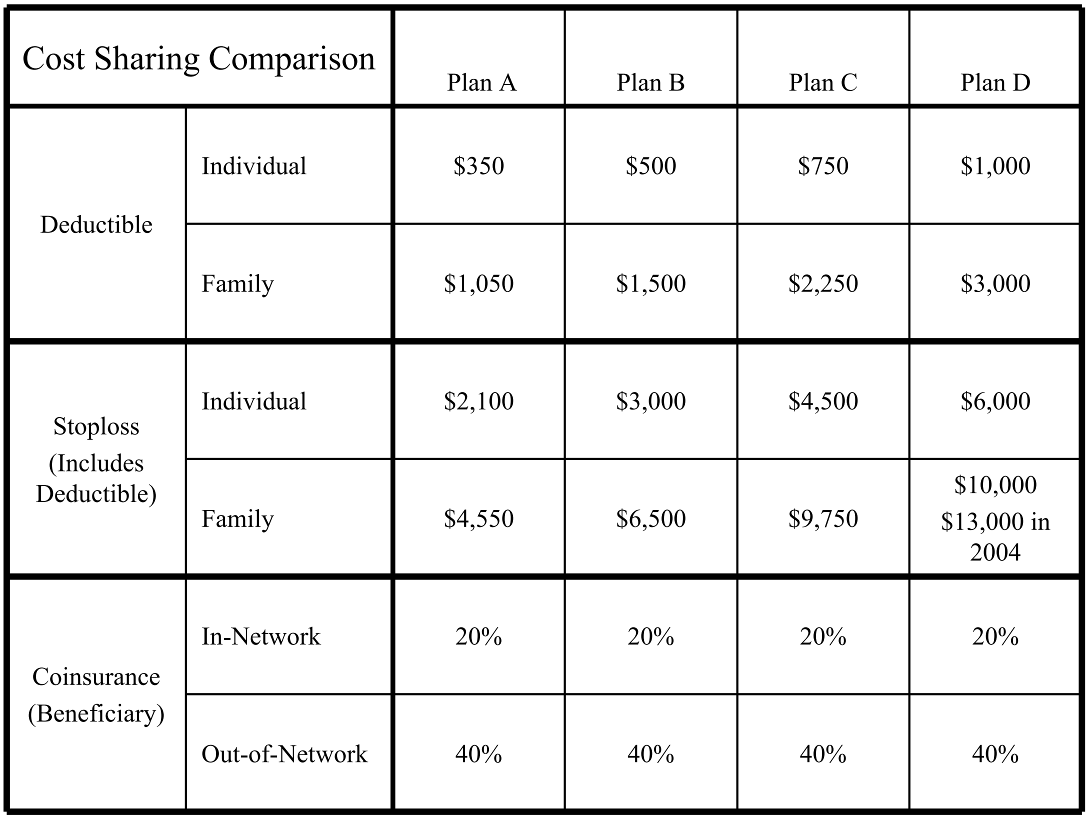
3.1 Data Description

I use recent proprietary data from a US firm with over 500,000 insured employees. The data for my analysis are merged together from several databases compiled and distributed by Medstat (2003). In my merged data set, in addition to observing inpatient, outpatient, and prescription drug claims, I also observe characteristics of the offered plans and associated enrollment characteristics. The Medstat claims data are particularly well-suited to my analysis because the medical claims data identify the beneficiary and insurer contributions on each claim. Because beneficiaries must submit claims to receive reimbursement, and because the firms that pay the claims collect the data, incentives are aligned to ensure the accuracy and completeness of the claims data.

A major advantage of the Medstat data over stand alone claims data is that if beneficiaries do not file any claims or discontinue enrollment, I can still verify their coverage and observe their demographic characteristics in the enrollment database. These data represent an advantage over Eichner (1997, 1998). Although I predominantly use cross-sectional variation in the data, I can track individuals and their covered family members over time as long as the subscriber remains at the same firm. One limitation of the Medstat data is that I do not observe employees or family members who are not covered, and I do not observe health insurance options available outside the firm. However, according to the 2006 Kaiser Annual Survey of Employer Health Benefits, 82% of eligible workers enroll in plans offered by their employers, so I should observe a large majority of workers at the firm that I study.

I focus on data from one firm to isolate marginal price variation from other factors that could vary by firm and plan. This firm is in the retail trade industry. The main advantage of the firm I that I study is that the four plans that it offered in 2003 and 2004 varied only in the deductible and stoploss. Furthermore, one of the offered plans has a $1,000 deductible, which is coincidentally the initial qualifying amount for a plan to be considered "high deductible" by 2003 legislation. Plan selection issues should not invalidate my identification strategy because it relies on within-plan price variation. However, plan-related local average treatment effects are possible, and I investigate them by comparing behavior across plans.

Table 1: Cost Sharing Comparison



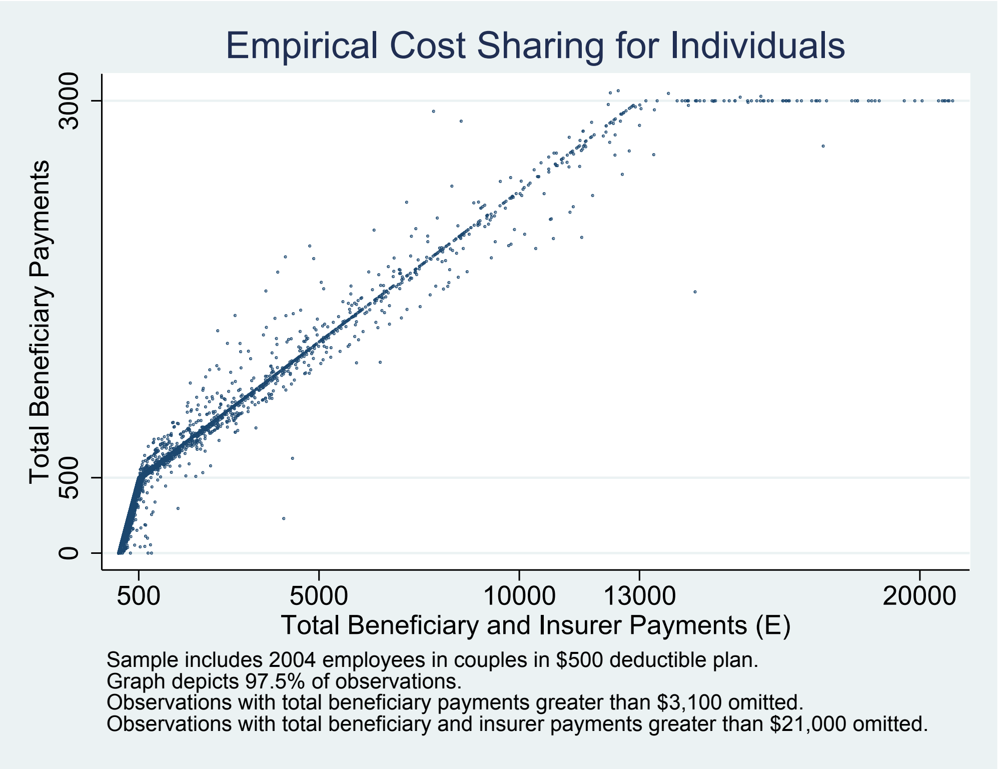
Cost Sharing Comparison		Plan A	Plan B	Plan C	Plan D
Deductible	Individual	$350	$500	$750	$1,000
	Family	$1,050	$1,500	$2,250	$3,000
Stoploss (Includes Deductible)	Individual	$2,100	$3,000	$4,500	$6,000
	Family	$4,550	$6,500	$9,750	$10,000 $13,000 in 2004
Coinsurance (Beneficiary)	In-Network	20%	20%	20%	20%
	Out-of-Network	40%	40%	40%	40%

Table 1 presents a comparison of the cost sharing parameters across plans. The individual deductibles vary from $350 to $1,000, and the family deductible is always three times the individual deductible, as in the example described above. Net of deductibles, the family stoplosses are always twice as large as the individual stoplosses.

The simple cost sharing parameters introduced above provide a very accurate description of the marginal prices that consumers face at this firm. Almost all covered medical spending counts toward the deductible and stoploss, except for spending on prescription drugs, which I do not include in my analysis because it is covered separately. Unlike in many medical plans, there is no fixed per-visit payment.

The only complication in the cost sharing structure at the firm that I study is that the plans offer incentives for beneficiaries to go to providers that are part of a network. All four plans are preferred provider organization (PPO) plans. According to the Kaiser 2006 Annual Survey of Employer Health Benefits, 60% of workers with employer-sponsored health insurance are covered by PPO plans. PPO plans do not require a primary care physician or a referral for services, and there are no capitated physician reimbursements. However, there is an incentive to visit providers in the network because there is a higher coinsurance rate for expenses outside of the network. In the firm that I study, the general coinsurance rate is 20%, and the out-of-network coinsurance rate is 40%. The network itself does not vary across plans. In the data, there are no identifiers for out-of-network expenses, but, as demonstrated by Figure 2, which plots beneficiary expenses on total expenses for a sample of individuals, beneficiary expenses follow the in-network schedule with a high degree of accuracy, indicating that out-of-network expenses are very rare. Accordingly, in my analysis,

Figure 2: Empirical Cost Sharing for Individuals



Sample includes 2004 employees in couples in $500 deductible plan.
Graph depicts 97.5% of observations.
Observations with total beneficiary payments greater than $3,100 omitted.
Observations with total beneficiary and insurer payments greater than $21,000 omitted.

I assume that everyone who has met the deductible faces the in-network marginal price for care. My main results do not change when I exclude the small number of beneficiaries whose out-of-pocket payments deviate from the in-network schedule.

3.2 Sample Selection

Although selection into the firm that I study could be a cause for concern, 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. Since these data are so recent, they should provide an accurate description of current health insurance offerings and usage. Because the covered population consists of active, non-union employees in the retail trade industry, my findings should have widespread external validity.

Within the firm, the main selection criterion that I apply is a continuous enrollment restriction. 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 sample if their entire families, with the exception of newborns, are enrolled for the entire plan year. I retain families with newborns on the grounds that child birth is an important medical expense. Care before death is also an important medical expense, but I cannot make an exception for individuals who die because I only observe in-hospital deaths, and there are none recorded in the unselected sample. 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.

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. 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. Taken together, these steps eliminate less than seven percent of individuals from the continuously enrolled sample.

In this clean sample, just over 25% of employees with other insured family members are insured in families of four or more. The 2004 main estimation sample includes 127,119 individuals from 29,010 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. In a robustness test, I examine employee-spouse couples precisely because price interactions are not possible.

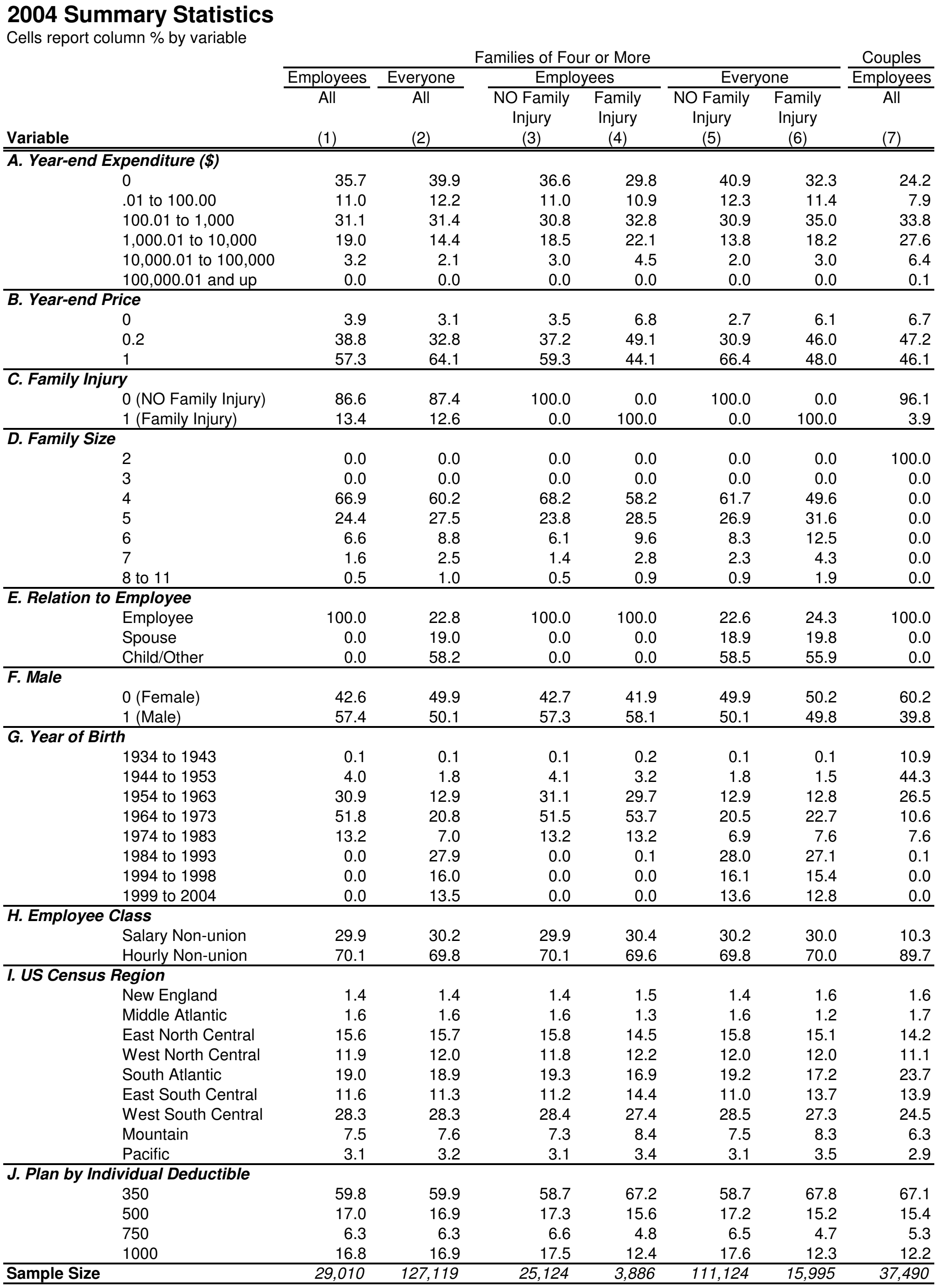
To better control for unobservables, I limit my estimation sample to the employee in each family, and I use other family members only in the determination of the instrument. In some specifications, I also include individuals identified as spouses in the estimation sample. Restricting the sample to employees or employees and spouses sacrifices power because it does not take the price responsiveness of all family members into account, but it arguably 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 possible correlations in price responsiveness among family members.

