WORKING PAPERS

N° 1477

September 2023

"The Demand Elasticity of Health Care Spending for Low-Income Individuals"

Angelique Acquatella

Toulouse School of Economics

The Demand Elasticity of Health Care Spending for Low-Income Individuals

Angelique Acquatella
Toulouse School of Economics *

September 29, 2023

Low-income individuals are typically the most price sensitive segment of the market, but this is not true in the market for health care services. I show that low-income individuals have a smaller demand elasticity of medical spending with respect to coinsurance, relative to their higher income counterparts, using data from the RAND Health Insurance experiment. The null effect is driven by disproportionate share of low-income individuals who consume zero health care. The key insight is that low-income individuals may optimally consume zero health care because, when marginal utility of consumption is high, forgoing non-medical consumption becomes very costly.

Keywords: income effects, health care demand elasticity, corner solution
JEL Codes: I12, I14, D11

*Contact: angelique.acquatella@tse-fr.eu. I am grateful to David Cutler, Marty Gaynor, Maggie Shi, Rena Conti, Jim Rebitzer, Matthew Zahn, Michael Ricks, Valentina Reig, Victoria Marone, Ben Vatter, Jean-Marie Lozachmeur, François Salanié, Christian Hellwig, and Samuel Arenberg for helpful comments and suggestions.

1 Introduction

While low-income individuals are typically the most price sensitive segment of the market, this does not hold true in the market for health care services. Low-income individuals are less responsive to health care subsidies[1] and use fewer health care resources[2] than their higher income counterparts. Recent literature suggests that the socioeconomic gradient in utilization would prevail even in the presence of generous public health care subsidies (Shepard, Baicker, and Skinner, 2020). Some potential explanations to these puzzling empirical patterns include that health insurance has little value for low-income individuals (as measured by their willingness to pay for insurance; see Finkelstein, Hendren, and Shepard, 2019), or that behavioral biases are particularly acute among low-income populations (Bhargava, Loewenstein, and Sydnor, 2017).

Using data from the RAND Health Insurance experiment, I show that the reason low-income individuals appear less responsive to prices, on average, is because a disproportionate share consume zero health care. Low-income individuals who *do* choose to consume some positive amount of care *are* price sensitive, as in other markets. Moreover, the choice to consume no health care can be considered *optimal*. The key insight is that low-income individuals will optimally consume no health care because forgoing non-medical consumption is very costly when the marginal utility of consumption is high. Albeit a simple and foundational point, income effects in the demand for health care services (conditional on a particular choice of insurance) are largely ignored in empirical work, with one notable exception (Finkelstein, Hendren, and Luttmer, 2019). These income effects have important implications for estimates of moral hazard and of the fiscal costs of expanding subsidized health care (e.g. Medicaid) for low-income individuals.

In this article, I argue theoretically and show empirically that low-income individuals exhibit a small demand elasticity for health care services, on average. Theoretically, I illustrate the simple point that the marginal cost of forgoing consumption is large for low-income individuals, which makes them more likely to choose zero health care spending (the corner solution). Empirically, I use data from the RAND Health Insurance Experiment to show that the causal effect of health care subsidies on spending is near zero for the lowest-income tercile, driven by a disproportionate share of low-income individuals who consume no health care (even when it is free). While dated, the RAND experiment allows us to study the relationship between income and health care demand without confounding supply-side factors that also affect low-income consumption of health care services (e.g. limited provider entry in low-income areas, insufficient Medicaid payments, or availability of unobserved alternatives such as uncompensated or charitable care).

Basic economics tells us that optimal medical spending trades off the marginal benefits against the marginal costs of forgoing consumption. Because of health insurance, the costs of forgoing consumption are a function of the relative (coinsurance) price of health care, which is equal to the

[1] The standard literature estimate for the demand elasticity of health care is -0.2 (Manning et al., 1987), yet literature estimates of the demand elasticity for health care services among the low-income subgroup range around -0.1 and -0.16 (Lavetti, De Leire, and Ziebarth, 2023; Chandra, Gruber, and McKight, 2014).

[2] See McClellan and Skinner, 2006; Bhattacharya and Lakdawalla, 2006; Gornick et al., 1996; Link et al., 1987.

out-of-pocket amount for which the consumer is liable. When the utility function is concave in consumption, the marginal cost of forgoing consumption is equal to the relative price of the health care *scaled* by the marginal utility of consumption. Models of health care demand often make the convenient assumption that utility is quasi-linear in some numeraire good, which allows one to quantify the relative price of health care by looking only at the marginal out-of-pocket share. Given the complexity and non-linearities in health insurance contracts, quantifying the marginal out-of-pocket cost is seldom a trivial exercise. Quasi-linear utility, however, shuts down income effects which have important implications for observed patterns of health care consumption along the income gradient.

In particular, income effects imply that low-income individuals are more likely to choose the corner solution of zero health care along a range of out-of-pocket costs. This implies that reductions in out-of-pocket costs may not affect behavior at all. However, for individuals at the middle of the income distribution who are already consuming some positive amount of health care, reductions in out-of-pocket costs should have a large positive impact on their health care spending. Thus, we should expect to see a small demand elasticity of medical spending (with respect to cost-sharing) for individuals at the bottom of the income distribution, and a larger elasticity for individuals in the middle. Reexamining data from the RAND Health Insurance Experiment corroborates this prediction: I find that individuals at the bottom of the income distribution have a small (and statistically insignificant) elasticity of medical spending with respect to out-of-pocket costs relative to their higher income counterparts.

This is the first paper to highlight that low-income individuals appear to have a *lower elasticity* for health care services relative to higher income individuals.[3] However, the finding is consistent with prior literature. Chandra, Gruber, and McKnight (2014) estimate an elasticity of -0.16 for a low-income population, and Lavetti, DeLeire, and Ziebarth (2023) estimate an elasticity of -0.1 for low-income enrollees in the Affordable Care Act marketplaces. Both estimates are within the range of the confidence interval for the -0.2 estimate from the RAND experiment (Manning et al., 1987), yet it is nonetheless surprising that the elasticity is not larger than the average. Brot-Goldberg et al. (2017) similarly find an income gradient in empirical responses to cost-sharing, where the average response among middle-income individuals is smaller than the response of higher income individuals (see online Appendix).

Finkelstein, Hendren, and Luttmer (2019) are the first to formally account for income effects in the demand for health care services, and my article can be viewed as complementary to their work by highlighting that the corner solution dimension of income effects has important implications for elasticity estimates. In Finkelstein, Hendren, and Luttmer (2019), the authors quantify the welfare effects of Medicaid by estimating the willingness to pay for this program. A key component of willingness to pay is the marginal rate of substitution between health care and consumption, and

[3]This finding contrasts the literature on the demand of low-income individuals for health *insurance*, where evidence supports that low-income individuals are quite price sensitive to premiums (Krueger and Kuziemko, 2013; Cutler and Reber, 1998; Nyman, 2003).

the authors correctly point out that the marginal utility of consumption is higher for low-income individuals. Thus, they scale down their calibrated marginal rate of substitution of health.

This article proceeds as follows. In Section 2, I lay out a general model of health care spending and characterize the tradeoffs in the consumer's decision. In Section 3, I present evidence of heterogeneous responses to cost-sharing along the socioeconomic gradient that are consistent with the theory. In Section 4, I conclude by discussing the implications of income effects for estimates of health care demand.

