Myopia, Price Perception, and Moral Hazard in Health Insurance

Iris SooJin Park

August 14, 2026

Click here for the most recent version.

Abstract

I show that spending responses to cost sharing in health insurance depend not only on contractual incentives but also on household myopia and price perception. Using administrative claims data from a large employer plan, I study a natural experiment that introduced a deductible and raised the out-of-pocket maximum. Households on average reduce spending modestly, but the reduction rises steeply with baseline health risk and is concentrated among the highest-risk households. Because high-risk households should readily anticipate reaching the out-of-pocket maximum under either contract, a structurally estimated forward-looking model predicts that they respond least, contrary to the observed risk gradient. To resolve this puzzle, I develop and estimate the first structural model of moral hazard that jointly incorporates household myopia and imperfect spot-price perception. In the model, myopic households respond to spot prices inferred from medical bills, with two frictions biasing perceived spot prices upward: (1) bills arrive with delays, causing price beliefs to lag behind actual prices, and (2) households may infer the price of subsequent care from the average price on a bill rather than the marginal price. I provide evidence for both mechanisms using administrative data and an incentivized survey experiment. The estimated model matches the modest aggregate response and captures the steep risk gradient that the forward-looking model cannot. More broadly, the model provides a framework for understanding spending responses under nonlinear health insurance contracts when households learn their spot prices gradually and imperfectly.

Keywords: health insurance, insurance demand, cost sharing, price perception

1 Introduction

Controlling high health care spending while protecting households from medical expenditure risk is a central challenge in health insurance design. A large economic literature evaluates this tradeoff by modeling how households respond to the nonlinear prices created by health insurance contracts (Keeler et al., 1977; Ellis, 1986; Cardon and Hendel, 2001; Einav et al., 2013; Kowalski, 2015; Einav et al., 2015; Ho and Lee, 2023; Diaz-Campo, 2022). These models typically assume that households fully understand the contract and make forward-looking utilization decisions.

Yet a growing literature documents household responses to spot prices that are consistent with myopia (Aron-Dine et al., 2015; Brot-Goldberg et al., 2017; Guo and Zhang, 2019; Dalton et al., 2020; Simonsen et al., 2021). Related evidence that medical bills update beliefs about out-of-pocket spending and affect subsequent utilization points to an additional role for imperfect spot-price perception (Hoagland et al., 2022). Understanding and modeling how households perceive and respond to spot prices is therefore important for interpreting moral hazard and evaluating the tradeoff between controlling spending and protecting households from financial risk.

In this paper, I provide new administrative and experimental evidence on spot-price perception and develop a structural model of ex post moral hazard that jointly incorporates myopia and imperfect spot-price perception. I focus on why high-risk households cut spending the most when deductibles rise even though their expected end-of-year marginal price changes little, a pattern documented by Brot-Goldberg et al. (2017) that also emerges in my analysis. Myopia can provide part of the answer because myopic households respond to spot prices rather than the price they will face once spending accumulates. I further show that myopic spending responses depend on how quickly and accurately households learn their spot prices. Using administrative claims, I document substantial billing delays and incorporate the resulting gap between true and perceived cost-sharing positions into the model. I supplement the administrative evidence with an incentivized survey experiment to examine how respondents infer the marginal price of subsequent care from explanations of benefits and document systematic reliance on average rather than marginal prices. Incorporating delayed and biased price perceptions allows the model to match the aggregate spending response and account for much of its steep gradient across health risk.

I examine household responses to cost sharing using administrative data from the State of Wisconsin Group Health Insurance Program, a large employer plan covering state and university employees and their dependents. In 2016, the program introduced a \$250 deductible into a previously zero-deductible plan and increased the out-of-pocket maximum from \$500 to \$1,250, while holding the coinsurance rate fixed. The spending response exhibits two central empirical patterns: (1) a modest average reduction, consistent with the modest change in cost sharing, and (2) a steep risk gradient, with the largest reductions among the highest-risk households. Despite the much smaller increase in cost sharing, the risk gradient closely resembles the pattern documented by Brot-Goldberg et al. (2017), who study a required transition from a plan with no out-of-pocket cost sharing to a high-deductible plan and likewise find the largest spending reductions among predictably sick households. The recurrence of this pattern across reforms of very different magnitudes means that a model must explain not only the average spending reduction but also why reductions increase with health risk.

I begin by evaluating whether a standard forward-looking model can explain both features of the spending response. Following the structural modeling approaches developed by Cardon and Hendel (2001), Einav et al. (2013), and Ho and Lee (2023), I estimate this benchmark model on the Wisconsin data using simulated method of moments, asking it to match both the average response and its variation across health risk. The model reproduces the modest aggregate reduction in spending, but requires an extremely large moral-hazard parameter, implying that health care spending under full insurance would be more than eight times that under no insurance. Even with this degree of price responsiveness, the model cannot match the gradient across health risk, and this mismatch persists in a more flexible specification that allows the moral-hazard parameter to vary with health risk. High-risk households expect to reach the out-of-pocket maximum under either contract, leaving their end-of-year marginal price at zero. The model therefore predicts the smallest response among the households that respond most in the data.

Next, I develop a dynamic model of moral hazard under myopia by adapting the standard forward-looking framework to repeated medical-spending decisions throughout the year. At each point of care, households respond to the spot price they face without internalizing how current spending lowers future prices. This structure allows high-risk households to remain responsive to cost sharing before they reach the out-of-pocket maximum, even if they are likely to reach it later in the year.