3.3 Summary Statistics

In the 2004 sample, mean year-end medical expenditure by the beneficiary and the insurer is \$1,485 in the sample of employees and \$1,135 in the sample that also includes spouses and dependents. However, the mean is not a very informative summary statistic for medical expenditures because many people consume zero care, and the distribution of medical spending among those who do consume care traditionally has a long right tail. 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 94.5% of expenditures. Given this skewness, I analyze the logarithm of expenditure instead of the level.

The first panel of Table 2 summarizes the expenditure distribution across bins that follow a logarithmic scale. As shown in the first column, excluding individuals with zero expenditure, the distribution of positive expenditure among employees follows an approximately lognormal distribution, with 31.1% of individuals in the expenditure range between \$100 and \$1,000, and smaller percentages of individuals

Table 2: 2004 Summary Statistics



2004 Summary Statistics

Cells report column % by variable

	Families of Four or More						Couples
	Employees	Everyone	Employees		Everyone		Employees
	All	All	NO Family Injury	Family Injury	NO Family Injury	Family Injury	All
Variable	(1)	(2)	(3)	(4)	(5)	(6)	(7)
A. Year-end Expenditure ($)							
0	35.7	39.9	36.6	29.8	40.9	32.3	24.2
.01 to 100.00	11.0	12.2	11.0	10.9	12.3	11.4	7.9
100.01 to 1,000	31.1	31.4	30.8	32.8	30.9	35.0	33.8
1,000.01 to 10,000	19.0	14.4	18.5	22.1	13.8	18.2	27.6
10,000.01 to 100,000	3.2	2.1	3.0	4.5	2.0	3.0	6.4
100,000.01 and up	0.0	0.0	0.0	0.0	0.0	0.0	0.1
B. Year-end Price							
0	3.9	3.1	3.5	6.8	2.7	6.1	6.7
0.2	38.8	32.8	37.2	49.1	30.9	46.0	47.2
1	57.3	64.1	59.3	44.1	66.4	48.0	46.1
C. Family Injury							
0 (NO Family Injury)	86.6	87.4	100.0	0.0	100.0	0.0	96.1
1 (Family Injury)	13.4	12.6	0.0	100.0	0.0	100.0	3.9
D. Family Size							
2	0.0	0.0	0.0	0.0	0.0	0.0	100.0
3	0.0	0.0	0.0	0.0	0.0	0.0	0.0
4	66.9	60.2	68.2	58.2	61.7	49.6	0.0
5	24.4	27.5	23.8	28.5	26.9	31.6	0.0
6	6.6	8.8	6.1	9.6	8.3	12.5	0.0
7	1.6	2.5	1.4	2.8	2.3	4.3	0.0
8 to 11	0.5	1.0	0.5	0.9	0.9	1.9	0.0
E. Relation to Employee							
Employee	100.0	22.8	100.0	100.0	22.6	24.3	100.0
Spouse	0.0	19.0	0.0	0.0	18.9	19.8	0.0
Child/Other	0.0	58.2	0.0	0.0	58.5	55.9	0.0
F. Male							
0 (Female)	42.6	49.9	42.7	41.9	49.9	50.2	60.2
1 (Male)	57.4	50.1	57.3	58.1	50.1	49.8	39.8
G. Year of Birth							
1934 to 1943	0.1	0.1	0.1	0.2	0.1	0.1	10.9
1944 to 1953	4.0	1.8	4.1	3.2	1.8	1.5	44.3
1954 to 1963	30.9	12.9	31.1	29.7	12.9	12.8	26.5
1964 to 1973	51.8	20.8	51.5	53.7	20.5	22.7	10.6
1974 to 1983	13.2	7.0	13.2	13.2	6.9	7.6	7.6
1984 to 1993	0.0	27.9	0.0	0.1	28.0	27.1	0.1
1994 to 1998	0.0	16.0	0.0	0.0	16.1	15.4	0.0
1999 to 2004	0.0	13.5	0.0	0.0	13.6	12.8	0.0
H. Employee Class							
Salary Non-union	29.9	30.2	29.9	30.4	30.2	30.0	10.3
Hourly Non-union	70.1	69.8	70.1	69.6	69.8	70.0	89.7
I. US Census Region							
New England	1.4	1.4	1.4	1.5	1.4	1.6	1.6
Middle Atlantic	1.6	1.6	1.6	1.3	1.6	1.2	1.7
East North Central	15.6	15.7	15.8	14.5	15.8	15.1	14.2
West North Central	11.9	12.0	11.8	12.2	12.0	12.0	11.1
South Atlantic	19.0	18.9	19.3	16.9	19.2	17.2	23.7
East South Central	11.6	11.3	11.2	14.4	11.0	13.7	13.9
West South Central	28.3	28.3	28.4	27.4	28.5	27.3	24.5
Mountain	7.5	7.6	7.3	8.4	7.5	8.3	6.3
Pacific	3.1	3.2	3.1	3.4	3.1	3.5	2.9
J. Plan by Individual Deductible							
350	59.8	59.9	58.7	67.2	58.7	67.8	67.1
500	17.0	16.9	17.3	15.6	17.2	15.2	15.4
750	6.3	6.3	6.6	4.8	6.5	4.7	5.3
1000	16.8	16.9	17.5	12.4	17.6	12.3	12.2
Sample Size	*29,010*	*127,119*	*25,124*	*3,886*	*111,124*	*15,995*	*37,490*

in the bins above and below this range. The distribution of expenditures in the full sample, summarized in the second column, is similar. Table A1 presents analogous summary statistics for the 2003 samples. In a previous version of this paper, Kowalski

(2008), I report a comparison of the skewness between my sample and the nationally-representative 2004 Medical Expenditure Panel Survey (MEPS). On a percentage basis, the skewness in my sample is relatively comparable to the skewness in the MEPS, but my sample has a slightly more concentrated right tail.

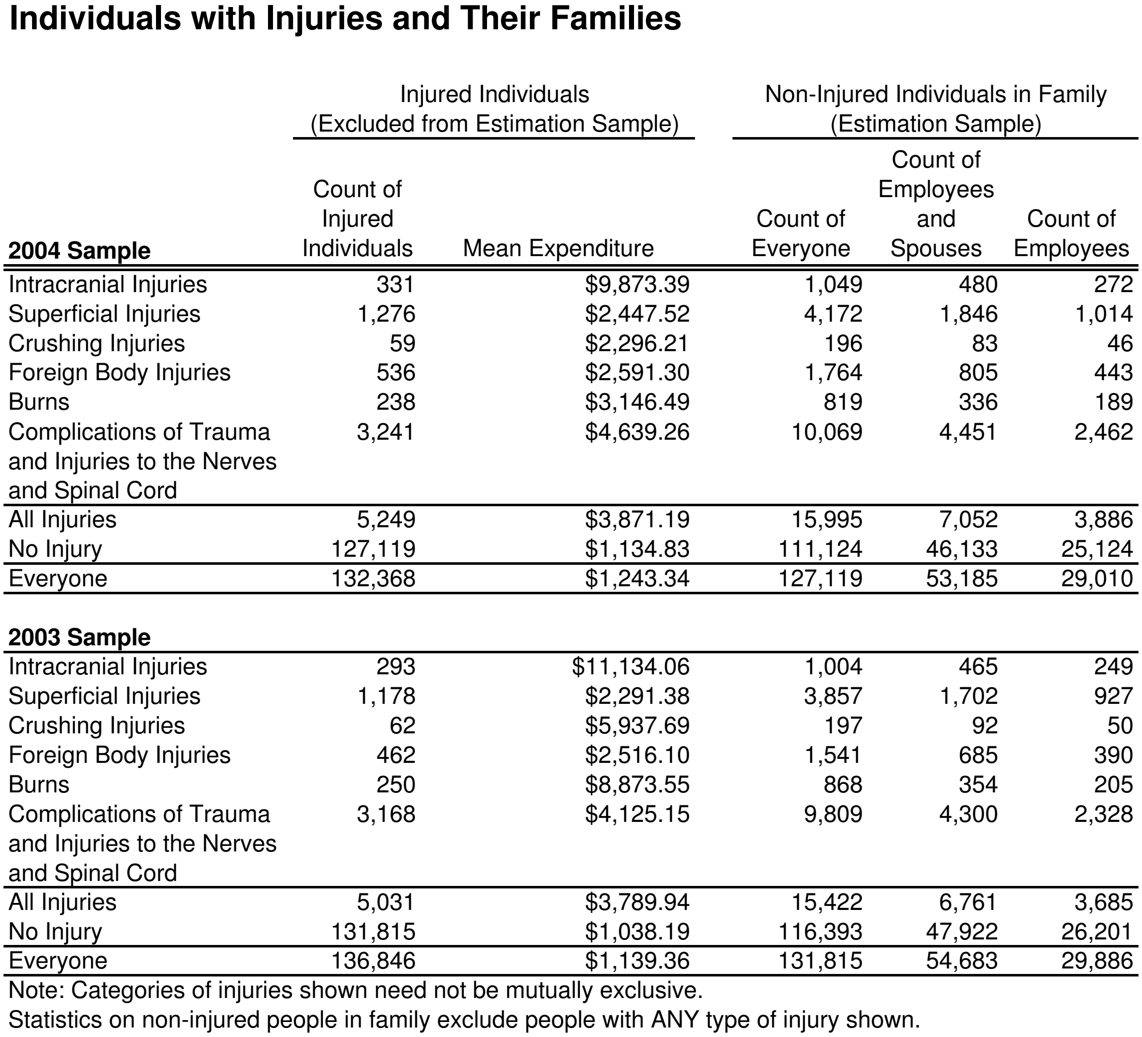
The second panel in Table 2 depicts the distribution of the endogenous variable, the marginal price for the next dollar of care at the end of the year. I calculate the marginal price to reflect the spending of the individual and his family members. If the individual has not consumed any care and the 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, 57.3% of beneficiaries face a marginal price of one, 38.3% of employees face the coinsurance rate of 0.2, and 3.9% 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 13.4% 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, 12.6% of individuals in the full sample have an injury in the family.

Even though injured people are excluded from all estimation samples, I report statistics on the injured people in Table 3. 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 first column. Complications of trauma and injuries to the nerves and spinal cord are the most prominent. The distribution of injuries across 2003 and 2004 is remarkably stable, which could indicate that the firm is large enough that injuries are not idiosyncratic. In the second column, 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. The last three columns of Table 3 show the number of affected family members in each estimation sample by injury category.

Panels D through J of Table 2 summarize the distribution of covariates. Family size varies from four to eleven, with 60.2% of people 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

Table 3: Individuals with Injuries and Their Families



Individuals with Injuries and Their Families

	Injured Individuals (Excluded from Estimation Sample)		Non-Injured Individuals in Family (Estimation Sample)		
2004 Sample	Count of Injured Individuals	Mean Expenditure	Count of Everyone	Count of Employees and Spouses	Count of Employees
Intracranial Injuries	331	$9,873.39	1,049	480	272
Superficial Injuries	1,276	$2,447.52	4,172	1,846	1,014
Crushing Injuries	59	$2,296.21	196	83	46
Foreign Body Injuries	536	$2,591.30	1,764	805	443
Burns	238	$3,146.49	819	336	189
Complications of Trauma and Injuries to the Nerves and Spinal Cord	3,241	$4,639.26	10,069	4,451	2,462
All Injuries	5,249	$3,871.19	15,995	7,052	3,886
No Injury	127,119	$1,134.83	111,124	46,133	25,124
Everyone	132,368	$1,243.34	127,119	53,185	29,010
2003 Sample					
Intracranial Injuries	293	$11,134.06	1,004	465	249
Superficial Injuries	1,178	$2,291.38	3,857	1,702	927
Crushing Injuries	62	$5,937.69	197	92	50
Foreign Body Injuries	462	$2,516.10	1,541	685	390
Burns	250	$8,873.55	868	354	205
Complications of Trauma and Injuries to the Nerves and Spinal Cord	3,168	$4,125.15	9,809	4,300	2,328
All Injuries	5,031	$3,789.94	15,422	6,761	3,685
No Injury	131,815	$1,038.19	116,393	47,922	26,201
Everyone	136,846	$1,139.36	131,815	54,683	29,886

Note: Categories of injuries shown need not be mutually exclusive.
Statistics on non-injured people in family exclude people with ANY type of injury shown.

because the sample includes parents and their children. Panel H shows that 70.1% 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 vs. hourly classification could serve 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.

The final panel 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.

A comparison of the plan distribution between the employee sample and the full sample shows that larger families do not select differentially into plans. Almost 60% of employees and families are enrolled in the most generous plan, which has a $350 deductible. 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. Indeed, I find that the first stage coefficient is the largest in this plan. However, elasticity estimates are very similar in magnitude in separate specifications by plan.