2 Tradeoffs in the health care consumption decision

I present a simple model of health care demand where utility is concave in consumption. Individuals choose how much to spend on health care by trading off the marginal benefits of care against their marginal costs of forgoing consumption, which consist of their marginal out-of-pocket costs scaled by the marginal utility of non-medical consumption. The model implies that low-income individuals consume less health care, overall, and that health care subsidies have differential effects on the health care demand of the high- versus low-income, resulting in low-income individuals being most inelastic on average.

To build intuition, consider a low-income individual with chronic back pain. While she may unquestionably derive positive benefits from receiving care for pain management, forgoing her monthly rent payment may be too costly for her to choose any care. Her marginal cost of forgone consumption is large because her low level of income (consumption) put her at the steepest part of her utility function. High marginal costs of forgone consumption imply that agents may not be able to equate marginal benefit to marginal cost, thus they may optimally choose to consume zero health care.

Even as insurance becomes more generous, the tradeoff for a low-income individual may hardly be altered. This is because marginal reductions in the out-of-pocket share are scaled by the (large) marginal utility of consumption; e.g., its still not worthwhile to forgo the monthly rent payment when she has to pay 20% versus 15% of a physical therapy visit. Even as care becomes nominally free, the hassle costs associated with consuming health care (e.g., parking or taking the bus, taking time off from work, scheduling the appointment) get scaled up by the marginal utility of consumption, which implies that some low-income individuals may not be induced to consume any care when the price is zero.

For high-income individuals, however, the marginal utility of consumption is small and close to zero. This implies that they tradeoff (presumably positive) marginal health benefits of additional care against a very small marginal cost of forgone consumption and they will have higher levels of health care expenditures (Acemoglu, Finkelstein, and Notowidigdo, 2013). Moreover, the demand elasticity with respect to coinsurance for high-income individuals should be small: a $50 copay for chronic back pain management is essentially free for a high-income individual, and behavior is unlikely to differ with or without a copay.

2.1 A model of health care demand

Consider individuals who derive utility from two goods: health and consumption of a numeraire good. Individuals cannot directly purchase health but can incur medical expenditures that affect their health. Individuals differ in type, denoted by θ_i, which determines the extent to which medical expenditures impact the individual's health. Assume that the marginal gains of medical spending are bounded at zero spending. For consumption c, health $H(m, \theta_i)$, and medical expenditures m, individual utility is given by:

$$u(c, H(m; \theta_i)). \tag{1}$$

where utility is increasing and concave in consumption, $u_c \geq 0$ and $u_{cc} \leq 0$, and health, $u_h \geq 0$ and $u_{hh} \leq 0$. Assume that the first unit of consumption gives infinite marginal utility at any health state $u_c(0, H(m; \theta_i)) \to \infty$; but that the first unit of health care gives finite marginal health gains $H'(0; \theta_i) \ll \infty$.

Individuals make optimal choices about consumption and medical spending subject to a budget constraint. For income y and out-of-pocket costs $oop(m)$, the budget constraint is:

$$y = c + oop(m). \tag{2}$$

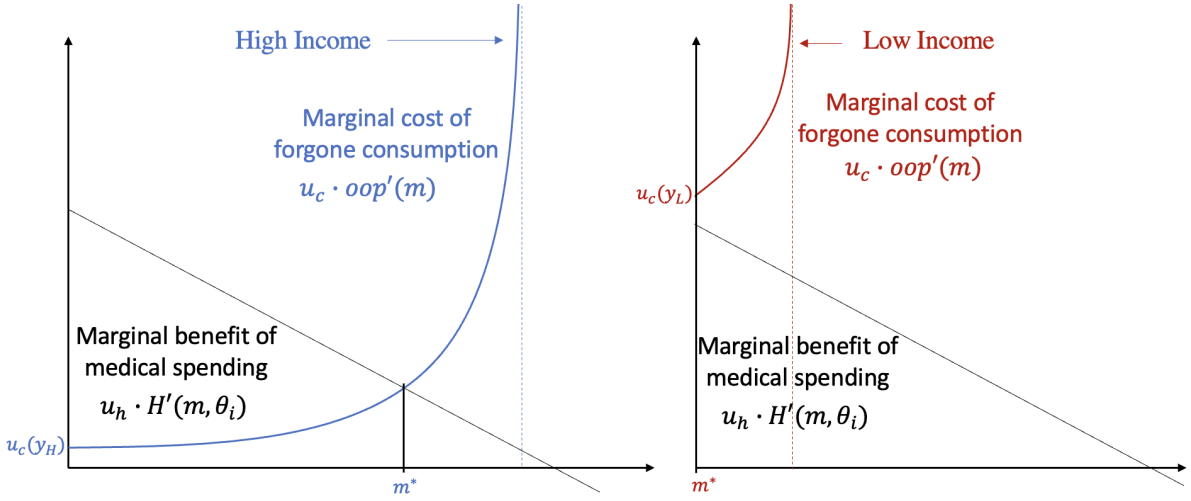
The optimal medical spending decision maximizes individual utility (1) subject to the budget constraint (2), a non-negative medical spending constraint $m \geq 0$, and a non-negative consumption constraint $c \geq 0$. The solution will either involve positive medical spending in which the agent equates marginal benefit to marginal cost, or a corner solution of zero medical spending. The Kuhn-Tucker conditions tell us that the solution will satisfy either:

$$\begin{aligned} & & m > 0 \quad & \text{and} \quad & \overbrace{u_h \cdot H'(m; \theta_i)}^{\text{marginal benefit of health care}} = \overbrace{u_c \cdot oop'(m)}^{\text{marginal cost of forgone consumption}} & \quad (3) \\ \text{or} \quad & & m = 0 \quad & \text{and} \quad & u_h \cdot H'(m; \theta_i) < \ u_c \cdot oop'(m). & \quad (4) \end{aligned}$$

The optimality conditions reveal two important insights: first, that income plays a key role in determining the consumer's medical spending tradeoff via the marginal cost of forgone consumption. If our aim is to understand differential behavior across individuals with different incomes, we need to account for differences in the marginal utility of consumption across high and low-income individuals, and moving away from quasi-linear utility is a step in the right direction. Because the costs of forgoing consumption are scaled by the marginal utility of consumption, health care will appear more costly to a low-income individual, relative to a high-income individual.

Second, the zero medical spending choice can be informative about economic primitives. To illustrate, consider a setting where the marginal benefit of care is homogeneous across individuals (e.g. insulin for diabetics), but the marginal costs of forgoing consumption are heterogeneous (due

Figure 1: Medical Spending Tradeoff for High v. Low-Income



Notes: This figure shows the marginal benefit curve of medical spending and the marginal cost of forgone consumption for a high- and low-income individual, respectively. Both individuals derive the same marginal benefits from medical spending, but differ in their levels of income and thus their costs of forgoing consumption are different. In both panels, the marginal cost of forgoing consumption $m = 0$ is equal to the marginal utility of consuming the entirety of income, and the cost asymptotes to infinity as $oop(m) \to y$. The panel on the left shows in blue the costs for a high-income individual, and the panel on the right shows in red the costs for a low-income individual.

to income and concavity differences). If we observe income and are willing to assume individuals are making rational decisions, then the share of individuals who do not use any care will be informative about the marginal cost of forgone consumption, and thus about concavity of the utility function with respect to non-medical consumption. Moreover, if extensive margin utilization is monotonic in income such that there is a single income threshold above which individuals do not use any care, then there will be a unique concavity parameter that rationalizes observed choices. This can be useful for structural work using moment inequalities.