Within this framework, however, spending decisions depend on the spot price households perceive, which may differ from the price implied by their true position on the cost-sharing schedule. The first mechanism I examine is delayed claims processing, which affects when recent care is reflected in households' perceived cost-sharing positions. Using administrative claims data, I show that billing delays are substantial and difficult to predict even among similar claims. About 20 percent of claims are not processed and paid by the health plan until more than twelve weeks after care. I also find that average billing delays increase with health risk, from 38 days in the lowest-risk quartile to 61 days in the highest-risk quartile. Consistent with bills updating perceived prices, utilization falls after a bill indicates that subsequent care remains subject to the deductible, whereas the response is much smaller after the deductible has been exhausted. Together, these findings suggest that substantial and difficult-to-predict billing delays create gaps between true and perceived cost-sharing positions and that these delays are most pronounced among higher-risk households.

Billing delay determines which claims are reflected in the price a household perceives, but it leaves open how households interpret the bills they have received. An explanation of benefits reports the total charge and total patient responsibility. This information makes the average price paid on the claim, measured as the ratio of patient responsibility to the claim amount, salient while leaving the marginal price of subsequent care unstated. When a claim crosses the deductible or out-of-pocket maximum, however, the average price share exceeds the marginal price of subsequent care, so relying on the average would lead households to perceive the spot price as higher than it is. This source of upward bias is especially relevant for high-risk households, whose claims are more likely to cross these thresholds. To measure the prevalence of average-price reasoning, I conduct an incentivized survey experiment in which respondents interpret explanation-of-benefits statements and report what they expect to pay for a subsequent visit. A finite-mixture model estimates that 16 to 18 percent of respondents are average-price responders, and I use these estimates to set the shares of price-perception types in the structural model.

I then use simulated method of moments to estimate a structural model of moral hazard that incorporates myopia, claims-processing delays, and biases in spot-price perception. Even with price responsiveness held constant across households, the model matches the aggregate spending reduction and generates larger reductions among higher-risk households, though it understates the steepness of the gradient at the top. I also estimate a more flexible specification in which price responsiveness varies with health risk, as in the forward-looking benchmark. Under this specification, the model accounts for much of the especially large spending reduction among the highest-risk households. Compared with the forward-looking benchmark, the model implies a substantially more moderate degree of underlying moral hazard, with a move from no insurance to full insurance increasing desired medical spending by less than 8 percent. More broadly, the results suggest that responses to nonlinear cost sharing depend not only on the incentives embedded in the contract but also on how quickly and accurately households learn the spot prices they face.

The main contribution of this paper is to explain why high-risk households respond strongly to cost-sharing changes even when standard forward-looking models predict that they should respond little. The modeling contribution is to incorporate myopia and imperfect spot-price perception into the workhorse structural framework of health insurance demand developed by Cardon and Hendel (2001), Einav et al. (2013), and Ho and Lee (2023). This framework and closely related models are widely used to estimate moral hazard and evaluate health insurance coverage and plan design (Marone and Sabety, 2022; Hoagland et al., 2022; Ho and Lee, 2023; Diaz-Campo, 2022). The closest work on myopic responses is Dalton et al. (2020). They identify myopia among Medicare Part D enrollees who are likely to enter the coverage gap, finding that spending remains flat before entry but falls sharply once the higher spot price applies. My model shares their emphasis on spot-price responses and embeds myopia in the workhorse framework for comprehensive medical spending under a conventional nonlinear health insurance contract. In my model, myopia generates the direction of the risk gradient, and delays and biases in spot-price perception help account for its magnitude. To my knowledge, this is the first paper to jointly model myopic spot-price behavior and imperfect spot-price perception under a conventional nonlinear health insurance contract.

A second contribution concerns how households form perceptions of spot prices under nonlinear cost sharing. Most closely related, Hoagland et al. (2022) show that delayed medical bills update beliefs about out-of-pocket spending and affect subsequent utilization. They incorporate this learning process into a model of imperfect moral hazard, with underlying price responsiveness calibrated from prior estimates. I instead use the cost-sharing reform to estimate underlying price responsiveness and assess whether imperfect-spot price perception can explain spending responses across health risk. The model represents medical spending as a sequence of discrete claim arrivals and links perceived spot prices to both claim processing and bill interpretation. More broadly, the paper contributes to the literature on responses to nonlinear price schedules, which shows that consumers may misperceive marginal incentives at kinks and thresholds (Ito, 2014; Rees-Jones and Taubinsky, 2020; Gelber et al., 2025). The incentivized survey experiment provides a new design that uses explanations of benefits to measure whether consumers infer the price of subsequent care from average or marginal cost shares.

These findings have implications for both the design and evaluation of cost sharing. Standard analyses treat an insurance contract as a mapping from accumulated spending into current and future prices. In practice, households face a second mapping from processed claims and billing statements into a perceived position and price. When claims are delayed or statements make average price shares more salient than marginal prices, the incentive households perceive can differ from the incentive written into the contract. Because billing delays and threshold crossings vary systematically with health risk, these discrepancies can change not only the aggregate spending response but also which households respond most. Policy evaluations based only on contractual prices may therefore misstate the distributional effects of a reform, while faster claims processing and clearer reporting of subsequent prices could bring perceived incentives closer to those intended by plan designers. Accounting for these frictions can therefore change how we evaluate the central tradeoff in insurance design between controlling spending and protecting households from financial risk.

The remainder of the paper proceeds as follows. Section 2 describes the Wisconsin program, the contract change, and the data. Section 3 documents the spending response and its gradient across health risk. Section 4 presents and estimates the perfect-foresight benchmark. Section 5 introduces and estimates the myopic model of spot-price decisions. Section 6 uses administrative billing data and an explanation-of-benefits experiment to examine how the timing and interpretation of medical bills shape perceived spot prices. Section 7 combines these mechanisms in the perceived-price model and evaluates spending responses across health risk. Section 8 concludes.