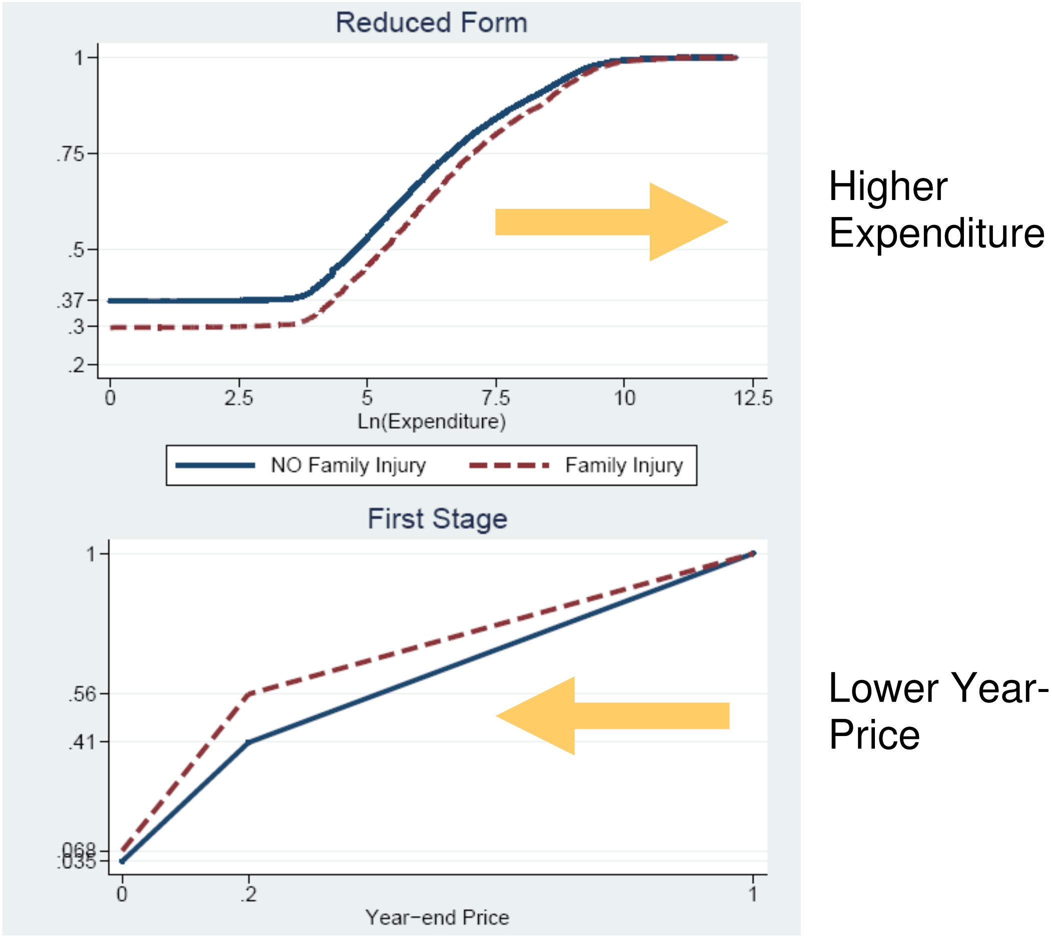
4 Results

4.1 Preliminary Evidence

The raw variation in the data that drives my instrumental variable approach is so pronounced that it can be discerned graphically, without the assistance of complex estimators. In instrumental variable parlance, the effect of family injury on expenditure is the "reduced form," and the effect of family injury on the year-end price is the "first stage." The simple instrumental variable estimate is the ratio of the reduced form to the first stage. To show the variation that drives the instrumental variable strategy, I present graphical depictions of the reduced form and the first stage in the 2004 sample of employees.

To demonstrate the reduced form, in the top panel of Figure 3, I present the cumulative distribution (cdf) of expenditure conditional on family injury. The cdf of expenditure for employees with no family injury is represented by a solid line, and the cdf of expenditure for employees with a family injury is represented by a dashed line. In this depiction, each quantile on the y axis is associated with a value of the logarithm of expenditure on the x axis. Median expenditure is $120 among employees with no family injuries and $203 among employees with family injuries. Since the lines never cross, it is clear from the figure that employees with family injuries have higher expenditures at all quantiles. Similarity in the curvature of the two cdfs provides reassurance that not all individuals with family injuries have extremely high expenditures, thus driving the results. The y intercepts of each line indicate that family injuries affect the extensive margin decision of whether or not to consume any care; only 30% of people with family injuries consume zero care, as opposed to 37% of people with no family injuries. To examine whether the difference between the lines at all quantiles is driven by effects on the extensive margin, I create

Figure 3: Reduced Form and First Stage



a similar figure, not shown here, that depicts cumulative distributions conditional on positive expenditure. The lines of the new figure do not cross, indicating that even among employees with positive expenditure, employees with family injuries have higher expenditure at each quantile. Columns 3-4 of Table 2 present the underlying conditional probability density functions in tabular form.

To demonstrate the first stage effect of family injury on the year-end price, in the bottom panel of Figure 3, I present the cumulative distribution of year-end price conditional on family injury. Since the year-end price takes on only three values, the cdf is a step function, but I connect the points of the step function with straight lines to aid in the visual interpretation. The lines in this figure do not cross, indicating that employees with family injuries are more likely to face lower prices than their counterparts without family injuries. Labels on the y axis show that 56% of employees with family injuries spend more than the deductible, while only 41% of employees without family injuries spend more than the deductible. Similarly, 6.8% of employees with family injuries spend more than the stoploss, while only 3.5% of employees without family injuries spend more than the stoploss.

The depiction in the bottom panel also allows us to assess which price change, the change from 1 to 0.2 or the change from 0.2 to 0, yields the most identification. Following Angrist and Imbens (1995), the vertical difference between the cdf's at the new price is proportional to the weight in an instrumental variable estimate formed from a weighted combination of separate Wald estimates for each price change. Since the difference in the cdf's is largest at the price of 0.2, the figure indicates that most identification comes from the price change between 1 and 0.2, and some identification comes from the price change between 0.2 and 0.

As a more formal alternative to the bottom panel of Figure 3, a simple ordinary least squares (OLS) regression of year-end price on family injury and a set of covariates discussed below indicates that having an injury in the family decreases the year-end price by 11 percentage points, with a standard error of .7 percentage points. The R-squared of this first stage regression with the covariates partialled out is 0.0096, implying a concentration parameter (defined as $NR^2/(1-R^2)$) of 281. Based on this evidence, "weak instruments bias" is unlikely to be a problem in this application.

The inclusion of control variables should not have a substantial impact on the estimate, but should merely make it more precise. One way to assess the importance of control variables to the instrumental variable strategy is to examine the distribution of each variable conditional on the values of the instrument. Ideally, in this setting, individuals who have any injured family member would be similar in all observable ways to those who do not have an injured family member.

Panels D through J of columns 3-4 and 5-6 of Table 2 give the distribution of covariates conditional on family injury. The distribution of family size shows that individuals in larger families are slightly more likely to have injuries in their families, as is to be expected if the incidence of injures is distributed evenly across individuals. Given this discrepancy, I include flexible controls for family structure in my formal estimates. Specifically, I include a dummy for the presence of a spouse on the policy, the year of birth of the oldest and youngest dependent, and the count of family members born in each of the year ranges in the table, with the 1999-2004 range saturated by year. In the remaining panels of Table 2, the distribution of the other control variables appears much less sensitive to the instrument. I control for them in my formal estimates because complex interactions between these variables might not be visible in the table.

4.2 CQIV Model and Estimation

I formalize the above preliminary evidence using the following censored quantile instrumental variable model:

$$
\begin{aligned}
\ln E &= \max(\ln E^*, C) = T((\ln E_i)^*) && \text{(1a)} \\
(\ln E)^* &= Q_{(\ln E)^*}(U|P, W, V) && \text{(1b)} \\
P &= \phi(V, W, Z) && \text{(1c)}
\end{aligned}
$$

where lnE, is the logarithm of observed year-end medical expenditure, and $T(x) \equiv \max(x, C)$ is the transformation function that censors the unobserved uncensored value of $(\ln E_i)^*$ at C, where C is lower than the smallest nonzero value of ln E. P is the year-end marginal price of medical care, W are covariates described above, Z is an indicator for family injury (the instrumental variable), V is a latent unobserved regressor called the "control function," and U is a Skorohod disturbance that satisfies the independence assumption

$$U \backsim U(0,1)|P, W, C, V.$$

This independence assumption is stronger than the mean independence assumption required by models of the conditional mean, but it should be plausible given the discussion in Section 2.2. It reflects the exclusion restriction that one family member's injury cannot affect another family member's expenditure outside of its effect on marginal price. The *quantiles* of any distribution always follow a uniform distribution, so the uniform distribution is completely general and is not a parametric assumption of the model.

For computational efficiency, I estimate a linear model. The functional form of the model that I estimate is very flexible, in that it allows for random coefficients that vary with the quantiles of the expenditure distribution:

$$
\begin{aligned}
(\ln E)^* &= \alpha(U)P + W'\beta(U) + \gamma(U)V \\
&= X'\beta(U), \qquad X = (P, V, W)
\end{aligned}
$$

where $\alpha(U)$ are the random coefficients of interest.

As in traditional models, we can interact the marginal price variable with an observed covariate to examine heterogeneity in price responsiveness along an observed dimension. The quantile model also allows us to examine heterogeneity in price responsiveness along the unobserved dimension U. In what follows, I maintain the agnostic interpretation that the coefficients are a function of the quantiles of unobserved heterogeneity. For stronger interpretations, we can make assumptions about the heterogeneity represented by U. For example, in this application, income is not observed, and if we assume that income is the only dimension of unobserved heterogeneity, the estimated coefficients will allow us to examine price responsiveness at varying quantiles of the income distribution. Alternatively, U could represent the quantiles of unobserved health or hypochondria. If unobserved heterogeneity is one dimensional and the quantiles of unobserved heterogeneity are the same as the quantiles of the expenditure distribution conditional on covariates, the estimated coefficients at the highest quantiles will yield price responsiveness for individuals who spend the most.

I estimate this model using the censored quantile instrumental variable (CQIV) estimator, developed in detail in Chernozkukov, Fernandez-Val, and Kowalski (2008). Here, I provide more intuition for the advantages of the CQIV estimator relative to other models, and I provide practical implementation details. I have already shown that the CQIV model allows the coefficients to vary with the quantiles of interest. In addition, CQIV handles censoring nonparametrically, and it allows for endogeneity.

Censoring induces attenuation bias in quantile regression much in the same way it induces bias in mean regression: when C is observed in the place of a value that should be much smaller, a line that fits the observed values will be biased toward zero. Since quantile regression uses information from the entire sample to generate the estimate at each quantile, if some observations on $\ln E$ are censored, the quantile regression lines can be biased toward zero at *all* quantiles. The Powell (1984) estimator overcomes this difficulty by incorporating censoring directly into the estimator as follows:

$$\widehat{\beta}(\tau) \text{ minimizes } \sum_{i=1}^{n} \rho_\tau(\ln E_i - T(X_i'\beta(\tau))).$$

where $\rho_\tau(u) = \{(1-\tau)1(u<0) + \tau 1(u>0)\}|u|$. Despite its theoretical appeal, this model is rarely used in practice because the function $T(x)$ induces nonconvexities in the objective function that present computational difficulties.

Chernozhukov and Hong (2002) devised a tractable computational censored quantile regression (CQR) algorithm for Powell's estimator based on the idea that Powell's censored regression model estimates the coefficients using observations that are not likely to be censored. The algorithm is a three-step procedure that predicts which observations are least likely to be censored and estimates the coefficients based on those observations. The first step involves a parametric prediction of the probability of censoring based on a probit or logit model. A set fraction of observations that are unlikely to be censored are retained for estimation via quantile regression in the second step. After the second step, a larger set of observations is retained based on the predicted values of the dependent variable. This sample gets asymptotically close to the ideal sample of non-censored observations, and consistent estimates are obtained through a third step of quantile regressionon this sample. The CQIV computational algorithm uses an analog of the Chernozhukov and Hong (2002) algorithm to handle censoring, with an additional pre-step to handle endogeneity.

The CQIV estimator uses a control function approach to handle endogeneity in the tradition of Hausman (1978). The control function approach is based on the observation that endogeneity between the price variable P and the expenditure variable $(\ln E)^*$ results in a lack of orthogonality between P and the structural disturbance U. Given this lack of orthogonality, estimates based only on Equation (1b) would be inconsistent. However, if Z is orthogonal to U conditional on covariates W, then the structural disturbance U is a function of the first stage disturbance V as follows: $E(U|V) = \delta V + \eta$. By construction, $E(\eta|V) = 0$. Therefore, when $\widehat{V}$ is included along with P in Equation (1b), the new structural error term is mean independent of the price variable P. The conditions for strict independence can be derived similarly. The estimated first stage error term $\widehat{V}$ is referred to as the estimated "control function," because it "controls" for endogeneity in the structural equation.

One advantage of the control function approach to endogeneity, in contrast to the moment condition approach to endogeneity used by Chernozhkov and Hansen (2008) in their quantile instrumental variable estimator, is that the control function approach does not require a rank invariance condition on the structural equation. However, one disadvantage is that the assumptions necessary for the control function approach are less likely to be satisfied when the endogenous variable is discrete, as it is in this application. In practice, however, estimates based on a variation on the CQIV estimator that uses a Chernozhukov and Hansen (2008) moment condition approach to endogeneity, reported in Kowalski (2008), are almost identical to those presented here to the number of reported decimal places.

In the reported estimates, I follow the standard practice of obtaining an estimate of the control term by predicting the OLS residuals from the first stage equation. I obtain

95% confidence intervals on the coefficients via bootstrapping. In practice, I report the mean of the confidence interval as the point estimate because the discreteness of the covariates can hinder convergence of the quantile estimator at specific combinations of covariates.