2.2 Equilibrium medical spending for high- versus low-income individuals

Figure 1 illustrates the tradeoffs in the medical spending decision for a high- versus low-income individual. Because medical spending reduces consumption, the marginal cost of forgone consumption is exactly the reflection of the marginal utility of consumption. The costs of forgoing consumption asymptote to infinity as the out-of-pocket costs approach income (and consumption approaches zero). When medical spending is zero, the costs of forgoing consumption are equal to the marginal utility of consumption for an individual who consumes the entirety of their income. Thus, when utility is concave and the marginal benefit of care is bounded at zero, we should not expect to see low-income individuals utilizing any health care.

How does insurance (i.e. cost-sharing) change the tradeoffs for low-income individuals? In general, lowering the out-of-pocket costs will increase health care utilization along the extensive margin: the likelihood that the marginal benefit of care exceeds the marginal costs of forgone consumption will increase as out-of-pocket costs are reduced. However, the number of individuals

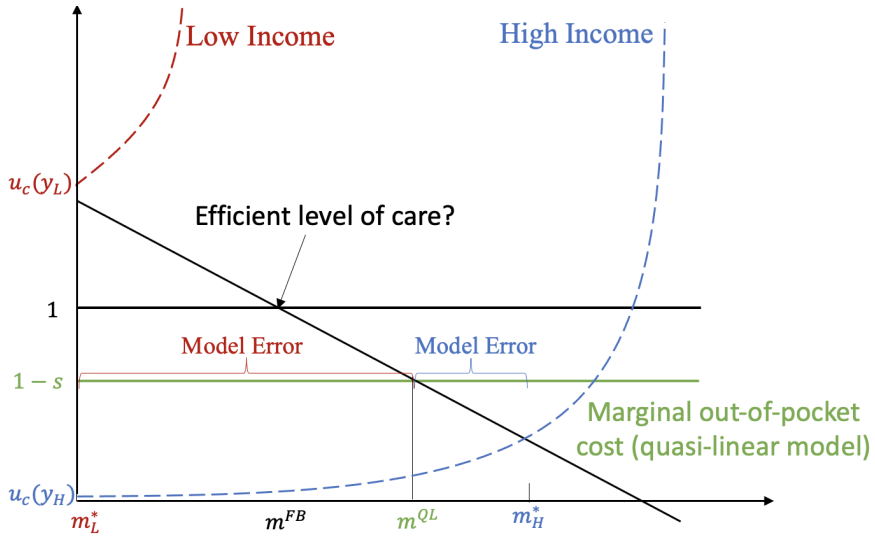
who switch from consuming no care to consuming some positive amount of care will depend on the relative magnitude of the costs of forgoing consumption. If the marginal utility is close to infinity, reductions in the out-of-pocket cost will still be scaled up by a very large number and may not change behavior at all.

2.3 Quasi-linear misspecification

Shutting down income effects implies that the marginal cost of forgone consumption is fully determined by the marginal out-of-pocket costs (regardless of whether the individual is high- or low-income). This means that for any two individuals facing the same out-of-pocket price and sickness profile, medical spending should be the same, and unexplained differences will be absorbed by some empirical error term. When utility is assumed to be quasilinear in consumption (e.g., $u(c, H(m)) = c + H(m)$), the marginal utility of consumption is equal to one for everyone, $u_c = 1$.

To illustrate, suppose the true model of utility were $u^{True}(c, H(m;\theta_i)) = v(c) + H(m;\theta_i)$ for $v(c)$ concave, and consider a misspecified quasi-linear utility model, $u^{QL}(c, H(m;\theta_i)) = c + H(m;\theta_i)$. Suppose there is some level of cost sharing so that out-of-pocket costs are $(1-s)m$ for $s \in [0,1]$. Figure 2 shows the equilibrium choices of medical spending for a low- and high- income individual, plotted against the equilibrium choice of m implied by the quasi-linear utility model.

Figure 2: Quasi-linear Misspecification



Notes: This figure shows the marginal benefit and marginal cost curves implied by the quasi-linear utility model, overlaid against the equilibrium medical spending choices for a low- and high-income individual, m_L^* and m_H^* of the same health type, θ_i. In equilibrium, each individual equates their marginal benefit of care against their marginal cost of forgoing consumption. The quasi-linear model implies that the equilibrium choice is m^{QL} for both high- and low-income individuals. If the efficient benchmark constituted the no insurance case, the efficient level of medical spending m^{FB} equates the marginal benefit of care to one, the marginal cost of care without insurance.

Figure 2 shows that low-income individuals will appear to underconsume "high-value care", and high-income individuals will appear to overconsume "low-value care," where high- (low-) refers to care for which the marginal benefit is greater (less) than the marginal cost of care implied by the

model. If the efficient benchmark constituted the case without insurance ($s = 0$), low- and high-income individuals would still appear to deviate from the efficient allocation, and ignoring income effects might lead one to conclude that either (1) the marginal benefit of care was different across these two individuals, or (2) that low-income individuals underestimated the benefits of care and that high-income individuals overestimated them.

Non-rational models of health care decisions can also explain empirical patterns of underuse or overuse of health care services (Newhouse, 2006; Gruber et al., 2020; Baicker, Mullainathan, and Schwartzstein, 2015). While one can hardly argue that behavioral frictions do not play a major role in medical spending decisions, it is important to understand the how income affects demand for health care services.

The standard approach to deal with zero medical spending involves additional modeling of how individuals select into zero consumption (Manning et al., 1987; Newhouse, 1996), but we do not need additional modeling when we interpret the zeroes as a corner solutions to the individual's optimization problem: the cost of forgoing consumption must have exceeded the marginal benefit of health care.

3 Medical spending and elasticities along the income gradient

The model of health care demand with concave consumption utility has three main empirical predictions. First, low-income individuals should be the least likely to consume any care. Second, the average response to reductions in cost-sharing should be smaller, relative to higher income counterparts, when such reductions fail to induce individuals to consume some positive amount of care. The (small) average response for low-income individuals will be driven by the inframarginal individuals for whom the marginal costs of forgoing consumption are large. Third, low-income individuals who do consume some positive amount of health care will be highly price sensitive.

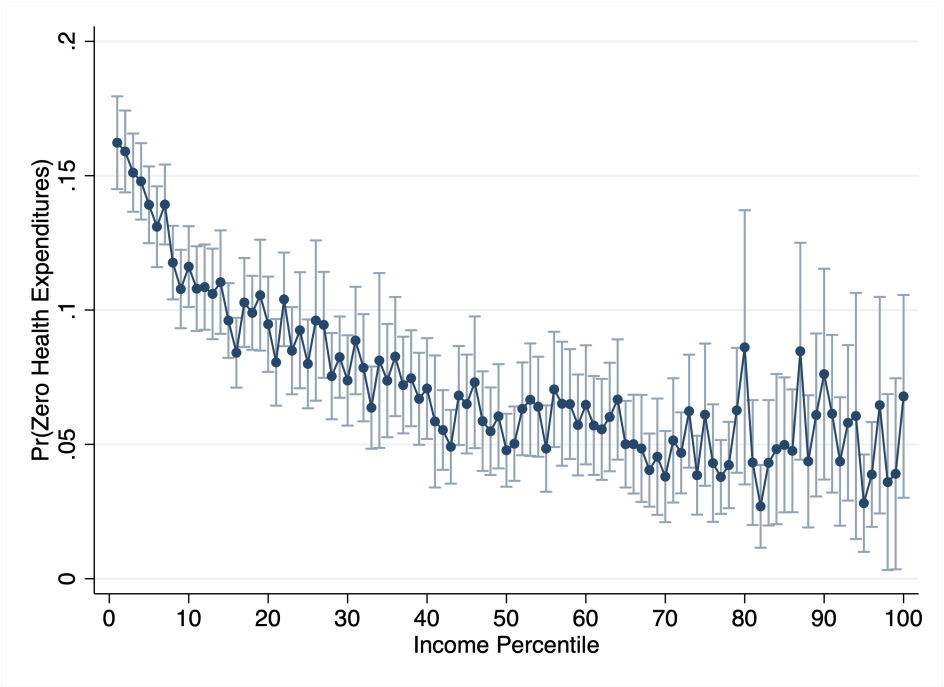
To examine the income gradient along the extensive margin to consume health care, I use data from the Medical Expenditure Panel Survey (MEPS). To study demand responses to cost-sharing by income, I use data from the RAND Health Insurance Experiment. I find that individuals at the bottom of the income distribution have a smaller average response to reductions in cost-sharing, while individuals at the middle have a significantly negative response. I estimate heterogeneous treatments effect of cost-sharing across income groups and find that the zero is driven by a disproportionate share of low-income individuals who consume zero care, independent of whether it is free or with cost-sharing.