2 Empirical Setting

The paper studies responses to cost-sharing changes using a natural experiment from the State of Wisconsin Group Health Insurance Program, which provides state and university employees and their dependents with fully insured coverage through several insurers. Employees choose among insurers with different provider networks, but the financial features of coverage are standardized, and the cost-sharing change I study applied uniformly across plans. Employees are pooled for premium purposes, so premiums do not vary with expected medical costs and differ only by coverage tier and dental benefits.

2.1 Plan Change

Prior to 2016, the program's traditional non-high-deductible health plan (non-HDHP) carried no deductible and only modest cost sharing, with 10 percent coinsurance and out-of-pocket maximums of \$500 for single and \$1,000 for family coverage. In 2016, the state introduced a deductible of \$250 for single coverage and \$500 for family coverage into the non-HDHP, with out-of-pocket maximums rising to $1,250 for single and $2,500 for family coverage,[1] while coinsurance remained unchanged at 10 percent. Table 1 reports the full details of the plan design changes. A high-deductible health plan (HDHP) was introduced in 2015 but attracted less than 1 percent of enrollees, and more than 96 percent of households continued to enroll in the non-HDHP through 2016 (Appendix Table A.1)

Table 1: Non-HDHP Plan Design Changes, 2015-2016

Feature	**2015**	**2016**
Annual Deductible		
Single	$0	$250
Family	$0	$500
Annual Maximum Out-of-Pocket (OOP)		
Single	$500	$1,250
Family	$1,000	$2,500
Coinsurance Rate	10%	10%
Premium		
Single	$92	$86
Family	$230	$217
Copay		
Primary care visit	10%	$15
Specialist visit	10%	$25

Note: The table summarizes changes in the State of Wisconsin employees' non-HDHP plan design between 2015 and 2016. Deductibles and out-of-pocket maximums are annual and shown for single and family coverage. Premiums reflect the monthly employee premium contribution.

2.2 Data

The data consist of de-identified enrollment and claims records for all State of Wisconsin employees and their dependents enrolled in the program from 2015 through 2017. On the enrollment side, I observe each member's plan name, coverage tier (single or family), insurer, and month enrolled in each plan year. On the claims side, I observe the total allowed medical cost and the amount paid out-of-pocket under the applicable cost-sharing rules, along with

[1] Family coverage also included embedded individual cost-sharing limits. In 2015, each member faced an individual out-of-pocket maximum of $500. In 2016, each member faced an individual deductible of $250 and an individual out-of-pocket maximum of $1,250.

individual demographics including age, gender, and relationship to the policyholder. Each daily claim also records the provider code, procedure code, diagnosis code, clinical condition, and paid date.[2]

The data also include a prospective Diagnostic Cost Group (DCG) risk score, a measure of predicted annual spending based on demographics and diagnoses recorded in the prior year, including chronic conditions. Because it uses only information available before the current plan year, the score is predetermined with respect to current-year outcomes. The score is unavailable for new enrollees without a prior-year record, so I restrict the sample to individuals with non-missing risk score .

For the primary analysis, I restrict the sample to individuals continuously enrolled in the non-HDHP between 2015 and 2017 and families with a constant number of covered members over the same period. The resulting sample contains 86,987 individuals from 34,036 households, and the analysis is conducted at the individual level. These restrictions ensure that observed spending changes reflect responses to the reform rather than mechanical changes due to plan switching or changes in the number of covered family members. As shown in Table 2, these restrictions reduce the sample by roughly half, excluding enrollees whose coverage changes due to life events such as marriage, childbirth, or transitions to Medicare. The full and primary samples have similar distributions of medical spending and demographics, indicating that the restrictions do not introduce systematic bias along these key dimensions.

3 Empirical Evidence

With the setting established, I document the spending response to the 2016 reform, focusing on two patterns that prove central to the analysis. Figure 1a plots mean monthly medical spending in the primary sample, comparing raw spending to spending adjusted for medical price trends and the aging of the sample.[3] Mean spending declines modestly at the start of 2016 and remains persistently lower over the subsequent months, with the pattern visible in both the raw and adjusted series. To quantify the average response, Column (1) of Table 3 estimates that real annual individual spending falls by $209, or approximately 4 percent relative to 2015. This response is substantially smaller than the 11.8 to 13.8 percent

[2]The paid date records when the insurer processes the claim and determines the patient's cost-sharing obligation.

[3]The adjusted series removes changes in spending associated with the sample's aging and deflates nominal spending using the medical care component of the Consumer Price Index for the Midwest, following the approach in Brot-Goldberg et al. (2017).

Table 2: Summary Statistics, 2015

	Full Sample	Primary Sample
Number of Households	64,409	34,036
Number of Members	163,584	86,987
Male	48.4%	48.6%
Family Coverage	86.9%	89.5%
Number of Family Members		
1	14.0%	12.8%
2	14.9%	16.5%
3	18.3%	19.8%
4+	52.8%	51.0%
Age: Member Level		
< 18	27.3%	29.6%
$18 - 29$	18.0%	12.1%
$30 - 54$	41.2%	45.5%
≥ 55	13.5%	12.8%
Projected Annual Cost: Individual Level		
Mean	\$4,547	\$4,347
25th percentile	\$1,485	\$1,481
Median	\$2,727	\$2,641
75th percentile	\$5,116	\$5,042
95th percentile	\$13,389	\$12,781
99th percentile	\$29,254	\$26,818