4.3 Main Results

Since so many individuals have zero expenditure, there is not enough empirical expenditure and price variation to obtain precise estimates below the .65 quantile of the expenditure distribution in my data. At conditional quantiles where zero expenditure is likely, the marginal price can have an effect on two margins - the decision to spend anything at all, and the decision to change spending conditional on spending a positive amount. If changes in price and other factors are not sufficient to induce people to visit the doctor at all, it is not possible to estimate the effect of small changes in price. With approximately 40% censoring in the data, it seems reasonable that CQIV coefficients are not reliable at the median. Estimates can be obtained below the .65 quantile, but they are not very precise.

For the .65 quantile and above, Table 4 reports the coefficient on year-end price and the associated lower and upper bounds of the 95% confidence interval. Year-end price is not specified in logarithmic form because it can take on a value of zero. Thus, the estimated coefficient must be transformed into an elasticity estimate. I transform the estimated coefficient into an elasticity using the following arc elasticity formula:

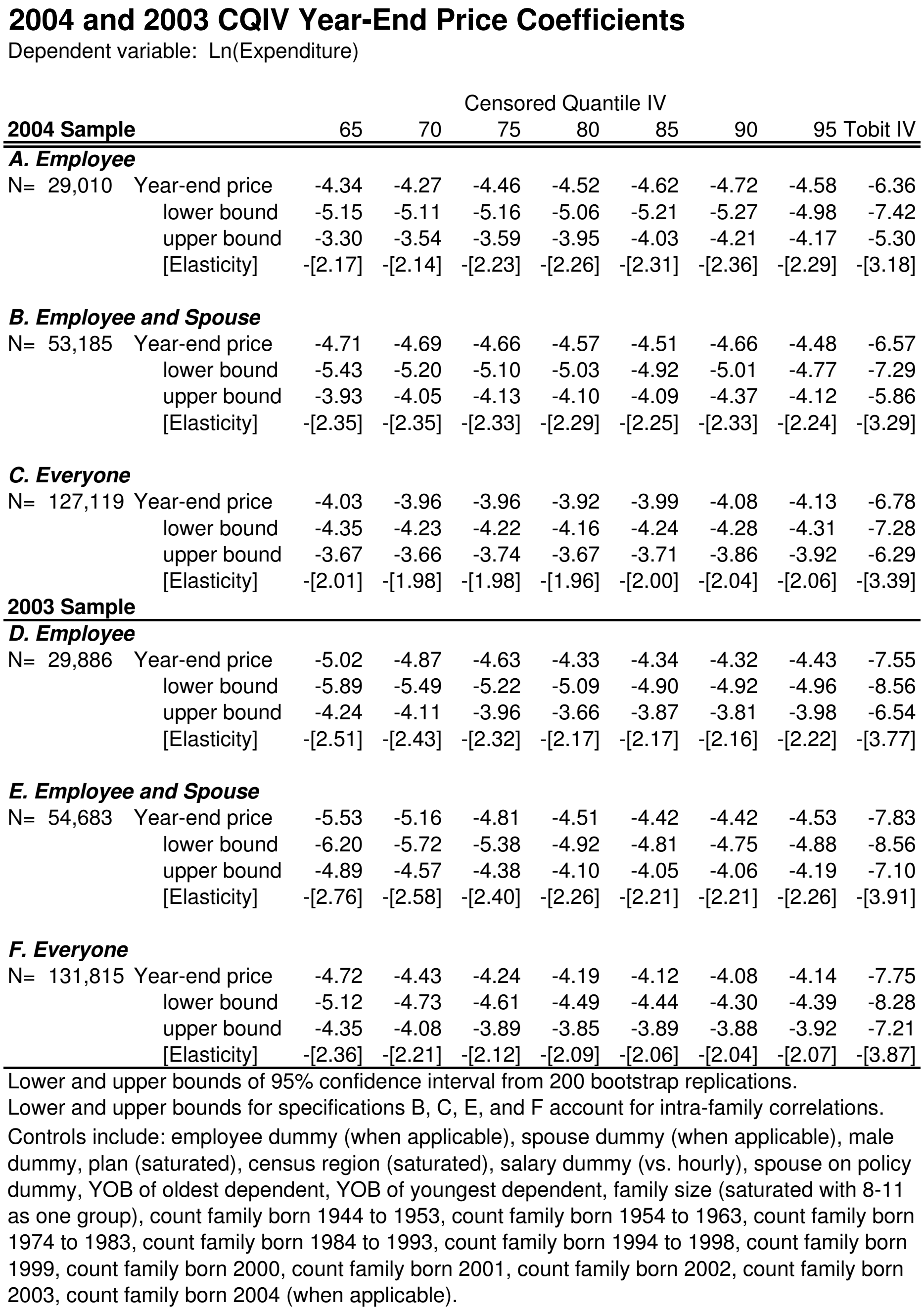
$$\eta_{arc} = \frac{\ln(\frac{y_a}{y_b})}{\ln(\frac{a}{b})}.$$

I use an arc elasticity instead of a point elasticity because, as discussed above, identification comes mainly from the large price drop from 1 to 0.2. Specifically, as a function of the estimated coefficient $\widehat{\alpha}$ at each quantile, and the prices of interest, the transformation that I use is as follows:

$$\widehat{\eta} = \frac{(\ln \widehat{E|P = 0.2}) - (\ln \widehat{E|P = 1})}{\ln(\frac{0.2}{1})} = \frac{\widehat{\alpha}(0.2-1)}{\ln(\frac{0.2}{1})} \approx .50\widehat{\alpha}.$$

This formula yields the "price elasticity of expenditure." By subtracting one from the expenditure elasticity, I could arrive at the price elasticity of demand for medical care. However, since the literature generally reports expenditure elasticities, I report expenditure elasticities in brackets under each coefficient. The upper and lower bounds of the bootstrapped 95% confidence interval can be transformed similarly.

Table 4: 2004 and 2003 CQIV Year-End Price Coefficients



2004 and 2003 CQIV Year-End Price Coefficients

Dependent variable: Ln(Expenditure)

		Censored Quantile IV							
2004 Sample		65	70	75	80	85	90	95	Tobit IV
A. Employee									
N= 29,010	Year-end price	-4.34	-4.27	-4.46	-4.52	-4.62	-4.72	-4.58	-6.36
	lower bound	-5.15	-5.11	-5.16	-5.06	-5.21	-5.27	-4.98	-7.42
	upper bound	-3.30	-3.54	-3.59	-3.95	-4.03	-4.21	-4.17	-5.30
	[Elasticity]	-[2.17]	-[2.14]	-[2.23]	-[2.26]	-[2.31]	-[2.36]	-[2.29]	-[3.18]
B. Employee and Spouse									
N= 53,185	Year-end price	-4.71	-4.69	-4.66	-4.57	-4.51	-4.66	-4.48	-6.57
	lower bound	-5.43	-5.20	-5.10	-5.03	-4.92	-5.01	-4.77	-7.29
	upper bound	-3.93	-4.05	-4.13	-4.10	-4.09	-4.37	-4.12	-5.86
	[Elasticity]	-[2.35]	-[2.35]	-[2.33]	-[2.29]	-[2.25]	-[2.33]	-[2.24]	-[3.29]
C. Everyone									
N= 127,119	Year-end price	-4.03	-3.96	-3.96	-3.92	-3.99	-4.08	-4.13	-6.78
	lower bound	-4.35	-4.23	-4.22	-4.16	-4.24	-4.28	-4.31	-7.28
	upper bound	-3.67	-3.66	-3.74	-3.67	-3.71	-3.86	-3.92	-6.29
	[Elasticity]	-[2.01]	-[1.98]	-[1.98]	-[1.96]	-[2.00]	-[2.04]	-[2.06]	-[3.39]
2003 Sample									
D. Employee									
N= 29,886	Year-end price	-5.02	-4.87	-4.63	-4.33	-4.34	-4.32	-4.43	-7.55
	lower bound	-5.89	-5.49	-5.22	-5.09	-4.90	-4.92	-4.96	-8.56
	upper bound	-4.24	-4.11	-3.96	-3.66	-3.87	-3.81	-3.98	-6.54
	[Elasticity]	-[2.51]	-[2.43]	-[2.32]	-[2.17]	-[2.17]	-[2.16]	-[2.22]	-[3.77]
E. Employee and Spouse									
N= 54,683	Year-end price	-5.53	-5.16	-4.81	-4.51	-4.42	-4.42	-4.53	-7.83
	lower bound	-6.20	-5.72	-5.38	-4.92	-4.81	-4.75	-4.88	-8.56
	upper bound	-4.89	-4.57	-4.38	-4.10	-4.05	-4.06	-4.19	-7.10
	[Elasticity]	-[2.76]	-[2.58]	-[2.40]	-[2.26]	-[2.21]	-[2.21]	-[2.26]	-[3.91]
F. Everyone									
N= 131,815	Year-end price	-4.72	-4.43	-4.24	-4.19	-4.12	-4.08	-4.14	-7.75
	lower bound	-5.12	-4.73	-4.61	-4.49	-4.44	-4.30	-4.39	-8.28
	upper bound	-4.35	-4.08	-3.89	-3.85	-3.89	-3.88	-3.92	-7.21
	[Elasticity]	-[2.36]	-[2.21]	-[2.12]	-[2.09]	-[2.06]	-[2.04]	-[2.07]	-[3.87]

Lower and upper bounds of 95% confidence interval from 200 bootstrap replications.

Lower and upper bounds for specifications B, C, E, and F account for intra-family correlations.

Controls include: employee dummy (when applicable), spouse dummy (when applicable), male dummy, plan (saturated), census region (saturated), salary dummy (vs. hourly), spouse on policy dummy, YOB of oldest dependent, YOB of youngest dependent, family size (saturated with 8-11 as one group), count family born 1944 to 1953, count family born 1954 to 1963, count family born 1974 to 1983, count family born 1984 to 1993, 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 (when applicable).

In all of the estimated quantiles, the CQIV expenditure elasticities are an order of magnitude larger than those in the literature. For example, at the .85 quantile of the expenditure distribution, the implied expenditure elasticity is -2.3, which indicates that a one percent increase in price would decrease spending at the .85 quantile of the expenditure distribution by 2.3 percent. This elasticity estimate is fairly stable across the quantiles from .65 to .95, indicating that price responsiveness, though strong, does not tend to vary widely among people in the highest quantiles of the expenditure distribution. Estimates at quantiles between the reported quantiles are similar.

Specifications B and C in Table 4 present coefficients estimated on samples that include spouses and other dependents in each family. 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. The elasticities estimated on the 2003 sample, presented in the bottom panels of Table 4, 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. Even though price responsiveness does not tend to vary across the estimated quantiles, as reported in Table A2, coefficients on some covariates have plausible signs but vary dramatically, indicating that a random coefficients model is appropriate in this application.

4.4 Comparison With Other Estimators

For comparison with previous literature, I compare my CQIV estimates to mean estimates. However, quantile estimators and 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 mean estimates, CQIV estimates are less sensitive to extreme values, and they are not based on parametric assumptions about censoring.

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, these conditions hold, and 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 elasticity, to a CQIV median elasticity estimate, a 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 last column of Table 4 presents Tobit IV coefficients. In all specifications, the estimated Tobit IV coefficient is more negative than all of the quantile coefficients, and the 95% confidence intervals barely overlap, indicating a rejection of the null hypothesis that the assumptions required by Tobit IV hold.

It should be noted that the Tobit IV coefficients imply an even larger elasticity than the CQIV coefficients, 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: a truncated model and a two-part model. The truncated elasticity estimate is -0.8, and the two-part model elasticity estimate is -1.6. As with the Tobit IV estimate, these estimates are generally much larger than those in the literature. However, Eichner (1998) reported a Tobit IV elasticity of -0.8 (the Eichner (1997) elasticity estimates varied from -0.22 to -0.32).

Given the insurance-induced mechanical relationship between price and expenditure, we expect mean elasticity estimates that do not account for endogeneity to be even larger. To assess the impact of endogeneity on the estimates, I compare the Tobit IV estimates to Tobit estimates. This comparison is similar to the comparison of IV to OLS estimates, but it is more appropriate in this context given the censoring and the logarithmic specification. As expected, the Tobit estimate that does not account for endogeneity yields an even larger elasticity estimate of -4.1 with a 95% confidence interval of -4.2 to -4.0. Censored quantile regression elasticity estimates that do not account for endogeneity do not exhibit such a large increase in magnitude relative to CQIV estimates, indicating that endogeneity could have less of an impact on quantile estimates relative to mean estimates in this context.

As another method of comparison between the CQIV estimates and mean estimates in the literature, I use conservative assumptions to transform the CQIV estimates into a single mean estimate. Assume, based on the CQIV estimates, that the expenditure elasticity is constant at -2.3 from the .65 quantile to the top of the expenditure distribution. Since we cannot accurately measure price responsiveness at other quantiles of the distribution, make the conservative assumption that the true elasticity at these quantiles is zero, although it is likely to be negative. We obtain a mean elasticity estimate by weighting the 0 and -2.3 elasticity estimates over all quantiles as follows: $(1 - 0.65) \times -2.3 = -0.805$. This conservative estimate, which implies that the true mean elasticity is more negative than -0.805, is still much larger than the RAND elasticity estimate of -0.2.

In the appendix, I discuss potential differences between the RAND estimates and my estimates. In short, the RAND researchers assume a myopic response to price, and I assume a forward-looking response to price. I provide evidence of forward-looking behavior in my data, and I provide simulation evidence that suggests that my estimates would be an order of magnitude smaller if I ignored forward-looking behavior.

5 Robustness and Specification Tests

5.1 Couples Data

Using data beyond my estimation sample, I conduct an indirect test of the exclusion restriction: one family member's injury can only affect another family member's expenditure through its effect on his marginal price. Specifically, I show that in families in which cost sharing interactions *cannot* occur, one family member's injury does *not* appear to be related to another family member's medical expenditure. At the firm that I study, in insurance policies for families of two ("couples"), one family member's spending has no mechanical effect on another family member's marginal price. 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 supports the validity of the exclusion restriction in the main sample.