3.1 The probability of consuming health care by income

The MEPS data are public survey data that contain information about medical expenditures and health conditions of individuals. The data consist of repeated cross-sections from 1996 to 2020, collected annually, and individuals are followed for only two years. I estimate the probability that

an individual uses some positive amount of health care services conditional on income, controlling for insurance type (private, public, or none), age, sex, race, education, chronic conditions, and impairments to physical and cognitive functioning. Standard errors are adjusted to account for survey design with non-random sampling. I assign individuals to income percentiles at the yearly level and based on their reported income in the first year. Children and retirees are dropped from the sample. Figure 3 shows the results.

Figure 3: Probability of Any Medical Spending by Income Percentile



Notes: Figure shows the predicted share of individuals who spent zero on health care over a period of two years. The probability that an individual uses some positive amount of health care services conditional on income is estimated in a regression where the dependent variable is a binary indicator for whether the individual consumed zero care, and the independent variables are 100 categorical dummies for the individual's income percentile. Regressions are estimated at the individual level and controls include insurance type (private, public, or none), age, sex, age squared, age by sex interactions, race, education, chronic conditions, and impairments to physical and cognitive functioning. Standard errors are adjusted to account for survey design with non-random sampling. N=186,895.

The main descriptive takeaway from Figure 3 is that low-income individuals are more likely than higher income individuals to consume no health care, even controlling for observable health and demographic characteristics, and type of insurance. An important caveat for interpretation is that Figure 3 describes equilibrium medical consumption, which is a product of both demand and supply side factors. Thus, while the patterns in the data are consistent with income effects in health care demand, they could also be explained by supply side constraints that are correlated with income, such as limited provide entry in low-income neighborhoods.

The RAND Health Insurance Experiment has the big advantage that it allows to study health care demand in isolation, in a setting where we do not need to worry so much about supply side constraints. The RAND investigators paid health care provider for their submitted charges, which means providers did not have an (insurer driven) incentive to undertreat or dismiss patients. More-

over, the RAND experiment was conducted across six urban areas where physical access health care providers was similar across individuals with varying incomes. Thus, I conduct all further analyses using data from the RAND Health Insurance Experiment.

The share of low-income individuals who consume no health care can help us learn about the concavity of the utility function with respect to consumption if one considers importing methods from other literature. With additional assumptions, the simple model presented in this article can connect the concavity of the utility function to share of individuals individuals who choose to consume no care. That is, echoing Chetty (2012) and following Todd and Wolpin (2023), one could consider parametrizing the utility function with a concavity parameter in consumption and health and using the RAND experiment data to recover structural parameters. This would require imposing assumptions on the marginal benefit of health care and interpreting the unobserved heterogeneity as differences in the marginal costs of forgoing consumption.

3.2 The elasticity of demand by income: revisiting the RAND experiment

The original RAND investigators were very interested in the question of whether behavioral responses to cost sharing differed across high- and low-income individuals. This was particularly salient at the time because it was a time of active Medicare policy reform in the United States, and the investigators were actively looking to test the theories of moral hazard that were underlying in the political debate (Feldstein, 1971, 1973). After the experiment and extensive analyses, the investigators found some evidence that extensive margin responses to care varied by income, but ultimately concluded that "although cost sharing in the experiment was reduced for the poor, decreases in expenditure were similar across the poor and nonpoor" (Newhouse p. 79, 1996). Due to power limitations, I am similarly unable to reject that low-income individuals have a statistically different response relative to their higher income counterparts, but my article highlights the important fact that the average response among the low-income subgroup is *near zero*.

The experiment randomly assigned individuals to health insurance plans with varying levels of cost-sharing. It ran from 1974 to 1981, and enrolled approximately 1,900 families and 5,8000 individuals. Individuals were assigned to one of six plans[4] with varying levels of cost-sharing: 95%, 50%, 25%, or 0%. The experiment ran in six different locations, and random assignment was random conditional on location and starting month. The investigators oversampled low-income families in the design: 17% of families in the experiment had incomes below the Federal Poverty Line (FPL), which was 6 percentage points greater than the national share of families below the poverty line in 1973.

The RAND Health Insurance Experiment data has the advantage of spanning a broad range of incomes: as a percent of the Federal Poverty Line (FPL), the poorest RAND family earned

[4]Six plans refers to six categories of cost-sharing which applied to either all services or sub-categories of services (e.g. dental or outpatient psychiatric care). Formally, there were twenty four unique plans which varied along two dimensions: cost-sharing and out-of-pocket maximum (referred to as maximum dollar expenditure, or MDE, limit by the experimenters).

19% FPL, while the richest family earned 11,200% FPL. In dollars and normalized to a family of four, the poorest family earned $860 while the richest family earned $25,000 in 1973 ($6,000 and $176,000 in 2023 dollars, respectively). The investigators collected income data in the two years prior to enrollment, and excluded individuals with incomes greater than $25,000 in 1973 dollars (which corresponds to approximately $175,00 in 2023 dollars). Because the experiment was carried out in six different locations, income data was collected on different years.

To discretize individuals into income terciles, I use the normalization provided by the RAND investigators: incomes where scaled by family size and adjusted for inflation for comparability to a four person household (1973 dollars). As a percentage of the Federal Poverty Line in 1973, the income cutoff for the bottom tercile was 203% FPL, for the middle tercile 336% FPL, and the top tercile had incomes ranging from 336% to 11,200% FPL. For the lowest income tercile, medical spending levels were not statistically different in the cost-sharing plans, relative to the free care plan. The middle income tercile had the largest spending differences across cost-sharing across plans, with an average spending of$2.3 thousand in the 0% coinsurance plan and $1 thousand in the 95% coinsurance plan. More detail about average spending levels across income terciles and plans can be found in Appendix Table

I estimate elasticities by income tercile using the experimental variation in cost-sharing from the RAND experiment. They key finding is that low-income individuals have an average elasticity of zero. The null result is driven by low-income individuals who consume zero medical spending, regardless of whether health care is free or costly. Given the relative spending differences across plans for low-income, it is not surprising that the estimated demand elasticity for this income group is small. I estimate pairwise elasticities as percent change in total spending, divided by the percent change in cost-sharing. Given the set plans offered, I can estimate pairwise elasticities by taking pairwise comparisons of the 0%, 25%, 50%, 95% coinsurance plans. For any two levels of cost-sharing, $(1-s)$ and $(1-s')$, the pairwise elasticity $\eta(s,s')$ is defined as:

$$\eta(s,s') \equiv \frac{\mathbb{E}[m_i|(1-s')] - \mathbb{E}[m_i|(1-s)]}{\mathbb{E}[m_i|(1-s)]} \cdot \frac{s}{(1-s)' - (1-s)} \tag{5}$$