Notes: The table reports summary statistics for 2015. The full sample includes members observed in 2015 after restricting the data to eligible employees and dependents enrolled in the IYC Health Plan with single or family coverage. The primary sample additionally requires a nonmissing prospective risk score; twelve months of enrollment in the same plan and coverage tier within each year; continuous enrollment under the same family ID from 2015 through 2017; a valid coverage structure; twelve member months in each analysis year; and a balanced 2015–2016 panel. The primary sample also requires stable retained family composition between 2015 and 2016: an entire family is excluded if any retained member is not present under that family ID in both years. Household counts are based on unique family IDs. Member counts, gender, age, family size, and projected annual costs are calculated at the member level. Projected annual cost is the first nonmissing value observed in 2015 and is not summed across months.

reduction documented by Brot-Goldberg et al. (2017) following a required transition from a plan with no out-of-pocket cost sharing to a high-deductible plan, consistent with the more modest change in cost sharing in this setting.

Figure 1b plots the change in real annual individual spending from 2015 to 2016 across the full distribution of baseline risk scores. The magnitude of the spending reduction rises steeply with the baseline risk score, with little response among low-risk enrollees and large declines concentrated at the top of the distribution. The risk gradient is consistent with the findings of Brot-Goldberg et al. (2017), who document that spending reductions following a substantial deductible increase are largest among the highest-risk quartile.[4] Table 3 quantifies these spending responses. Columns (1) and (2) show that real allowed spending falls from 2015 to 2016, both for total spending and for non-preventive spending.[5] Column (3) shows that the non-preventive spending reduction is concentrated among the highest-risk enrollees, with a large and statistically significant decile in the top risk quartile.

Table 4 examines the spending response by clinical category and shows that reductions are concentrated in discretionary care. Preventive care spending falls across all risk quartiles, and the gradient is steepest at the top of the risk distribution, where the highest-risk enrollees reduce preventive spending by an additional \$91 relative to the lowest-risk quartile. Mental health spending follows a similar gradient, with larger reductions among high-risk enrollees. Spending on serious illness and injuries, by contrast, shows little systematic response across risk quartiles. This lack of response likely reflects the more limited scope for patient discretion in care for conditions such as cancer, heart failure, or acute injuries.

Together, these results establish two facts that prove central to the analysis. First, the introduction of a modest cost-sharing generated a modest overall spending reduction. Second, the spending response exhibits a steep risk gradient increasing nearly monotonically across the predicted spending distribution and concentrated among the highest-risk enrollees.

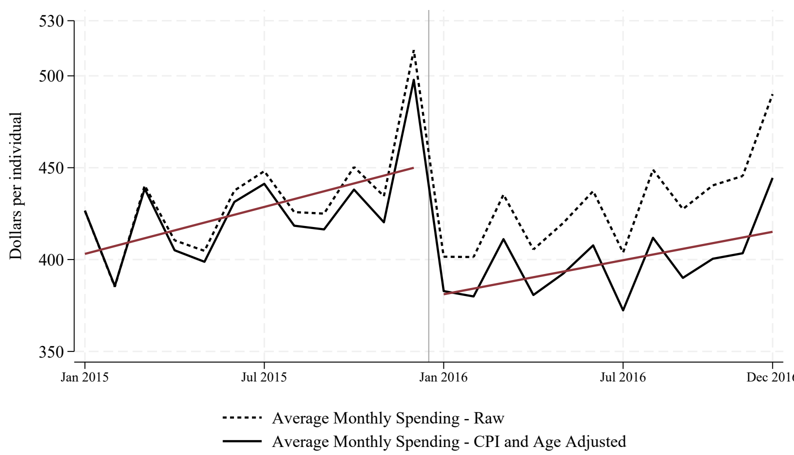
4 Perfect Foresight Model

The health insurance literature typically treats medical utilization as a forward-looking decision. Under nonlinear contracts, care used today can affect the marginal prices consumers face later in the coverage year (Einav et al., 2013; Ho and Lee, 2023; Diaz-Campo, 2022).

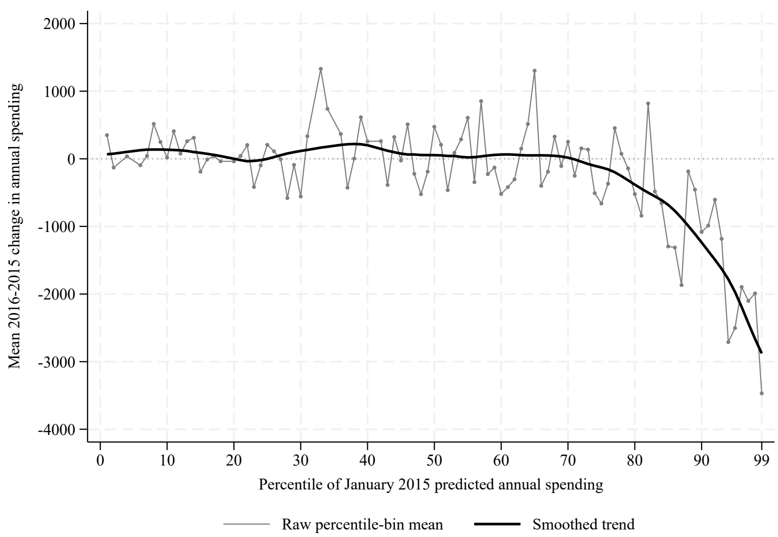
[4] Their setting involves a deductible increase from \$0 to \$3,750.

[5] I focus on non-preventive spending in the structural model because preventive care is generally covered without out-of-pocket cost sharing. Although preventive utilization also changes after the plan redesign, these services do not provide the same marginal-price variation that is central to the model.