To formalize this test, I use the following model, which I estimate with censored quantile regression:

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

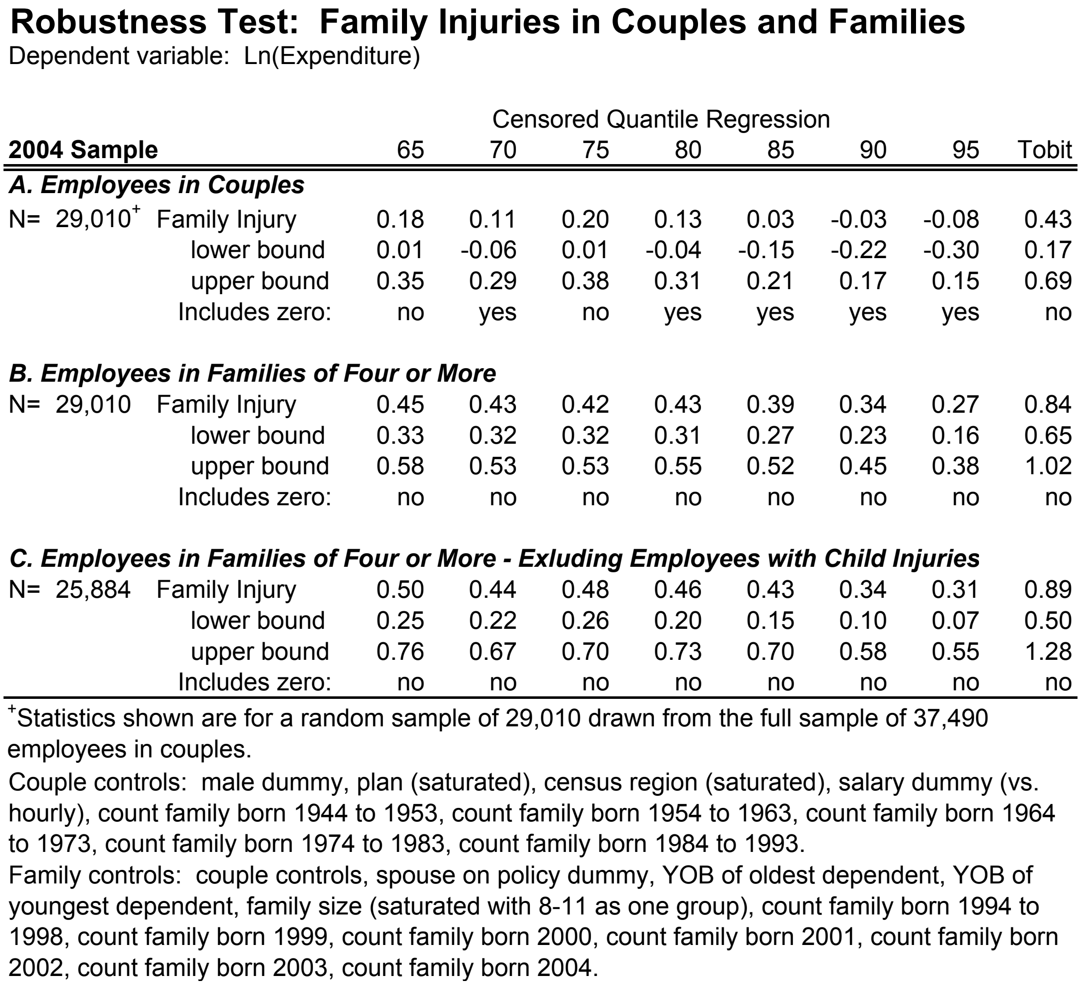
where the regressors are defined above. This specification differs from the main specification only in that, in instrumental variable terminology, it examines the reduced form effect of the family injury on *lnE* directly. A traditional instrumental variable specification would not be informative here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

I first estimate this specification on the "couples" sample of 2004 employees in employee-spouse families of two. Column 7 of Table 2 presents summary statistics on the couples sample. Comparison with Column 1 shows that the couples population consists mostly of older "empty nesters" and young couples without children. Furthermore, only 24% of employees in couples consume zero care, as opposed to 36% in families of four or more. Employees in couples have much higher average expenditures on medical care than their counterparts in families of four or more ($2,883 vs. $1,485). Given that employees in the couples sample consume more medical care, we should be more likely to observe spurious effects of other family injuries on spending in the couples sample than in the family sample. Since the couples sample is much larger than the family sample, to remove effects of sample size from the comparison, I conduct the estimation in 100 random subsets of the couples sample of the same size as the family sample.

The results in specification A of Table 5 show that the effect a spouse's injury on own expenditure is not statistically different from zero. In the 100 random couples samples taken together, the median point estimate at each quantile is generally not statistically different from zero.

For comparison, I estimate the same specification on employees in families of four or more, where family price interactions *can* occur. As shown in specification B of Table 5, the coefficients in the family specification suggest that employees with an injured spouse or child spend 0.27 to .45 percent more on their own medical care. In the family specification, the 95% confidence intervals never include zero, but they do include zero at almost all quantiles in the couples specifications. In the couples point estimates shown, even though the point-wise confidence intervals at the .65 and .75 quantiles do not include zero, a conservative calculation of a uniform confidence interval over all quantiles would include zero, given that the lower bounds at these quantiles are already so close to zero. At most quantiles, the entire confidence

Table 5: Robustness Test: Family Injuries in Couples and Families



Robustness Test: Family Injuries in Couples and Families
Dependent variable: Ln(Expenditure)

		Censored Quantile Regression							
2004 Sample		65	70	75	80	85	90	95	Tobit
A. Employees in Couples									
N= 29,010[+]	Family Injury	0.18	0.11	0.20	0.13	0.03	-0.03	-0.08	0.43
	lower bound	0.01	-0.06	0.01	-0.04	-0.15	-0.22	-0.30	0.17
	upper bound	0.35	0.29	0.38	0.31	0.21	0.17	0.15	0.69
	Includes zero:	no	yes	no	yes	yes	yes	yes	no
B. Employees in Families of Four or More									
N= 29,010	Family Injury	0.45	0.43	0.42	0.43	0.39	0.34	0.27	0.84
	lower bound	0.33	0.32	0.32	0.31	0.27	0.23	0.16	0.65
	upper bound	0.58	0.53	0.53	0.55	0.52	0.45	0.38	1.02
	Includes zero:	no	no	no	no	no	no	no	no
C. Employees in Families of Four or More - Exluding Employees with Child Injuries									
N= 25,884	Family Injury	0.50	0.44	0.48	0.46	0.43	0.34	0.31	0.89
	lower bound	0.25	0.22	0.26	0.20	0.15	0.10	0.07	0.50
	upper bound	0.76	0.67	0.70	0.73	0.70	0.58	0.55	1.28
	Includes zero:	no	no	no	no	no	no	no	no

[+] Statistics shown are for a random sample of 29,010 drawn from the full sample of 37,490 employees in couples.

Couple 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 controls: couple 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.

interval for the family point estimates exceeds the entire 95% confidence interval for the couples. Tobit coefficients, shown in the last row for comparison, do not include zero in the confidence interval, but they are substantially smaller in the couples specification than they are in the family specification. Overall, this comparison lends strong support to the validity of the exclusion restriction.

One concern with the comparison of the couples specification to the family specification is that identification in the couples specification comes only from injures to spouses, and identification in the family specification comes from injuries to children as well as spouses. To address this concern, I drop employees with injured children from the sample and re-estimate the family specification. Relative to the full family specification, I eliminate 3,126 employees with injured children, leaving 760 employees with injured spouses and 25,124 employees with no family injuries. Even though this restricted sample should have less power to produce coefficients statistically different from zero, as shown in specification C of Table 5, the confidence intervals do not include zero at any quantile, further reinforcing the validity of the exclusion restriction when compared to the couples specification. Moreover, the point estimates are stable across the two family specifications, suggesting that the identification strategy is robust to the source of the family injuries included in the instrument, an observation which I investigate more fully in Section 6.1.

5.2 Longitudinal Data

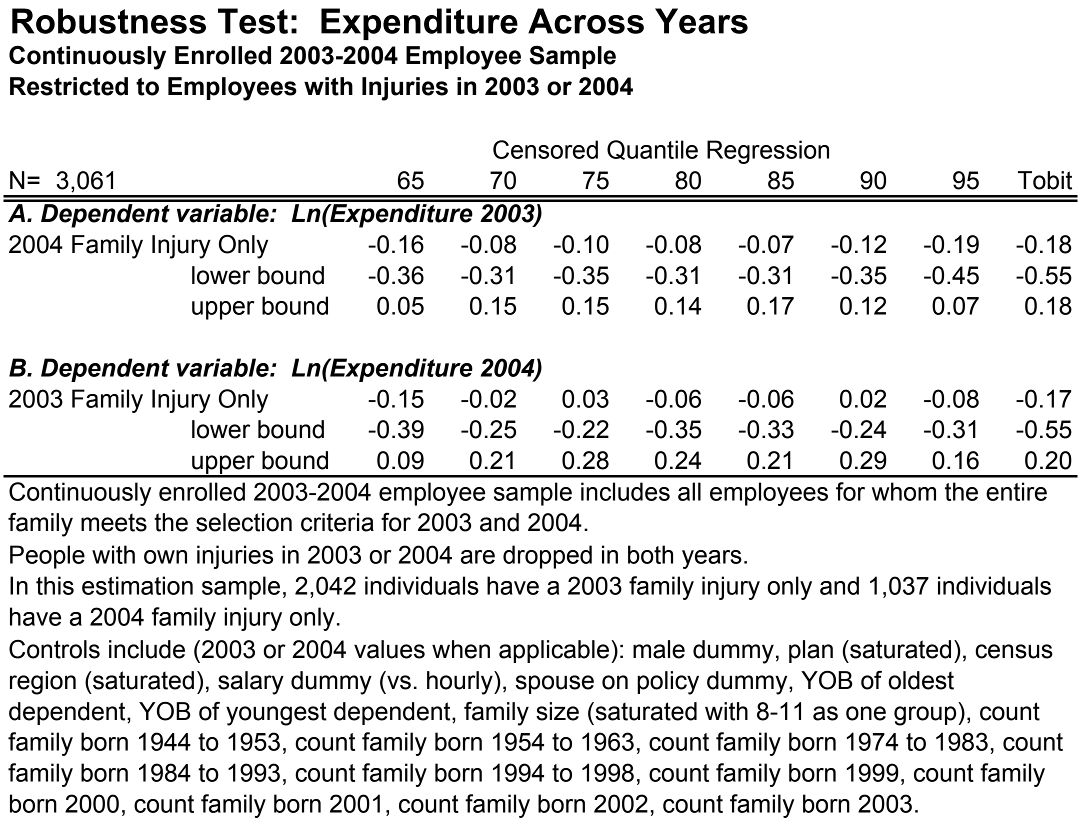
Using longitudinal data, I perform two related exercises: I conduct an indirect test of the exclusion restriction, and I investigate a potential source of the large magnitudes of my estimated elasticities. First, if the exclusion restriction holds, one family member's injury should only be related to another family member's spending through price interactions. The insurance benefits dictate that one family member's injury cannot affect another family member's marginal price in the previous year. Therefore, in 2003, employees with family injuries in 2004 should spend much less than employees with family injuries in 2003. Accordingly, I examine the reduced form effect of a family injury in 2004 on expenditure in 2003 in the sample of employees with family injuries in either year. As in the couples analysis above, I use a censored quantile specification rather than an instrumental variable specification because the first stage relationship between a family injury in one year and expenditure in another cannot exist.

I construct a longitudinal sample of employees in families of four or more in which every family member is continuously enrolled in 2003 and 2004. I exclude employees who have injuries themselves in either year, resulting in a sample of 18,743 individuals. In this sample, family injuries have limited persistence across years; only 295 employees have family injuries in both years. I further exclude individuals with no family injuries in either year and individuals with family injuries in both years, resulting in an estimation sample of 3,061 individuals.

The coefficients in specification A of Table 6 show that employees with family injuries in 2004 spend less in 2003 than employees with family injuries in 2003. The coefficients indicate that they spend 7 to 19 percent less across the .65 to .95 quantiles of the expenditure distribution. The Tobit coefficient in the last column indicates that mean spending is 18 percent less. The coefficients are not statistically significant, likely due to the relatively small sample size.

Second, I estimate a related specification that allows me to examine the claim

Table 6: Robustness Test: Expenditure Across Years



Robustness Test: Expenditure Across Years

Continuously Enrolled 2003-2004 Employee Sample

Restricted to Employees with Injuries in 2003 or 2004

	Censored Quantile Regression							
N= 3,061	65	70	75	80	85	90	95	Tobit
A. Dependent variable: Ln(Expenditure 2003)								
2004 Family Injury Only	-0.16	-0.08	-0.10	-0.08	-0.07	-0.12	-0.19	-0.18
lower bound	-0.36	-0.31	-0.35	-0.31	-0.31	-0.35	-0.45	-0.55
upper bound	0.05	0.15	0.15	0.14	0.17	0.12	0.07	0.18
B. Dependent variable: Ln(Expenditure 2004)								
2003 Family Injury Only	-0.15	-0.02	0.03	-0.06	-0.06	0.02	-0.08	-0.17
lower bound	-0.39	-0.25	-0.22	-0.35	-0.33	-0.24	-0.31	-0.55
upper bound	0.09	0.21	0.28	0.24	0.21	0.29	0.16	0.20

Continuously enrolled 2003-2004 employee sample includes all employees for whom the entire family meets the selection criteria for 2003 and 2004.

People with own injuries in 2003 or 2004 are dropped in both years.