Empirical Framework: I compute pairwise elasticities separately within each income tercile. Following the RAND investigators and Aron-Dine, Einav, and Finkelstein (2013), I use person-year as the primary unit of analysis. The average conditional spending component of the pairwise elasticity, $\mathbb{E}[m_i|(1-s)]$, is estimated based on regressing m_i on plan dummies λ_p and location by month fixed effects, clustering standard errors at the family level. Denoting individuals by i, calendar year by y, medical spending by m, the insurance plan by p, month by t, and location by l, the specification is:

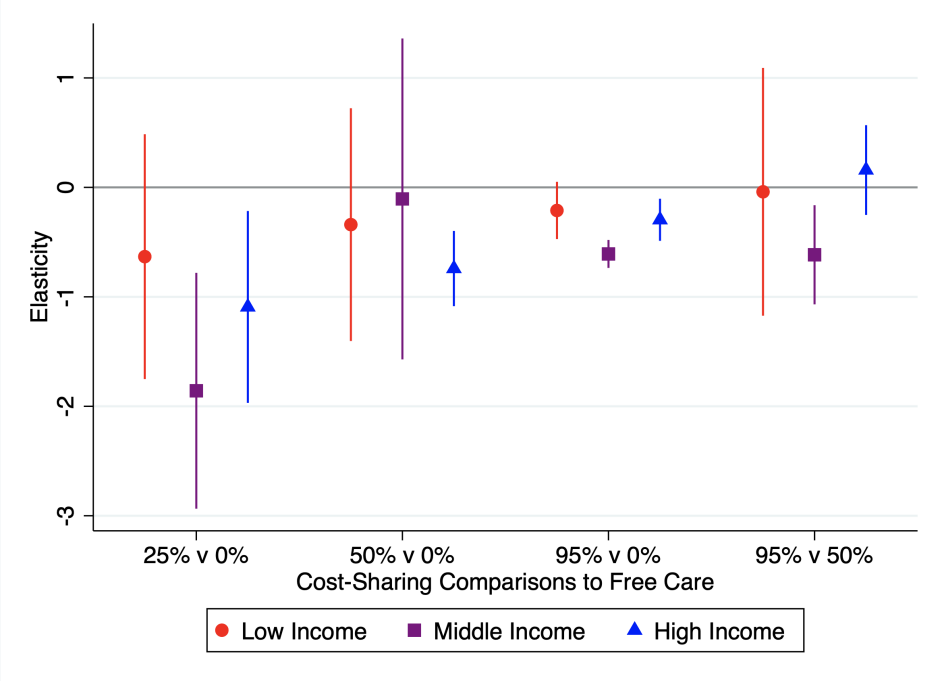
$$m_{i,y} = \lambda_p + \lambda_y + \lambda_{l,t} + \epsilon_{i,t} \tag{6}$$

Pairwise elasticities are then computed using equation (5) by taking $\lambda_p = \mathbb{E}[m_i|(1-s_p)]$, where

s_p corresponds to the coinsurance rate of plan p. Standard errors are bootstrapped based on 2,000 replications and clustered on family.

Figure 4 shows the estimated pairwise elasticities across plans and income tercile. Low-income individuals are shown in red, high-income in blue, and middle income in purple. The lowest income tercile were the least responsive to cost-sharing, and the middle income individuals were the most responsive. For the lowest income tercile, the proportional spending reduction relative to the free care plan was not statistically different from zero among any of the cost-sharing plans. The highest income tercile had a milder (significant and negative) response to cost-sharing, relative to the middle income group.

Figure 4: Pairwise Elasticities by Income Tercile



Notes: Regressions are estimated separately by income tercile. The unit of analysis is person-year and standard errors are clustered at the family level. The dependent variable is inflation-adjusted medical spending. The independent variables are RAND plan group dummies and year plus location by month fixed effects. Pairwise elasticities are computed by equation (5) using as inputs the plan dummy coefficients from the regression described by equation (6). Standard errors are bootstrapped based on 2,000 replications and clustered on family.

3.3 What's driving the small average elasticity for low-income individuals?

There were two additional implications from the theory for the demand response of low-income individuals: that the inframarginal individuals may not be induced to use any health care following reductions in coinsurance, and that conditional on utilization, low-income individuals will be quite price-sensitive. That is, we would expect the intensive margin response of low-income individuals who do move into the interior region of positive health care spending to be large.

Therefore, estimating the heterogeneous treatment effects of an exogenous change in cost-sharing among low-income individuals should look flat in the lower quantiles of the outcome

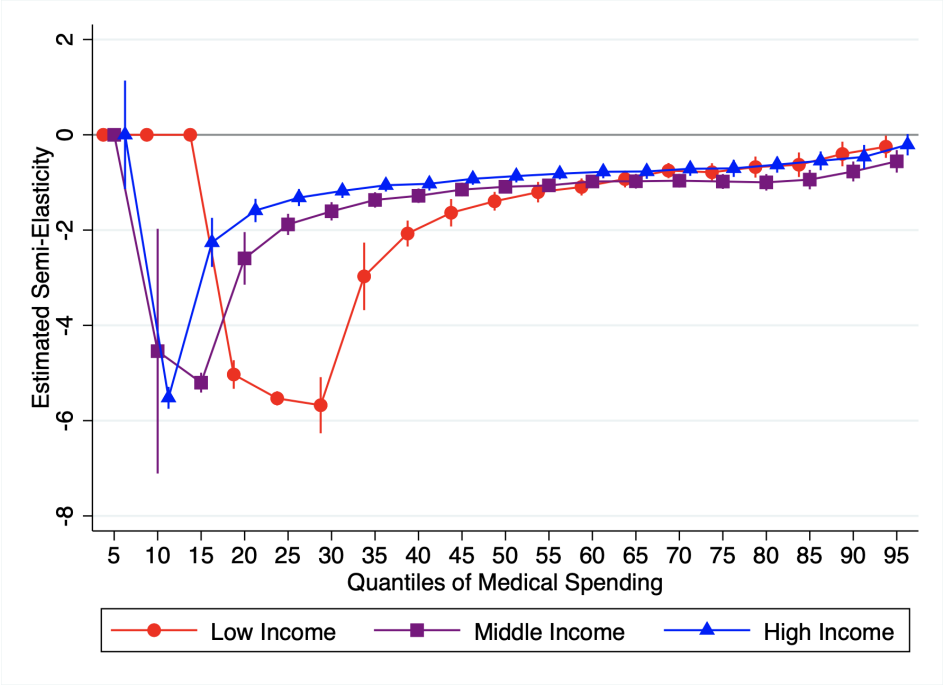
variable (medical spending), followed by a steep decline in medical spending. Leveraging random assignment in the RAND allows me to estimate the heterogenous treatment effect for different income groups.

To avoid dropping individuals assigned to the free care plan, I estimate semi-elasticities of medical spending with respect to coinsurance for each income group. Within each income tercile, I regress log-spending on the coinsurance level and year plus location by month fixed effects, clustering at the family level. Denoting individuals by i, calendar year by y, medical spending by m, the insurance plan by p, the coinsurance rate for that plan by $(1 - s_p)$, month by t, and location by l, the empirical model is:

$$\log(m_{i,y}|tercile) = \beta(1 - s_p) + \lambda_y + \lambda_{l,t} + \epsilon_{i,t} \tag{7}$$

where λ_y, and $\lambda_{l,t}$ denote calendar year, and location by start month fixed effects. Results are shown in Figure 5.