Figure 1: Medical Spending and Spending Response



(a) Monthly Medical Spending, Overall



(b) Spending Response by Baseline Health Risk

Note: Left panel plots mean monthly medical spending at the member level in the primary sample, with raw spending (dashed line) and spending adjusted for aging and medical price trends (solid line). Right panel plots the change in real annual individual spending from 2015 to 2016 by percentile of January 2015 predicted annual spending. Points show raw percentile-bin means and the solid line shows a LOWESS-smoothed trend.

Table 3: Effect of Plan Design on Real Individual Allowed Amounts

	(1) Total Medical Spending	(2) Non-preventive Medical Spending	(3) Non-preventive Spending by Risk Quartile
Year = 2016	-209.3***	-170.1***	113.7**
	(49.85)	(49.48)	(49.62)
Year = 2016 × Risk quartile 2			46.0
			(86.88)
Year = 2016 × Risk quartile 3			-2.7
			(108.37)
Year = 2016 × Risk quartile 4			-1,138.2***
			(152.63)
Mean spending in 2015	5,164	4,806	4,806
Observations	173,974	173,974	173,974
Adjusted R-squared	0.468	0.467	0.467
Person fixed effects	Yes	Yes	Yes

Notes: The table reports OLS estimates of real annual allowed medical spending at the individual level, expressed in 2015 dollars. The regression sample is the primary sample described in Table 2. It contains 86,987 individuals observed in both 2015 and 2016, yielding 173,974 person-year observations. The omitted year is 2015, so the coefficient on Year = 2016 in Columns (1) and (2) captures the change in real spending relative to 2015. Column (1) uses total allowed medical spending, including preventive care. Columns (2) and (3) use non-preventive allowed medical spending. Column (3) interacts the 2016 indicator with baseline risk quartile, with risk quartile 1 as the omitted group. Accordingly, the coefficient on Year = 2016 reports the change for risk quartile 1, while the interaction coefficients report differences relative to risk quartile 1. Risk quartiles are based on prospective risk measured in January 2015. All specifications include person fixed effects and the controls reported in the text. Real spending is deflated using the Midwest CPI-U Medical Care index. Standard errors, clustered by family, are in parentheses. * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Table 4: Changes in Real Spending by Clinical Category and Risk Quartile, 2016

	Preventive		Mental Health		High-cost		Injury	
	Base	Risk Q	Base	Risk Q	Base	Risk Q	Base	Risk Q
Year = 2016	-84.127***	-28.096***	-6.370	16.714***	20.252	-15.111	10.002	26.729
	(3.754)	(3.318)	(5.186)	(6.329)	(13.237)	(19.867)	(9.247)	(17.208)
Year = 2016 × Q2		-52.027***		5.547		70.491**		-15.129
		(7.166)		(8.929)		(33.196)		(24.305)
Year = 2016 × Q3		-83.199***		-10.480		53.775		-0.950
		(9.121)		(10.409)		(38.077)		(23.884)
Year = 2016 × Q4		-90.665***		-88.168***		16.599		-50.260*
		(10.805)		(19.176)		(68.654)		(28.270)
Observations	294,693	294,693	294,693	294,693	294,693	294,693	294,693	294,693
Adjusted R-squared	0.104	0.104	0.273	0.273	0.231	0.231	0.038	0.038

Notes: The table reports OLS estimates of annual person-level real allowed spending by clinical category. Preventive spending includes claims classified as preventive or administrative health encounters. Mental health spending includes anxiety, depression, bipolar disorder, psychoses, schizophrenia, and other mental health conditions. High-cost spending includes serious chronic or high-cost conditions, including congestive heart failure, renal failure, COPD, dementia, and cancer-related categories. Injury spending includes injury, fracture/dislocation, and skin burn categories. The omitted year is 2015. In the risk-quartile specifications, Q1 is the omitted category, so the coefficient on Year = 2016 gives the 2016 change for the lowest-risk quartile. All regressions include person fixed effects, prospective risk score, and coverage-tier controls. Standard errors are clustered at the family level and shown in parentheses. Controls, omitted categories, and constant terms are not reported. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Households' utilization incentives therefore depend on how their spending accumulates over the year and where it places them on the cost-sharing schedule. Building on the static framework in Ho and Lee (2023), I develop and estimate a perfect-foresight model in which households know their annual health needs and choose utilization under the full nonlinear contract. Although perfect foresight abstracts from uncertainty about future health shocks, the forward-looking mechanism central to this analysis also operates under uncertainty, as households still account for how current utilization affects the marginal prices they expect to face later in the year.

I use the estimated model to assess whether this standard benchmark can jointly account for the pre-reform spending distribution and the spending response to the reform across baseline health risk. The analysis that follows shows that the model does not fully reproduce this gradient and examines how households' anticipation of future marginal prices contributes to the mismatch.

4.1 Model setup

The benchmark adopts the annual utilization framework developed by Ho and Lee (2023), which builds on the multiplicative moral-hazard specification in Einav et al. (2013). Under this framework, households observe their annual health needs and choose annual medical spending under the full nonlinear cost-sharing schedule. The benchmark also retains the quadratic health-benefit specification used by Ho and Lee (2023). Conditional on positive medical spending, the household's optimal problem is therefore the same as in Ho and Lee (2023). The main departure concerns how the model generates zero spending. Rather than using a fixed hassle cost for positive spending, I introduce a lumpy claim-arrival process in which zero spending arises from the absence of realized claim events.