In this estimation sample, 2,042 individuals have a 2003 family injury only and 1,037 individuals have a 2004 family injury only.

Controls include (2003 or 2004 values when applicable): male dummy, plan (saturated), census region (saturated), salary dummy (vs. hourly), spouse on policy dummy, YOB of oldest dependent, YOB of youngest dependent, family size (saturated with 8-11 as one group), count family born 1944 to 1953, count family born 1954 to 1963, count family born 1974 to 1983, count family born 1984 to 1993, 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.

that my estimated price elasticity is so large because it captures forward inter-year shifting of expenditures in response to a family injury. If this explanation is true, in 2004, employees with family injuries in 2003 should spend much less than families with injuries in 2004. Formally, if inter-year forward shifting occurs, in a related specification that examines the effect of a 2003 family injury on expenditure in 2004, we expect the coefficients to be more negative than the coefficients at each quantile in specification A, which examines the effect of a 2004 family injury in 2003. However, the coefficients in specification B of Table 6 are, if anything, less negative at each quantile than the coefficients in specification A, but they are not statistically significant. These results are subject to the caveat that without a longer panel, I cannot rule out forward shifting of expenditure from a multi-year time horizon. Taken together, the point estimates from both specifications of Table 6 do not suggest a violation of the exclusion restriction or forward-shifting of expenditures across years.

6 Additional Specifications

In this section, I consider variations on the main specification that allow me to examine heterogeneous treatment effects. Specifically, I examine variation across injuries to children vs. spouses and variation across inpatient vs. outpatient spending. In each setting, the variation in the estimates is small relative to the magnitude of the main estimates.

6.1 Injuries to Spouses vs. Children

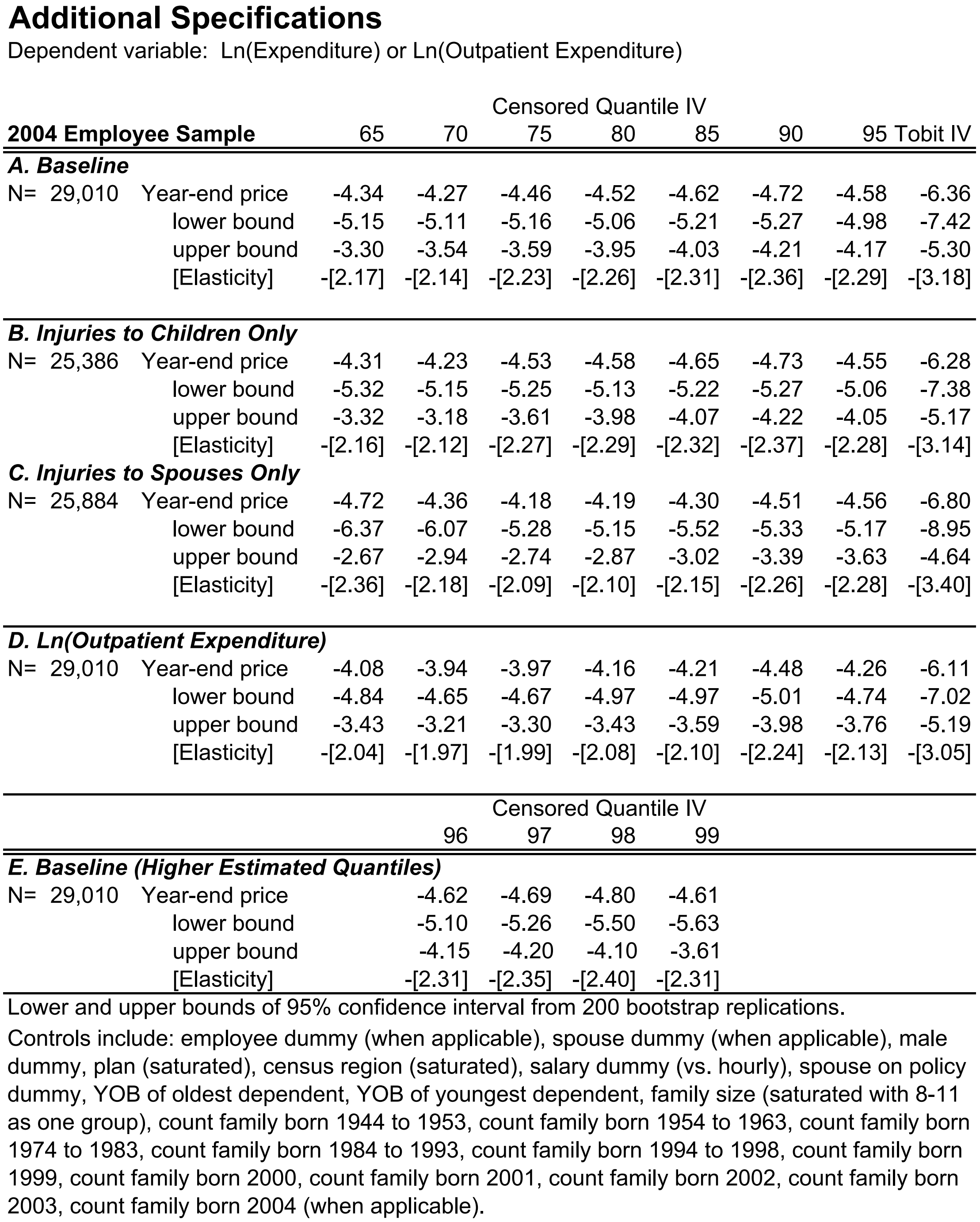
In my analysis, there is potential cause for concern if family injuries affect family income and family income affects expenditure. I cannot control for income directly because I do not observe it. However, I can informally investigate the role of income effects by estimating separate specifications 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.

I estimate two variations on the main specification: first I keep just employees with child injuries or no family injuries, and second I keep just the employees with spouse injuries or no family injuries. The second specification is similar to Specification C of Table 5, but is an instrumental variables specification instead of a reduced form specification. As shown in specifications B and C of Table 7, the 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 my 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. This suggests that variation in the estimates due to child vs. spouse injuries is not large relative to the main elasticity estimates.

6.2 Outpatient Spending vs. Total Spending

Quantile estimators are less sensitive to extreme values than mean estimators. However, to be sure that individuals with the highest expenditures are not driving the results, I estimate the main specification at the very highest quantiles, and I estimate another specification which includes only outpatient spending as the dependent variable. Since the potential for cross-substitution is so vast among the medical services covered by the plans that I study, and there is a great deal of judgment involved in categorizing different types of medical spending, I do not examine expenditure by

Table 7: Additional Specifications



Additional Specifications

Dependent variable: Ln(Expenditure) or Ln(Outpatient Expenditure)

		Censored Quantile IV							
2004 Employee Sample		65	70	75	80	85	90	95	Tobit IV
A. Baseline									
N= 29,010	Year-end price	-4.34	-4.27	-4.46	-4.52	-4.62	-4.72	-4.58	-6.36
	lower bound	-5.15	-5.11	-5.16	-5.06	-5.21	-5.27	-4.98	-7.42
	upper bound	-3.30	-3.54	-3.59	-3.95	-4.03	-4.21	-4.17	-5.30
	[Elasticity]	-[2.17]	-[2.14]	-[2.23]	-[2.26]	-[2.31]	-[2.36]	-[2.29]	-[3.18]
B. Injuries to Children Only									
N= 25,386	Year-end price	-4.31	-4.23	-4.53	-4.58	-4.65	-4.73	-4.55	-6.28
	lower bound	-5.32	-5.15	-5.25	-5.13	-5.22	-5.27	-5.06	-7.38
	upper bound	-3.32	-3.18	-3.61	-3.98	-4.07	-4.22	-4.05	-5.17
	[Elasticity]	-[2.16]	-[2.12]	-[2.27]	-[2.29]	-[2.32]	-[2.37]	-[2.28]	-[3.14]
C. Injuries to Spouses Only									
N= 25,884	Year-end price	-4.72	-4.36	-4.18	-4.19	-4.30	-4.51	-4.56	-6.80
	lower bound	-6.37	-6.07	-5.28	-5.15	-5.52	-5.33	-5.17	-8.95
	upper bound	-2.67	-2.94	-2.74	-2.87	-3.02	-3.39	-3.63	-4.64
	[Elasticity]	-[2.36]	-[2.18]	-[2.09]	-[2.10]	-[2.15]	-[2.26]	-[2.28]	-[3.40]
D. Ln(Outpatient Expenditure)									
N= 29,010	Year-end price	-4.08	-3.94	-3.97	-4.16	-4.21	-4.48	-4.26	-6.11
	lower bound	-4.84	-4.65	-4.67	-4.97	-4.97	-5.01	-4.74	-7.02
	upper bound	-3.43	-3.21	-3.30	-3.43	-3.59	-3.98	-3.76	-5.19
	[Elasticity]	-[2.04]	-[1.97]	-[1.99]	-[2.08]	-[2.10]	-[2.24]	-[2.13]	-[3.05]

		Censored Quantile IV			
		96	97	98	99
E. Baseline (Higher Estimated Quantiles)					
N= 29,010	Year-end price	-4.62	-4.69	-4.80	-4.61
	lower bound	-5.10	-5.26	-5.50	-5.63
	upper bound	-4.15	-4.20	-4.10	-3.61
	[Elasticity]	-[2.31]	-[2.35]	-[2.40]	-[2.31]

Lower and upper bounds of 95% confidence interval from 200 bootstrap replications.

Controls include: employee dummy (when applicable), spouse dummy (when applicable), male dummy, plan (saturated), census region (saturated), salary dummy (vs. hourly), spouse on policy dummy, YOB of oldest dependent, YOB of youngest dependent, family size (saturated with 8-11 as one group), count family born 1944 to 1953, count family born 1954 to 1963, count family born 1974 to 1983, count family born 1984 to 1993, 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 (when applicable).

therapeutic category. However, I recognize that an important area for future research is to determine which types of medical spending respond to marginal price. Here, I separate outpatient expenditure from total expenditure because the Medstat data clearly differentiates inpatient spending from outpatient spending and because RAND examined both types of spending separately.

Approximately 64% of the sample has some outpatient but no inpatient expenditure, and these individuals spend \$1,586 on average. In contrast, only 4% of the sample has some inpatient expenditure, and these individuals spend \$9,068 on average. Even though average spending is large, since such a small fraction of individuals have inpatient expenditures, it is unlikely that inpatient expenditures drive the CQIV results. As shown in specification E of Table 7, even at very high quantiles where we expect more inpatient expenditures, the estimated elasticities remain fairly stable around -2.3. Further, in specification D, which includes only outpatient expenditures, the elasticity estimates are not directly comparable because they represent the elasticity of outpatient expenditures only, but they are generally similar to the main estimates. It does not appear that the largest expenditures are responsible for the large magnitude of the main coefficients.

7 Conclusion

This paper makes several contributions. Using recent, detailed data and a rigorous identification strategy, I estimate the price elasticity of expenditure on medical care using a new censored quantile instrumental variables (CQIV) estimator. With the CQIV estimator, I go beyond standards in the literature by allowing the elasticity estimate to vary with the quantiles of the expenditure distribution, relaxing the distributional assumptions traditionally used to deal with censoring, and addressing endogeneity.

I find that across the .65 to .95 quantiles of the expenditure distribution, the price elasticity of expenditure is approximately -2.3. My estimated elasticities are an order of magnitude larger than those in the literature. I take several steps to compare my estimates to those in the literature, and I consider several sources of heterogeneity in the estimates. I conclude that the price elasticity of expenditure on medical care is much larger than the literature would suggest.

The task for my future research is to understand the welfare consequences of large price responsiveness. In ongoing research, I develop a structural model of the price elasticity of expenditure on medical care that is based on insurance-induced nonlinearities in consumer budget sets, following Hausman (1985). This model allows me to measure the welfare consequences of price-responsiveness and to examine the optimal nonlinear design of insurance.

A Comparison to RAND

A.1 Scope of Comparison

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 and 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. (For more background on the RAND experiment, see Newhouse et. al 1983.)

Below I discuss the calculation of the RAND estimates of the price elasticity of expenditure on medical care. I emphasize that the RAND methodology assumes a myopic response to contemporaneous marginal price, and my methodology assumes a forward-looking response to year-end marginal price. Next, I present simple suggestive evidence of forward-looking behavior among the individuals in my data. Lastly, I conduct a simulation in my data under conditions intended to mimic the plans and assumptions of the RAND experiment. The simulation shows that by assuming myopia when some individuals are forward-looking, it is possible to estimate an elasticity that is an order of magnitude smaller than the true elasticity.

A.2 Review of 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%, 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 with 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 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), page 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:

$$\eta_{midpoint} = \frac{(e_1 - e_2)/(e_1 + e_2)}{(p_1 - p_2)/(p_1 + p_2)} \tag{3}$$

where p denotes the coinsurance rate and e 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 as follows:

$$\eta_{RAND} = \frac{(71 - 55)/(71 + 55)}{(25 - 95)/(25 + 95)} \approx -0.22 \tag{4}$$

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. Another key difference between the RAND methodology and my methodology comes from the underlying treatment of myopia vs. foresight.

A.3 Evidence of Foresight

In the simple model of medical care expenditure on which I base my analysis, the most important parameter is the year-end marginal price. According to the model, if an individual expects to meet the stoploss by the end of the year, he will consume medical care all year as if his marginal price is zero, and expenditures paid at the randomized marginal rate will induce only an income effect. In contrast, by forecasting expenditures based on expenditure patterns before the stoploss is met, the Keeler and Rolph (1988) analysis assumes a strong form of myopia.