Figure 5: Decomposition of Semi-Elasticities by Income Tercile



Notes: Quantile regressions are estimated separately within each income tercile, block bootstrapped at the family level based on 500 replications. The dependent variable in each of these quantile regressions is log(1+medical spending). The independent variables are the continuous coinsurance rate associated with the RAND plan (e.g. sticker price), and year plus location by month fixed effects.

Figure 5 shows the semi-elasticity of medical spending with respect to coinsurance, estimated at different quantiles of the outcome variable (medical spending). There are three takeaways about the observed behavior of low-income individuals. First, the estimated zeros at the bottom quantiles of medical spending tell us that there are a lot of inframarginal individuals with zero medical spending, *and whose zero spending is unaffected* by cost-sharing reductions. Second, individuals that do move

into the interior region (positive medical spending) are as price sensitive as their higher income counterparts. This is captured by the steep negative semi-elasticity estimated off of the 15th and 30th medical spending percentiles. Third, at the highest medical spending quantiles, the semi-elasticity becomes almost zero: the very sick individuals will use care regardless of cost-sharing.

Internal validity and additional plan features: One potential concern with the specification in equation (7) might be that estimates of the semi-elasticity could be confounded with additional plan features–the RAND investigators designed plans that varied along two dimensions: cost-sharing and the out-of-pocket expenditure limit. The latter varied with income for a subset of participants,[5] which implies that, for the subgroup of individuals with income-dependent maxima, lower income individuals had lower out-of-pocket limits than their higher income counterparts. While this plan design feature would bias intensive margin elasticities conditional on utilization towards zero, it cannot explain differences in extensive margin utilization by income.

Since the primary objective of this analysis to document that low-income individuals are more likely to choose the corner solution of zero medical spending, my analysis focuses on the extensive margin component of utilization and how it affects estimates of the intensive margin response to cost-sharing. The main takeaway from Figure 5 is that estimates of the average intensive margin response to cost-sharing are attenuated by the extensive margin share of low-income individuals who choose to consume zero health care under both positive and zero cost-sharing. Under alternative specifications that address the non-linear feature of the health insurance contract,[6] the attenuation result still holds among low-income individuals. I point the reader to the Appendix for further evidence and discussion of these alternative specifications.

4 Discussion

In this fairly general model, I show that the marginal cost for forgone consumption is an important aspect of the medical spending decisions when we are trying to understand behavior along the income gradient. I empirically show that low-income individuals are not very responsive to cost-sharing in the RAND experiment, on average, but those who are induced to consume care are price sensitive.

Taken together, this analysis tells a story that is consistent with income effects being present in health care demand, and raises a point of caution against the quasi-linear utility model of utility. The key insight from the theory is that forgoing consumption is very costly to low-income individuals because their marginal utility of non-medical consumption is high. Thus, low-income individuals may not consume any care, *even* in the presence of large subsidies, and their elasticity of medical spending (with respect to out-of-pocket costs) will appear low, on average. A mis-specified model

[5] Individuals got randomly assigned to out-of-pocket expenditure limits of either 5, 10, or 15% of income, or $1,000 (whichever amount was smaller).

[6] Prior literature has addressed the concern that plan cost-sharing did not fully capture average or marginal out-of-pocket costs by using the average out-of-pocket share across the plan or instrumenting for the end-of-year price using plan assignment (Aron-Dine, Einav, and Finkelstein, 2013; Lin and Sacks, 2019).

without income effects would lead us to interpret the empirical patterns as evidence that either low-income individuals do not understand the benefits of care or that the benefits are small.

Formally accounting for income effects in empirical work about health care demand presents two foreseeable challenges. First, health insurance claims data are sensitive and data privacy rules often imply that income information is not available available at the individual level. However, geographic information could be used as an (imperfect) proxy for socioeconomic status. Second, accounting for income effects requires making modeling assumptions about the concavity of the utility function with respect to consumption (which determines the marginal utility of consumption). While there is a vast literature in across finance, public, and macroeconomics dedicated to estimating the coefficients of relative and absolute risk aversions, lack of consensus across estimates and structural assumptions introduce an additional dimension of model mis-specification that would certainly bias demand estimates.

Empirical work generally accounts for income effects within the context of health *insurance* demand, but largely ignores them in the context of health *care* demand. To my knowledge, Finkelstein, Hendren, and Luttmer (2019) are the first to incorporate empirically the idea that the marginal utility of consumption is higher for low-income individuals: they adjust the marginal rate of substitution between health care and consumption. If income effects are present in the choice of insurance, then to maintain internal consistency they must also be accounted for in the intensive and extensive margin choice of health care consumption. Finkelstein, Hendren, and Luttmer (2019) take significant steps towards this goal and provide a method to account for income effects that can be used in future empirical work.

References

Acemoglu, Daron, Amy Finkelstein, and Matthew J. Notowidigdo, "Income and Health Spending: Evidence from Oil Price Shocks," *MIT Press*, October 2013. Accepted: 2014-01-13T17:26:30Z Publisher: MIT Press.

Aron-Dine, Aviva, Liran Einav, and Amy Finkelstein, "The RAND Health Insurance Experiment, Three Decades Later," *Journal of Economic Perspectives*, February 2013, *27* (1), 197–222.

Baicker, Katherine, Sendhil Mullainathan, and Joshua Schwartzstein, "Behavioral Hazard in Health Insurance *," *The Quarterly Journal of Economics*, November 2015, *130* (4), 1623–1667.

Bhargava, Saurabh, George Loewenstein, and Justin Sydnor, "Choose to Lose: Health Plan Choices from a Menu with Dominated Option*," *The Quarterly Journal of Economics*, August 2017, *132* (3), 1319–1372.

Bhattacharya, Jay and Darius Lakdawalla, "Does Medicare benefit the poor?," *Journal of Public Economics*, 2006, *90* (1-2), 277–292. Publisher: Elsevier.

Brot-Goldberg, Zarek C., Amitabh Chandra, Benjamin R. Handel, and Jonathan T. Kolstad, "What does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics*," *The Quarterly Journal of Economics*, August 2017, *132* (3), 1261–1318.

Bureau, US Census, "Characteristics of the Low-Income Population: 1973." Section: Government.

Chandra, Amitabh, Jonathan Gruber, and Robin McKnight, "The impact of patient cost-sharing on low-income populations: Evidence from Massachusetts," *Journal of Health Economics*, January 2014, *33*, 57–66.

Chetty, Raj, "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply," *Econometrica*, 2012, *80* (3), 969–1018. Publisher: [Wiley, Econometric Society].

Cutler, David M. and Sarah J. Reber, "Paying for Health Insurance: The Trade-Off between Competition and Adverse Selection*," *The Quarterly Journal of Economics*, May 1998, *113* (2), 433–466.

Einav, Liran and Amy Finkelstein, "Moral Hazard in Health Insurance: What We Know and How We Know It," *Journal of the European Economic Association*, August 2018, *16* (4), 957–982.

Feldstein, Martin, "A new approach to national health insurance," *The Public Interest.*, 1971.

Feldstein, Martin S., "The Welfare Loss of Excess Health Insurance," *Journal of Political Economy*, 1973, *81* (2), 251–280. Publisher: University of Chicago Press.