4.1.1 Medical Spending Choice

The contract households face in year t is $j_t = (D_t, s_t, M_t)$, where D_t is the deductible, s_t is the coinsurance rate, and M_t is the out-of-pocket maximum. Under perfect foresight, household i observes its annual health need λ_i at the beginning of year t and chooses total annual spending m_{it} to maximize within-year utility, given by

$$u(m_{it}; \lambda_i, \omega_i, j_t) = h(m_{it}; \lambda_i, \omega_i) - \mathrm{OOP}_t(m_{it}). \tag{1}$$

The function $h(m_{it}; \lambda_i, \omega_i)$ captures the health benefit from medical care, while $\mathrm{OOP}_t(m_{it})$ is the out-of-pocket payment under contract j_t.

Following Ho and Lee (2023), I specify health benefits as a quadratic loss around annual health need:

$$h(m_{it}; \lambda_i, \omega_i) = (m_{it} - \lambda_i) - \frac{(m_{it} - \lambda_i)^2}{2\omega_i \lambda_i}, \tag{2}$$

where $\omega_i > 0$ governs household i's responsiveness of medical spending to the marginal price. Under the interior first-order-condition, ω_i equals the proportional increase in spending as the marginal price falls from no insurance to full coverage. The specifications below either impose a homogeneous value of ω or allow it to vary across households.

The out-of-pocket payment function is determined by the nonlinear cost-sharing schedule:

$$\mathrm{OOP}_t(m_{it}; D_t, s_t, M_t) = \begin{cases} m_{it}, & 0 \leq m_{it} \leq D_t, \\ D_t + s_t(m_{it} - D_t), & D_t < m_{it} < \bar{m}_t, \\ M_t, & m_{it} \geq \bar{m}_t, \end{cases} \tag{3}$$

where $\bar{m}_t = D_t + \frac{M_t - D_t}{s_t}$ is the annual spending level at which the out-of-pocket maximum is reached. Within a cost-sharing region with marginal out-of-pocket price p_t, the interior first-order condition on equation 2 implies

$$m^*_{it}(p_t) = \lambda_i \left[1 + \omega_i(1 - p_t)\right]. \tag{4}$$

Because the out-of-pocket schedule is piecewise linear, the household's problem is not solved by a single first-order condition. Instead, the household compares utility at the candidate spending level in each cost-sharing region, along with the kink points at D_t and $\bar{m}_t$, and chooses whichever spending level yields the highest utility over the full nonlinear budget set.

4.1.2 Health Needs and Claim Arrivals

Annual health needs vary across households according to observed baseline characteristics. As in Ho and Lee (2023), annual health needs are log-normally distributed, $\log \lambda_i \sim N(\nu_i, \sigma^2_\lambda)$. I parameterize the mean of log health needs as a quadratic function of prospective health risk: $\nu_i = \beta_0 + \beta_1 \log r_i + \beta_2 (\log r_i)^2$. Here, r_i denotes the individual's prospective DCG risk score. The parameter σ_λ captures dispersion in health need that is not explained by the risk score, so households with the same r_i can still differ in latent need.

Zero annual spending, however, need not imply an absence of underlying health needs. Because medical care occurs through discrete episodes, some individuals may experience no claim-generating event during the coverage year. I capture this extensive margin with a discrete claim-arrival process. Let n_i denote the number of realized claim events during the coverage year. Claim events arrive according to

$$n_i \sim \text{Poisson}(\mu_i), \tag{5}$$

where μ_i is individual i's expected annual number of claim arrivals. Since latent health need reflects the underlying demand for care, I allow the expected number of claim events, μ_i, to increase with λ_i.[6]

Specifically, I use empirical relationship between projected annual cost and active claim

[6] Ho and Lee (2023) instead generate zero spending by adding a fixed hassle cost, $-\zeta_i \mathbf{1}\{m_{it} > 0\}$, to the health-benefit function corresponding to equation 2. A household chooses zero spending when the benefit from positive medical care is insufficient to cover its out-of-pocket payment and this fixed cost. A related alternative used by Einav et al. (2013) specifies a shift log-normal distribution, $\log(\lambda_{it} - \kappa_{\lambda,i}) \sim N(\mu_{\lambda,it}, \sigma^2_{\lambda,i})$. When the shift parameter is negative, sufficiently low draws of λ_{it} can imply zero spending. Both approaches therefore generate zero spending primarily from the lower tail of the latent health-need distribution. In the present model, zero spending instead occurs when no claim event is realized, $n_i = 0$. Because $\Pr(n_i = 0 \mid \lambda_i) = \exp(-\mu_i)$ and μ_i increases with λ_i, zero spending remains more likely among individuals with lower health needs but is not determined by a fixed health-need threshold.

days to map latent health need into a relative claim-arrival rate. Let $g(\lambda_i)$ denote predicted active claim days at health-need level, λ_i:

$$g(\lambda_i) = \alpha_0 + \alpha_1\lambda_i + \alpha_2\lambda_i^2. \tag{6}$$

The coefficients $\alpha_0, \alpha_1, \alpha_2$ come from a regression of active claim days on projected annual cost, reported in Appendix Table B.1. I then set

$$\mu_i = \bar{\mu}\frac{g(\lambda_i)}{E[g(\lambda_i)]}, \tag{7}$$

where $g(\lambda_i)/E[g(\lambda_i)]$ determines household i's claim-arrival rate relative to the sample average, and $\bar{\mu}$ is estimated by SMM and determines the average annual claim-arrival rate. When $n_i = 0$, annual spending is zero. When $n_i > 0$, the household chooses total annual spending according to the perfect-foresight problem above, regardless of the realized number of claim events. Specifically,

$$m_{it}^{obs} = \begin{cases} 0, & n_i = 0, \\ m_{it}^*(\lambda_i, j_t), & n_i > 0. \end{cases} \tag{8}$$

This lumpy claim-arrival structure is useful for matching the extensive margin of annual spending. Rather than generating zero spending only from the bottom of the latent health-need distribution, as in a fixed hassle cost (Ho and Lee, 2023) or shifted-need specification, the model allows zero spending to arise from the absence of realized claim events. This distinction is also useful for the myopic model below, where realized claim timing shapes the marginal prices households face over the course of the year.