To investigate the validity of the myopia assumption, I conduct a simple test for forward-looking behavior in my data: if individuals are forward-looking, individuals who do not expect to meet the deductible should change their intra-year pattern of expenditures when a family injury occurs, but individuals who expect to meet the deductible should not. To examine people who plausibly expected to meet the deductible in 2004 regardless of a family injury, I identify individuals whose 2003 own spending exceeded the 2003 individual deductible as "High 2003." I identify all other individuals as "Low 2003". I restrict the sample to people with family injuries in 2004, and I compare average monthly expenditures before and after the month of the first family injury within these two spending categories. As in the main estimation sample, I exclude individuals with own injuries. I also omit individuals whose first family injuries occur in January or December so that it is always possible to observe spending before and after the family injury.

The top panel of Figure A1 presents the results from the sample of 2,265 employees with 2004 family injuries and complete 2003 expenditure data. (This sample is larger than the sample of employee with 2004 family injuries in Table 6 because I do not require the full family to be continuously enrolled in 2003.) A comparison of the two bars on the left to the two bars on the right shows that individuals with high 2003 spending spend more on average in 2004, regardless of the timing of the family injury. Within each set of bars, the comparisons provide evidence of forward-looking behavior. As expected, the left set of bars shows that employees with low 2003 spending spend more on average after the family injury than they did before the family injury. Also as expected, the right set of bars shows that employees with high 2003 spending do not appear to alter their spending patterns in response to the timing of a family injury.

Formally, neither difference is statistically significant. However, in the bottom panel, when I use the entire sample instead of just the employees, low spenders also spend more on average after the injury, and the difference in means is statistically significant for low 2003 spenders (t=-2.74) and is not statistically significant for high 2003 spenders (t=.4748). When I formalize this comparison of means using a Tobit model for the logarithm of expenditures, I find that low 2003 spenders spend 16% more after injuries in 2004 relative to high 2003 spenders, but this estimate is not statistically significant.

The question of whether consumers are myopic or forward-looking is complicated and interesting in its own right, and it should be investigated more completely. However, this test provides suggestive evidence against the Keeler and Rolph (1988) assumption of myopia. 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 simulation 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 to 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 care is free, which further attenuates elasticity estimates toward zero. The lack of experimental price variation among the highest spenders is unfortunate because, given the skewness in the distribution of medical expenditure, the price responsiveness of the highest spenders is very policy-relevant.

A.4 Simulation Exercise

To calculate expenditure elasticities, Keeler and Rolph (1988) simulated the expenditure response to plans with a higher stoploss than the true stoploss in their data. To illustrate potential bias in the Keeler and Rolph (1988) methodology, I conduct a theoretical reverse of the RAND exercise, in which I simulate the response to plans with lower stoplosses than the true stoplosses in my data. One advantage of my simulation over the RAND simulation is that it leads to within-sample predictions, whereas the RAND simulation led to out-of-sample predictions.

Since the RAND simulation included people who faced a zero effective year-end marginal price but attributed their behavior to a nonzero statutory marginal price, the RAND estimates should be biased toward zero. In my simulation, I simulate behavior governed by a zero effective marginal price, but I attribute this behavior to a nonzero statutory marginal price in the estimates, and I demonstrate the magnitude of the resulting bias toward zero. Under assumptions intended to mimic the conditions of the RAND experiment in my simulation, I estimate a simulated elasticity that is an order of magnitude smaller than the true elasticity.

The simulation steps are as follows:

1. Estimate the following specification using my data and my methodology:

$$\ln E = \alpha P + W'\beta + u \tag{5}$$

   where all variables are defined as above. Retain estimates for subsequent steps. In practice, I estimate my model in my data using Tobit IV, and I estimate a price elasticity of -3.2. I do not use CQIV for this simulation because I am interested in a mean estimate for comparison to RAND.

2. Predict log expenditure for all individuals using the estimated coefficients and the empirical values of P and X:

$$\widehat{\ln E} = \widehat{\alpha} P + W'\widehat{\beta} \tag{6}$$

3. To mimic the spending response to a new, lower stoploss than that in the actual plans, choose a group of individuals for whom the new stoploss will be low enough that they will reasonably expect to meet it. Calibrate the size of this group according to the percentage of individuals who met the stoploss in the RAND study. For this group, compute a simulated predicted expenditure, which assumes an effective marginal price of zero, even though the nominal year-end marginal price for these individuals in the actual plans is often nonzero:

$$\widetilde{\ln E} = \widehat{\alpha} * 0 + W'\widehat{\beta} \tag{7}$$

   Since $\widehat{\alpha} < 0$ and $P \geq 0$, it follows that $\widetilde{\ln E} > \widehat{\ln E}$. This makes intuitive sense because, given downward sloping demand, people who face a price of zero will spend *more* on medical care than they would if they faced a nonzero marginal price. For example, in the data, there is an individual who faces a year-end nominal marginal price of 0.2, and has total year-end spending of \$927. Based on his nominal marginal price and the values of his values of W, his predicted log spending is 5.7244, which by exponentiation, translates into \$306.25. In the simulation, when I predict his log spending based on a year-end effective price of zero, the new predicted value is 6.997, which by exponentiation, translates into \$1,093.

4. Re-estimate the price elasticity using my methodology on the data set of predicted expenditures and nominal marginal prices, and compare it to the "true" elasticity as computed by the price coefficient $\widehat{\alpha}$, estimated in the first step.

To determine whose expenditures to alter in the third step, I examine expenditures on the family level because the RAND stoplosses were on the family level. Since approximately 20% of subjects met the stoploss in the RAND study, I place approximately 20% of my sample into in hypothetical plans in which the effective marginal price is zero. Specifically, this subset includes 6,015 people with no family injuries whose total family spending exceeds \$5,500 (20.7% of the entire sample, and 23.9% of the sample with no family injuries).

It is plausible that families without injuries whose expenditures exceed \$5,500 would have met the \$1,000 stoploss in the RAND plans, even accounting for overall and medical inflation. In the least generous plan in my data, when family total beneficiary plus insurer spending is \$5,500, beneficiary spending is \$3,000+(\$5,500-\$3,000)*0.2=\$3,500. Similarly, in the most generous plan in my data, when family total beneficiary plus insurer spending is \$5,500, beneficiary spending is \$1,050+(\$5,500-\$1,050)*0.2= \$1,940. In my data, since the stoplosses are so much higher than they were in the RAND experiment, very small numbers of individuals meet the stoploss. Among the individuals whose expenditures I alter, the average statutory marginal price is .4 (29.4% at 1, 52.6% at 0.2, and 14.6% at 0).

When I re-estimate the model in the fourth step using predicted expenditures and nominal marginal prices, I estimate a price elasticity of -.34, which is an order of magnitude smaller than the original estimate of -3.2. It is possible to alter the expenditures of other plausibly-sized subsets of individuals to yield similar results. For example, when I alter the spending of a random 15% of individuals with no family injuries, I estimate a price elasticity of -.33. In addition, when I alter the spending of a random 50% of individuals with family spending that exceeds \$2,000 and no family injuries, I estimate a price elasticity of -0.28. The results of these simulation exercises suggest that if plausibly-sized groups of individuals are forward-looking, but they are assumed to be myopic, estimates of the price elasticity of expenditure on medical care could reflect a substantial bias toward zero.

References

[1] Angrist, Joshua, and Imbens, Guido. “Two-Stage Least Squares Estimation of Average Causal Effects in Models with Variable Treatment Intensity.” Journal of the American Statistical Association. 1995. 90(430), pp. 431-442.

[2] Angrist, Joshua, Imbens, Guido, and Rubin, Donald. “Identification of causal effects using instrumental variables.” Journal of the American Statistical Association. 1996, 91(434), pp. 444-455.

[3] Buntin, Melinda Beeukes, and Zaslavsky, Alan M. “Too Much Ado About Two-Part Models and Transformation? Comparing Methods of Modeling Health Care Costs.” Journal of Health Economics. 2004, 23, pp. 525-542.

[4] Chernozhukov, Victor, and Hansen, Christian. “Instrumental variable quantile regression: A robust inference approach.” Journal of Econometrics. January 2008. 142(1), pp.379-398.

[5] Chernozhukov, Victor, and Hong, Han. “Three-Step Quantile Regression and Extramarital Affairs.” Journal of The American Statistical Association. September 2002, 97(459). pp. 872-882.

[6] Chernozhukov, Victor, Fernandez-Val, Ivan, and Kowalski, Amanda. “Censored Quantile instrumental variable Estimation via Control Functions.” 2008. mimeo.

[7] Duan, Naihua, Manning, Willard G Jr., Morris, Carl N., Newhouse, Joseph P. “A Comparison of Alternative Models for the Demand for Medical Care.” Journal of Business & Economic Statistics, 1983, 1(2), pp. 115-126.

[8] Eichner, Matthew J. “Medical Expenditures and Major Risk Health Insurance,”Massachusetts Institute of Technology dissertation, chapter 1, 1997, 1-66.

[9] Eichner, Matthew J. “The Demand for Medical Care: What People Pay Does Matter.” The American Economic Review. Papers and Proceedings of the Hundred and Tenth Annual Meeting of the American Economic Association. May 1998. 88(2) pp. 117-121.

[10] Hausman, Jerry A. “Specification Tests in Econometrics.” Econometrica, 1978, 46(6), pp. 1251-71.

[11] Hausman, Jerry A. “The Econometrics of Nonlinear Budget Sets.” Econometrica, 1985, 53(6), pp. 1255-82.

[12] Kaiser Family Foundation and Health Research Educational Trust. “Employer Health Benefits Annual Survey 2006.”

[13] Keeler, Emmett and Rolph, John E. “The Demand for Episodes of Treatment in the Health Insurance Experiment.” Journal of Health Economics, 1988, 7, pp. 337-367.

[14] Kowalski, Amanda E. “Censored Quantile instrumental variable Estimates of the Price Elasticity of Expenditure on Medical Care.” Massachusetts Institute of Technology dissertation, chapter 1, 2008 pp. 1-96.

[15] Manning, Willard G., Newhouse, Joseph P., Duan, Naihua, Keeler, Emmett B., and Leibowitz, Arleen. “Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment.” The American Economic Review, Jun 1987, 77(3), pp. 251-277.

[16] “Medical Expenditure Panel Survey.” Agency for Healthcare Research and Quality. 2004.

[17] “MarketScan Database,” Ann Arbor,MI: The MEDSTAT Group Inc., 2004.

[18] Mullahy, John. “Much ado about two: reconsidering retransformation and the two-part model in health econometrics.” Journal of Health Economics, 1998, 17, pp. 247-281.

[19] Newey, Whitney K. “Efficient Estimation of Limited Dependent Variable Models with Endogenous Explanatory Variables.” Journal of Econometrics, 1987, 36, pp. 231-250.

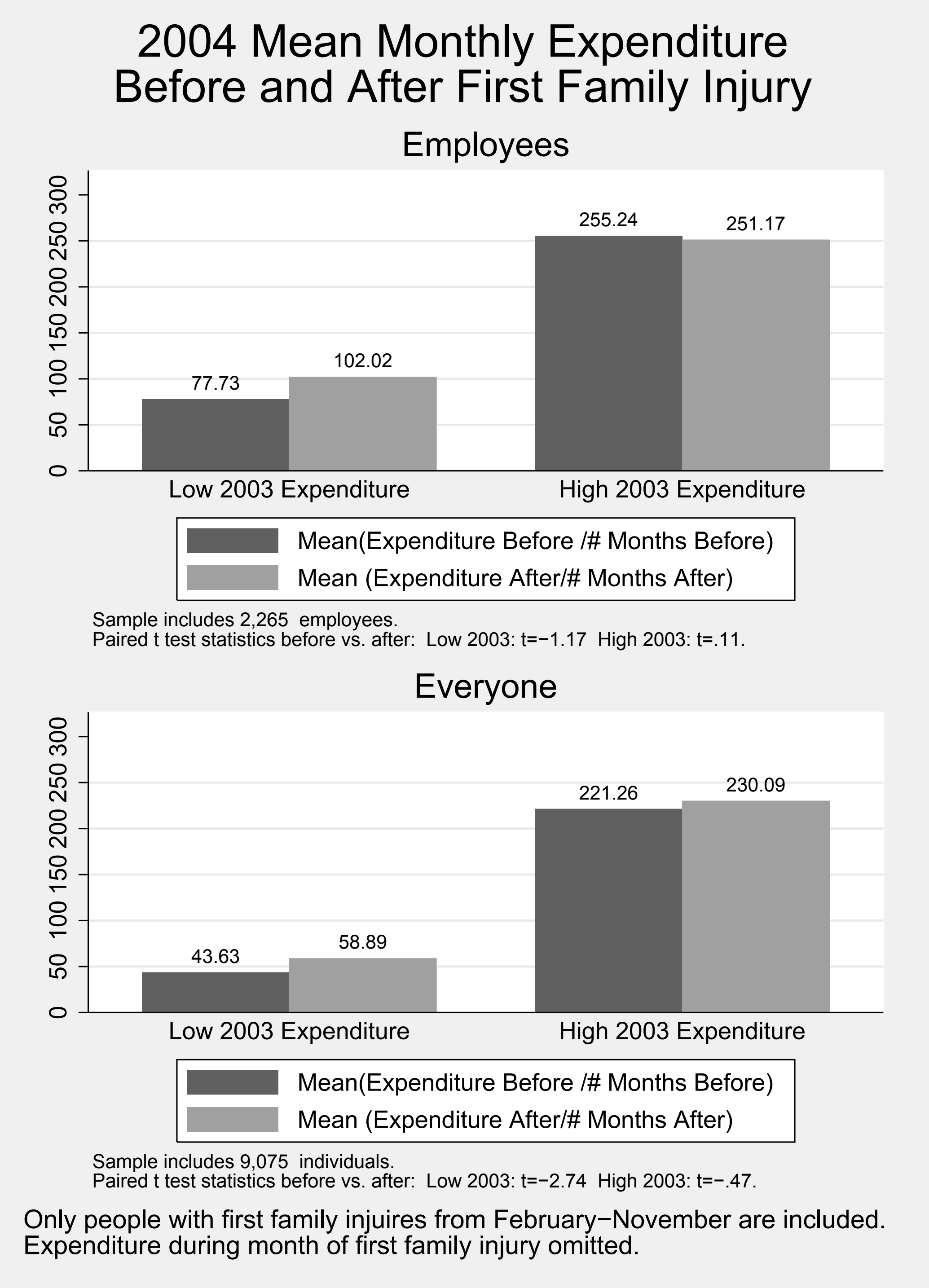
[20] Newhouse, Joseph P., Phelps, Charles E., and Marquis, M. Susan. “On Having Your Cake and Eating it too: Econometric Problems in Estimating the Demand for Health Services. ” Journal of Econometrics, 1980, 13, pp. 365-390.