Finkelstein, Amy, Nathaniel Hendren, and Erzo F. P. Luttmer, "The Value of Medicaid: Interpreting Results from the Oregon Health Insurance Experiment," *Journal of Political Economy*, December 2019. Publisher: The University of Chicago PressChicago, IL.

_ , _ , and Mark Shepard, "Subsidizing Health Insurance for Low-Income Adults: Evidence from Massachusetts," *American Economic Review*, April 2019, *109* (4), 1530–1567.

Gornick, M. E., P. W. Eggers, T. W. Reilly, R. M. Mentnech, L. K. Fitterman, L. E. Kucken, and B. C. Vladeck, "Effects of race and income on mortality and use of services among Medicare beneficiaries," *The New England Journal of Medicine*, September 1996, *335* (11), 791–799.

Gruber, Jonathan, Johanna Catherine Maclean, Bill Wright, Eric Wilkinson, and Kevin G. Volpp, "The effect of increased cost-sharing on low-value service use," *Health Economics*, October 2020, *29* (10), 1180–1201.

Krueger, Alan B. and Ilyana Kuziemko, "The demand for health insurance among uninsured Americans: Results of a survey experiment and implications for policy," *Journal of Health Economics*, September 2013, *32* (5), 780–793.

Lavetti, Kurt, Thomas DeLeire, and Nicolas R. Ziebarth, "How do lowincome enrollees in the Affordable Care Act marketplaces respond to costsharing?," *Journal of Risk and Insurance*, June 2022, *90* (1), 155–183.

Lin, Haizhen and Daniel W. Sacks, "Intertemporal substitution in health care demand: Evidence from the RAND Health Insurance Experiment," *Journal of Public Economics*, July 2019, *175*, 29–43.

Link, Charles R., Stephen H. Long, and Russell F. Settle, "Equity and the Utilization of Health Care Services by the Medicare Elderly," *The Journal of Human Resources*, 1982, *17* (2), 195–212. Publisher: [University of Wisconsin Press, Board of Regents of the University of Wisconsin System].

Manning, Willard G., Joseph P. Newhouse, Naihua Duan, Emmett B. Keeler, and Arleen Leibowitz, "Health Insurance and the Demand for Medical Care: Evidence from a Randomized Experiment," *The American Economic Review*, 1987, *77* (3), 251–277. Publisher: American Economic Association.

McClellan, Mark and Jonathan Skinner, "The incidence of Medicare," *Journal of Public Economics*, January 2006, *90* (1), 257–276.

Newhouse, Joseph P., *Free for All?*, Harvard University Press, February 1996.

Nyman, John A., *The Theory of Demand for Health Insurance*, Stanford: Stanford University Press, 2003.

Shepard, Mark, Katherine Baicker, and Jonathan Skinner, "Does One Medicare Fit All? The Economics of Uniform Health Insurance Benefits," *Tax Policy and the Economy*, May 2020, *34*, 1–41. Publisher: The University of Chicago Press.

Todd, Petra E. and Kenneth I. Wolpin, "The Best of Both Worlds: Combining Randomized Controlled Trials with Structural Modeling," *Journal of Economic Literature*, March 2023, *61* (1), 41–85.

A Empirical Appendix

A.1 Summary statistics from the RAND experiment data

Table 1 shows average spending levels across income terciles and across plans. For the lowest income tercile, medical spending levels were not statistically different in the cost-sharing plans, relative to the free care plan. The middle income tercile appears to have had the largest spending differences across cost-sharing across plans, with an average spending of $2.3 thousand in the 0% coinsurance plan and $1 thousand in the 95% coinsurance plan.

Table 1: Average Spending by Income and Plan

		Income Tercile		
	Overall	(1)	(2)	(3)
	Share	Average Spending (1973 $)		
0% Coinsurance	.34	$1,819	$2,294	$ 2,489
(Free Care)		(4,421)	(5,565)	(5,178)
25% Coinsurance	.12	1,392	1,488	1,806
		(3,764)	(5,321)	(4,004)
50% Coinsurance	.07	1,445	2,052	1,450
		(9,143)	(14,841)	(3,028)
95% Coinsurance	.18	1,326	1,008	1,756
		(4,916)	(2,525)	(4,849)
Mixed Coinsurance	.08	1,763	1,496	2,494
		(5,434)	(4,595)	(7,015)
Individual Deductible	.21	1,151	2,002	1,967
		(3,180)	(6,629)	(4,821)
Number of families	1,908	636	636	636

Notes: Table shows the shares of individuals enrolled across plans, and average individual spending by income tercile within each plan. Individuals were assigned to terciles at the family level. As a percentage of the Federal Poverty Line in 1973, the income cutoff for the bottom tercile was 203% FPL, for the middle tercile 336% FPL, and the top tercile had incomes ranging from 336% to 11,200% FPL.

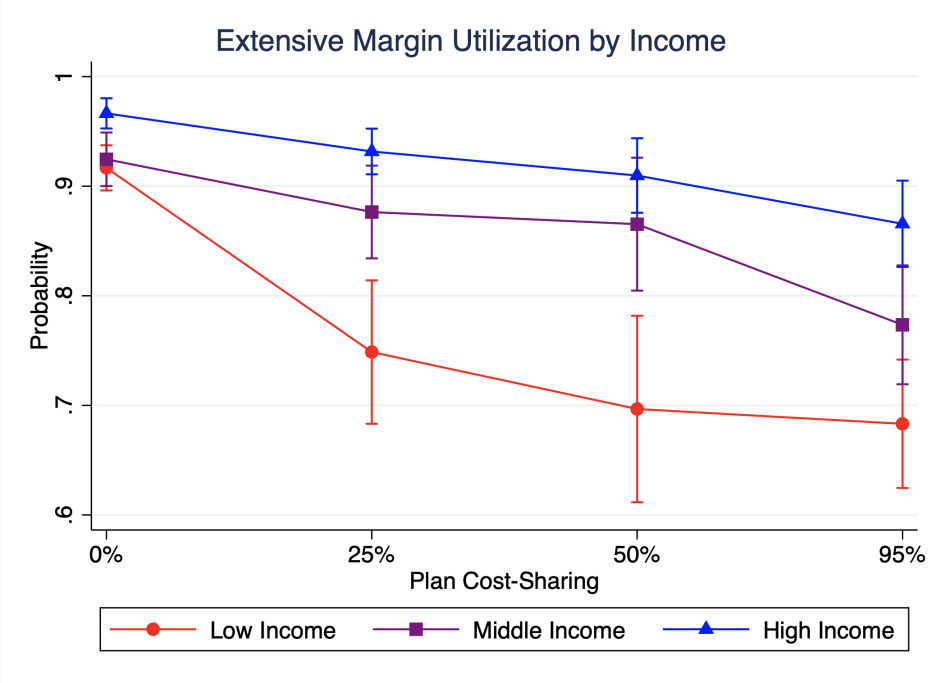
A.2 Alternative specifications: semi-elasticity of health care demand with respect to coinsurance

The following section contains additional empirical specifications that address concerns about potential confounders to empirical results. In particular, the RAND experimenters designed health insurance plans that varied over two dimensions: cost-sharing and out-of-pocket expenditure limits. The out-of-pocket expenditure limits varied with income for a subset of participants: individuals got randomly assigned to out-of-pocket expenditure limits of either 5, 10, or 15% of income,

or $1,000 (whichever amount was smaller). Taking into account both plan dimensions, here were twenty four unique plans offered during the experiment.