4.2 SMM Estimation of the Perfect-Foresight Model

I take the model structure above to the Wisconsin claims data using simulated method of moments. The estimated parameter vector is

$$\theta = \left(\omega_i, \beta_0, \beta_{\ln(\text{DCG})}, \beta_{\ln(\text{DCG})^2}, \bar{\mu}, \sigma_\lambda\right),$$

where ω_i governs the spending response to marginal prices, $\beta_0, \beta_{\ln(\text{DCG})}, \beta_{\ln(\text{DCG})^2}$ determine the relationship between baseline risk and latent health need, $\bar{\mu}$ controls the average annual claim-arrival rate, and σ_λ captures residual heterogeneity in latent health need. For the heterogeneous-ω specification, I replace the homogeneous parameter ω with the parameters $\rho_0, \rho_{\omega,1}, \rho_{\omega,2}$, and σ_ω, which govern the conditional distribution of ω_i. The coefficients α_0, α_1, and α_2 in equation 6 are estimated separately from the regression of active claim days on projected annual cost reported in Appendix Table B.1. I hold these coefficients fixed during SMM, so they are not included in θ.

To compute simulated moments, I generate R paths for each household and construct simulated annual spending as follows:

1. For each household i and path $r = 1, \ldots, R$, I draw $\varepsilon_{ir} \sim N(0, 1)$ and construct health need λ_{ir} using the linear and quadratic effects of log projected risk and the residual heterogeneity term σ_λ.

2. Given λ_{ir}, I compute the household-specific claim-arrival rate μ_{ir} and simulate the realized claim amount.

3. I apply the lumpy spending rule in equation 8, using household i's assigned value of ω_i, to obtain simulated annual spending.

I hold the simulation draws fixed across parameter evaluations and apply the same draws to the 2015 and 2016 contracts, so that simulated cross-year spending differences reflect the change in cost sharing rather than differences in random health-need and claim-arrival draws.

The SMM objective minimizes the weighted distance between simulated and empirical moments:

$$\hat{\theta} = \arg\min_{\theta} \left[\hat{m}^{sim}(\theta) - \hat{m}^{data}\right]' W \left[\hat{m}^{sim}(\theta) - \hat{m}^{data}\right]. \tag{9}$$

Here, W is a diagonal weighting matrix that scales moments by their empirical standard errors.[7] The targeted moments characterize the 2015 spending distribution, including its extensive margin and upper tail, as well as the regression-adjusted spending response from 2015 to 2016 overall and among individuals with higher baseline risk. The objective is minimized using a derivative-free optimizer, with multiple starting values used to reduce sensitivity to local minima.[8]

I estimate the benchmark under two specifications of ω_i. In the baseline specification, $\omega_i = \omega$ is homogeneous across households, so differences in the spending response arise entirely from differences in health need, claim arrivals, and prices households face. Because

[7] I impose lower bounds on the moment scales to prevent moments with very small standard errors from mechanically dominating the objective.

[8] I use Powell's method because the objective is non-smooth due to the kinked cost-sharing schedule and the claim-arrival extensive margin.

spending reductions in the data increase sharply with baseline risk, I also estimate a more flexible specification in which the response to cost sharing can vary both across the risk distribution and among households with the same measured risk. The purpose is to assess whether the perfect-foresight model can reproduce the observed gradient when given greater flexibility in household responsiveness, rather than to interpret baseline risk as a determinant of ω_i.

I model household heterogeneity in ω_i using a heterogeneous distribution whose conditional mean varies with baseline health risk. Let r_i denote household i's standardized log baseline risk score. I parameterize the conditional mean of ω_i as

$$\mu_\omega(r_i) \equiv E[\omega_i | r_i] = \exp(\rho_0 + \rho_{\omega,1} r_i + \rho_{\omega,2} r_i^2). \tag{10}$$

The linear and quadratic terms allow mean responsiveness to vary smoothly across the baseline risk distribution.

Conditional on baseline risk, ω_i follows a lognormal distribution:

$$\ln \omega_i \mid r_i \sim N\left(\ln \mu_\omega(r_i) - \frac{1}{2}\sigma_\omega^2, \sigma_\omega^2\right). \tag{11}$$

Here, σ_ω governs residual variation in ω_i among households with the same baseline risk. The adjustment $-\sigma_\omega^2/2$ ensures that $E[\omega_i|r_i] = \mu_\omega(r_i)$. This specification therefore allows ω_i to vary continuously across households without imposing discrete response types. Each household's value of ω_i is drawn once from its conditional distribution and held fixed across the 2015 and 2016 contracts. I apply the same specification to the models that follow.

4.3 Parameter Estimates and Model Fit

Table 5 reports parameter estimates for the homogeneous- and heterogeneous-ω specifications. Both specifications imply a large response to marginal prices. In the homogeneous-ω specification, the estimate of ω is 8.007, implying that a decline in the marginal price from one to zero raises desired spending approximately 800 percent relative to latent health need. In the heterogeneous-ω specification, the conditional mean of ω_i at the average baseline risk score is $\exp(2.144) = 8.53$, corresponding to an increase of approximately 853 percent. Although the heterogeneous-ω specification allows mean responsiveness to vary with baseline risk, the estimated conditional mean of ω_i varies only modestly across the risk distribution.