[21] Newhouse, Joseph P and the Insurance Experiment Group. Free for All? Lessons from the RAND Health Insurance Experiment. Harvard University Press. Cambridge: 1993.

[22] Powell, James L. “Censored Regression Quantiles.” Journal of Econometrics, 1986, 23, pp.143-155.

[23] Tobin, J. “Estimation of Relationships for Limited Dependent Variables.” Econometrica, 1958, pp. 24-36.

Figure A1: 2004 Mean Monthly Expenditure
Before and After First Family Injury



Sample includes 2,265 employees.
Paired t test statistics before vs. after: Low 2003: t=−1.17 High 2003: t=.11.

Sample includes 9,075 individuals.
Paired t test statistics before vs. after: Low 2003: t=−2.74 High 2003: t=−.47.

Only people with first family injuires from February−November are included.
Expenditure during month of first family injury omitted.

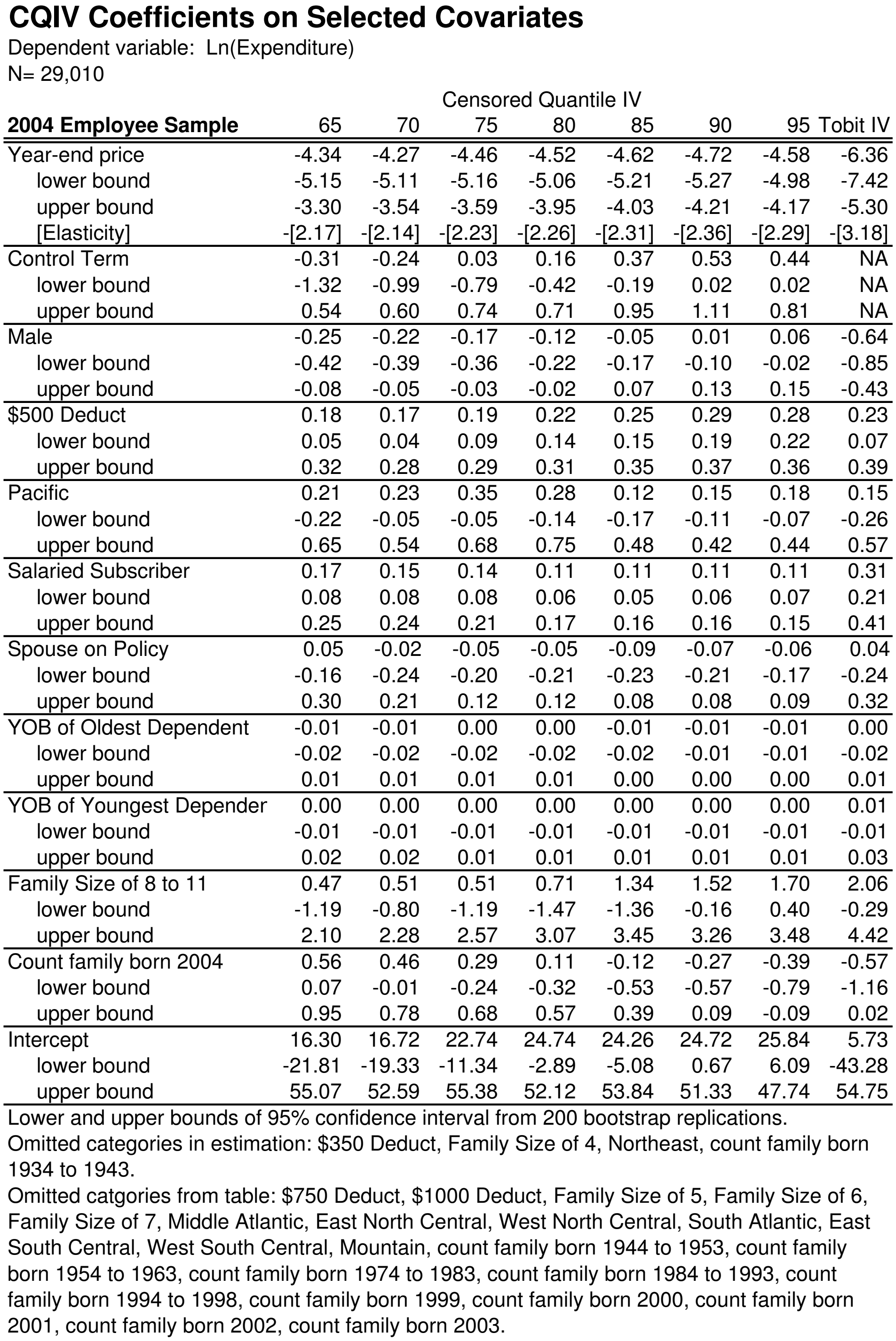
Table A1: 2003 Summary Statistics

2003 Summary Statistics

Cells report column % by variable

			Families of Four or More			
	Employees	Everyone	Employees		Everyone	
	All	All	NO Family	Family	NO Family	Family
			Injury	Injury	Injury	Injury
Variable	(1)	(2)	(3)	(4)	(5)	(6)
A. Year-end Expenditure ($)						
0	36.8	40.7	38.0	28.7	41.9	31.6
.01 to 100.00	11.4	12.5	11.4	11.4	12.5	12.2
100.01 to 1,000	31.1	31.7	30.7	34.4	31.1	36.2
1,000.01 to 10,000	17.8	13.3	17.2	22.3	12.7	17.5
10,000.01 to 100,000	2.8	1.8	2.7	3.3	1.7	2.4
100,000.01 and up	0.0	0.1	0.0	0.1	0.0	0.1
B. Year-end Price						
0	3.4	2.5	3.1	5.3	2.2	4.8
0.2	37.5	31.6	35.8	49.4	29.8	44.9
1	59.1	65.9	61.1	45.3	68.0	50.3
C. Family Injury						
0 (NO Family Injury)	87.7	88.3	100.0	0.0	100.0	0.0
1 (Family Injury)	12.3	11.7	0.0	100.0	0.0	100.0
D. Family Size						
2	0.0	0.0	0.0	0.0	0.0	0.0
3	0.0	0.0	0.0	0.0	0.0	0.0
4	66.1	59.2	67.2	58.1	60.5	49.3
5	24.9	27.9	24.3	29.0	27.4	32.1
6	6.8	9.1	6.4	9.5	8.6	12.8
7	1.7	2.6	1.5	2.4	2.4	3.9
8 to 12	0.6	1.2	0.6	0.9	1.1	1.9
E. Relation to Employee						
Employee	100.0	22.7	100.0	100.0	22.5	23.9
Spouse	0.0	18.8	0.0	0.0	18.7	19.9
Child/Other	0.0	58.5	0.0	0.0	58.8	56.2
F. Male						
0 (Female)	43.4	50.1	43.4	43.1	50.1	50.2
1 (Male)	56.6	49.9	56.6	56.9	49.9	49.8
G. Year of Birth						
1934 to 1943	0.2	0.1	0.2	0.2	0.1	0.1
1944 to 1953	4.7	2.1	4.8	4.4	2.1	2.0
1954 to 1963	33.9	14.1	34.0	33.3	14.0	14.6
1964 to 1973	50.2	20.0	50.0	51.5	19.8	21.6
1974 to 1983	10.9	6.7	11.0	10.6	6.7	6.8
1984 to 1993	0.0	30.4	0.0	0.0	30.5	29.8
1994 to 1998	0.0	15.5	0.0	0.0	15.6	14.8
1999 to 2003	0.0	11.0	0.0	0.0	11.4	10.2
H. Employee Class						
Salary Non-union	29.4	29.8	29.4	29.5	29.9	29.0
Hourly Non-union	70.6	70.2	70.6	70.5	70.1	71.0
I. US Census Region						
New England	1.4	1.4	1.4	1.6	1.4	1.7
Middle Atlantic	1.7	1.7	1.7	1.7	1.7	1.5
East North Central	15.5	15.6	15.4	16.7	15.5	16.9
West North Central	12.3	12.2	11.8	15.4	11.9	15.1
South Atlantic	18.5	18.4	19.2	13.5	19.1	13.2
East South Central	10.8	10.7	10.7	11.3	10.6	11.7
West South Central	29.1	29.1	29.2	28.7	29.2	28.5
Mountain	7.8	8.0	7.7	8.4	7.9	8.7
Pacific	2.9	2.9	2.9	2.8	2.9	2.8
J. Plan by Individual Deductible						
350	63.4	63.4	62.7	68.8	62.7	69.0
500	17.1	17.0	17.4	15.6	17.2	15.4
750	5.7	5.6	5.8	4.5	5.8	4.2
1000	13.8	13.9	14.1	11.0	14.2	11.4
Sample Size	*29,886*	*131,815*	*26,201*	*3,685*	*116,393*	*15,422*

Table A2: CQIV Coefficients on Selected Covariates



CQIV Coefficients on Selected Covariates

Dependent variable: Ln(Expenditure)
N= 29,010

	Censored Quantile IV							
2004 Employee Sample	65	70	75	80	85	90	95	Tobit IV
Year-end price	-4.34	-4.27	-4.46	-4.52	-4.62	-4.72	-4.58	-6.36
lower bound	-5.15	-5.11	-5.16	-5.06	-5.21	-5.27	-4.98	-7.42
upper bound	-3.30	-3.54	-3.59	-3.95	-4.03	-4.21	-4.17	-5.30
[Elasticity]	-[2.17]	-[2.14]	-[2.23]	-[2.26]	-[2.31]	-[2.36]	-[2.29]	-[3.18]
Control Term	-0.31	-0.24	0.03	0.16	0.37	0.53	0.44	NA
lower bound	-1.32	-0.99	-0.79	-0.42	-0.19	0.02	0.02	NA
upper bound	0.54	0.60	0.74	0.71	0.95	1.11	0.81	NA
Male	-0.25	-0.22	-0.17	-0.12	-0.05	0.01	0.06	-0.64
lower bound	-0.42	-0.39	-0.36	-0.22	-0.17	-0.10	-0.02	-0.85
upper bound	-0.08	-0.05	-0.03	-0.02	0.07	0.13	0.15	-0.43
$500 Deduct	0.18	0.17	0.19	0.22	0.25	0.29	0.28	0.23
lower bound	0.05	0.04	0.09	0.14	0.15	0.19	0.22	0.07
upper bound	0.32	0.28	0.29	0.31	0.35	0.37	0.36	0.39
Pacific	0.21	0.23	0.35	0.28	0.12	0.15	0.18	0.15
lower bound	-0.22	-0.05	-0.05	-0.14	-0.17	-0.11	-0.07	-0.26
upper bound	0.65	0.54	0.68	0.75	0.48	0.42	0.44	0.57
Salaried Subscriber	0.17	0.15	0.14	0.11	0.11	0.11	0.11	0.31
lower bound	0.08	0.08	0.08	0.06	0.05	0.06	0.07	0.21
upper bound	0.25	0.24	0.21	0.17	0.16	0.16	0.15	0.41
Spouse on Policy	0.05	-0.02	-0.05	-0.05	-0.09	-0.07	-0.06	0.04
lower bound	-0.16	-0.24	-0.20	-0.21	-0.23	-0.21	-0.17	-0.24
upper bound	0.30	0.21	0.12	0.12	0.08	0.08	0.09	0.32
YOB of Oldest Dependent	-0.01	-0.01	0.00	0.00	-0.01	-0.01	-0.01	0.00
lower bound	-0.02	-0.02	-0.02	-0.02	-0.02	-0.01	-0.01	-0.02
upper bound	0.01	0.01	0.01	0.01	0.00	0.00	0.00	0.01
YOB of Youngest Depender	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01
lower bound	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01	-0.01
upper bound	0.02	0.02	0.01	0.01	0.01	0.01	0.01	0.03
Family Size of 8 to 11	0.47	0.51	0.51	0.71	1.34	1.52	1.70	2.06
lower bound	-1.19	-0.80	-1.19	-1.47	-1.36	-0.16	0.40	-0.29
upper bound	2.10	2.28	2.57	3.07	3.45	3.26	3.48	4.42
Count family born 2004	0.56	0.46	0.29	0.11	-0.12	-0.27	-0.39	-0.57
lower bound	0.07	-0.01	-0.24	-0.32	-0.53	-0.57	-0.79	-1.16
upper bound	0.95	0.78	0.68	0.57	0.39	0.09	-0.09	0.02
Intercept	16.30	16.72	22.74	24.74	24.26	24.72	25.84	5.73
lower bound	-21.81	-19.33	-11.34	-2.89	-5.08	0.67	6.09	-43.28
upper bound	55.07	52.59	55.38	52.12	53.84	51.33	47.74	54.75

Lower and upper bounds of 95% confidence interval from 200 bootstrap replications.

Omitted categories in estimation: $350 Deduct, Family Size of 4, Northeast, count family born 1934 to 1943.

Omitted catgories from table: $750 Deduct, $1000 Deduct, Family Size of 5, Family Size of 6, Family Size of 7, Middle Atlantic, East North Central, West North Central, South Atlantic, East South Central, West South Central, Mountain, count family born 1944 to 1953, count family born 1954 to 1963, count family born 1974 to 1983, count family born 1984 to 1993, 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.