Because the out-of-pocket expenditure limits were a function of income for a randomly selected subset of participants, lower income individuals had lower out-of-pocket limits than their higher income counterparts among this subset. This implies that, conditional on being low-income, the differences between the plans with cost-sharing versus the free care plan were smaller, which would bias estimates of the pairwise spending comparisons to the free care plan towards zero. While this plan design feature would bias intensive margin elasticities *conditional on utilization* towards zero, it cannot explain differences in *extensive margin utilization* by income. If anything, lower-income individuals had an incentive to spend up to their (lower) out-of-pocket limit in order to reach the zero cost-sharing region. But in the RAND data, lower-income individuals were less likely to use any health care than their higher-income counterparts. This fact can be seen in Figure A.1, and was also extensively documented by the original RAND investigators documented (Newhouse, 1996).

Figure A.1: Probability of Any Medical Spending by RAND Plan and Income



Notes: Figure shows the predicted share of individuals who spent zero on health care over the RAND experiment period. The probability that an individual uses some positive amount of health care services conditional on income is estimated in a regression where the dependent variable is a binary indicator for whether the individual consumed non-zero care, and the independent variable are categorical dummies for the individual's cost-sharing plan assignment. Regressions are estimated at the person-year level and controls include location by month fixed effects. Standard errors are clustered at the family level. N= 13,323.

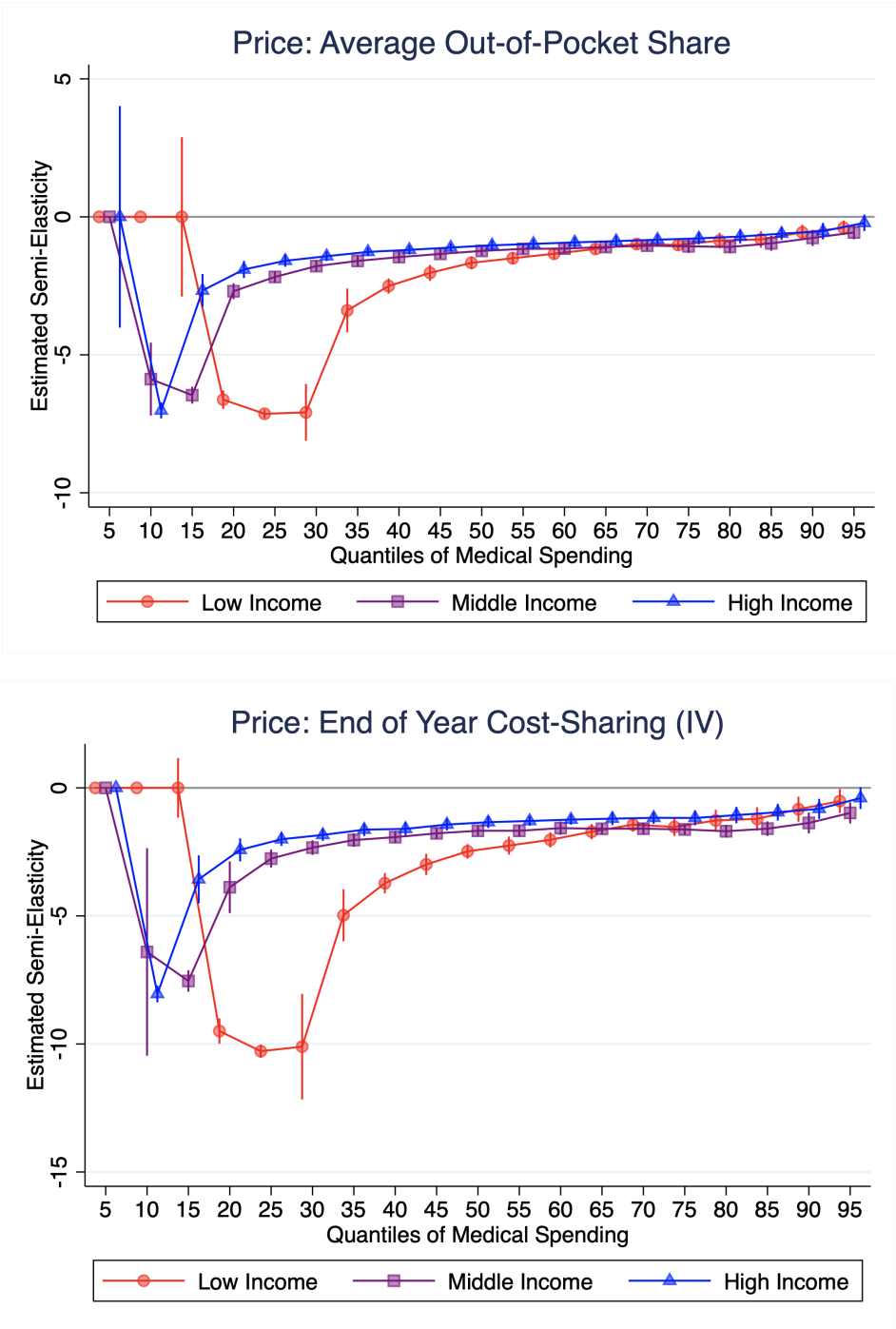
The intensive margin response to cost-sharing has been extensively examined in the literature. Aron-Dine, Einav, and Finkelstein (2013) carefully reexamine the validity of the experimental treatment effects in the RAND experiment and pay special attention to the challenges raised by the non-linear feature of health insurance contracts for estimating the demand elasticity of health care with respect to "price." When health insurance contracts are non-linear, the researcher has to make as-

sumptions about which "price" individuals respond to, which could include the current "spot" price of care, the expected end-of-year price, the actual end-of-year price, or some weighted average of prices paid over the year.

Under the assumption that price is either the actual end-of-year price, or some weighted average of prices paid over the year, random assignment may be invalid for low-income individuals. This is because lower-income individuals were systematically more likely to have lower "prices" than their higher income counterparts. Precisely because of this concern, I do not use income as a continuous measure interacted with prices, nor do I conduct formal testing of coefficients across income groups. Rather, I divide the RAND sample of individuals into three discrete subgroups based on income levels observed at time of enrollment which are held fixed throughout the analysis. Then I present descriptive evidence of differences among these subgroups that is consistent with the narrative that low-income individuals consume health case at the corner solution of zero.

Among each the three income-designated subgroups, the same concerns for validity in estimates of the demand elasticity raised in Aron-Dine, Einav, and Finkelstein (2013) apply. Methods to address these concerns are only impaired in that they are employed on smaller samples, and thus will have less precision power. Nonetheless, I argue that the quantile regressions are quite informative. I re-run the quantile regression figures using two different prices on the right hand side: (1) average out-of-pocket share within each plan *and income tercile* following Aron-Dine, Einav, and Finkelstein (2013); and (2) instrument for actual end-of-year price using plan assignment following Lin and Sacks (2019). Results are presented in Figure A.2 below.

Figure A.2: Decomposition of Alternative Estimates of Semi-Elasticities by Income Tercile



Notes: Quantile regressions are estimated separately within each income tercile, block bootstrapped at the family level based on 100 replications. The dependent variable in each of these quantile regressions is log(1+medical spending). The independent variables in the top figure consist of the average out-of-pocket share faced by all individuals within the RAND plan-income tercile, and year plus location by month fixed effects. The bottom figure presents the two-stage least squares estimates where the independent variables consist of the predicted end-of-year price (estimated by regressing end-of-year price on plan dummies and year plus location by month fixed effects), and year plus location by month fixed effects.