Table 6 reports the fit of both specifications to the targeted moments. Both specifica-

Table 5: Estimated Perfect-Foresight Model Parameters

	Estimate	
Parameter	Homogeneous-ω	Heterogeneous-ω
ω	8.007	–
ρ_0	–	2.144
ρ_1	–	0.071
ρ_2	–	0.002
σ_ω	–	0.027
β_0	5.026	5.294
$\beta_{\ln(\mathrm{DCG})}$	1.147	0.848
$\beta_{\ln(\mathrm{DCG})^2}$	-0.089	-0.018
$\bar{\mu}$	2.462	2.349
σ_λ	1.150	1.052

Notes: The table reports estimated perfect-foresight model parameters. Estimation uses non-preventive spending because preventive care faces zero cost sharing under both contracts.

tions broadly reproduce the pre-reform spending distribution, although the heterogeneous-ω specification more closely matches mean spending and the zero-spending share. They nevertheless understate the average spending response, predicting a reduction of approximately \$108 compared with \$170 in the data, even though this moment receives substantial weight in estimation. This discrepancy becomes much larger among high-risk individuals. Among individuals in the 75th-90th risk percentiles, the homogeneous- and heterogeneous-ω specifications predict spending reductions of \$138 and \$149, respectively, compared with \$401 in the data. The gap widens further among individuals in the top 10 percent of risk, for whom the predicted reductions range from \$148 to \$157, compared with \$1959 in the data. Thus, the added flexibility in ω_i does little to improve the perfect-foresight model's ability to reproduce the spending response among the highest-risk individuals.

Figure 2 reinforces the results in Table 6 by plotting the simulated spending response across percentiles of the baseline risk score for both specifications. Under both specifications, the predicted spending reduction reaches approximately \$170 between the 90th and 95th risk percentiles before diminishing sharply to approximately \$140 at the very top. The model therefore predicts that the spending response weakens at the top of the risk distribution, whereas the data show the largest reductions in this group.

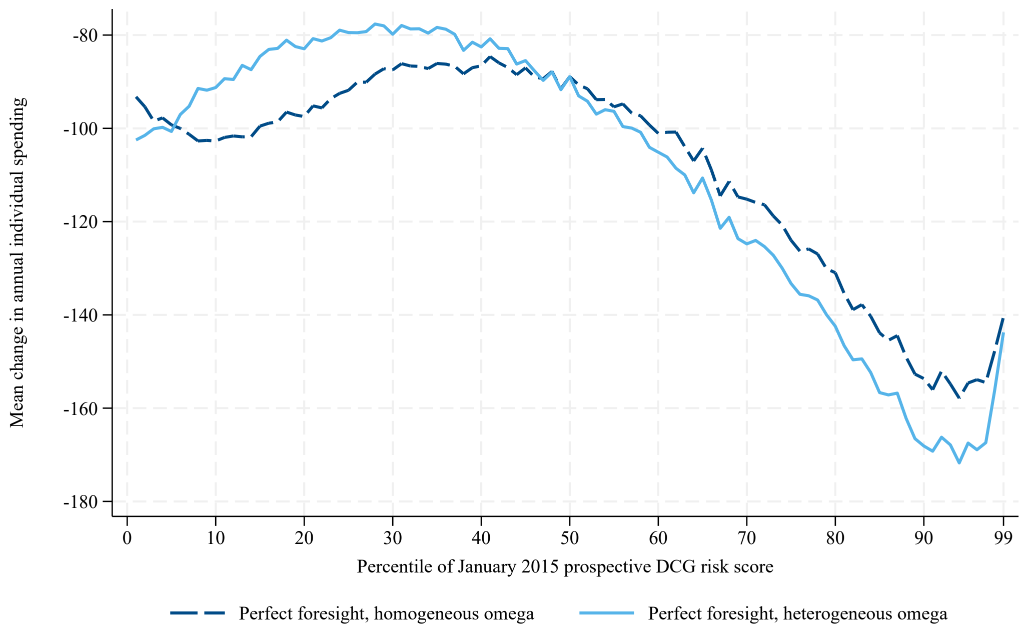
Figure 3 shows why the perfect-foresight model predicts little spending response among the highest-need households even when ω is large. The figure is constructed using the estimated homogeneous value $\omega = 8.007$, which already implies a large response to changes

Table 6: Targeted Moment Fit

Moment	Data	Homogeneous ω	Heterogeneous ω
Panel A. Pre-reform spending distribution			
Mean spending, 2015	4,805.9	4,396.3	4,634.2
Standard deviation of spending, 2015	13,587.0	13,790.5	13,519.6
Zero-spending share, 2015	0.104	0.087	0.097
Share reaching out-of-pocket maximum, 2015	0.189	0.191	0.212
Mean spending, 2015, 75th–90th risk percentiles	6,974.8	7,301.7	7,111.0
Mean spending, 2015, top 10% of risk	17,345.9	17,585.3	17,684.1
Panel B. Spending response to the reform			
Regression-adjusted spending change, 2016–2015	-170.1	-108.4	-108.7
Regression-adjusted spending change, 75th–90th risk percentiles	-400.6	-137.5	-149.2
Regression-adjusted spending change, top 10% of risk	-1,958.5	-147.6	-157.1

Notes: The table reports the empirical moments and their simulated counterparts under the homogeneous-ω and heterogeneous-ω perfect-foresight specifications. Spending amounts and spending changes are measured in dollars. Baseline risk groups are defined using predicted spending before the reform. Spending-change moments are regression-adjusted. All moments use non-preventive spending because preventive care faces zero cost sharing under both contracts.

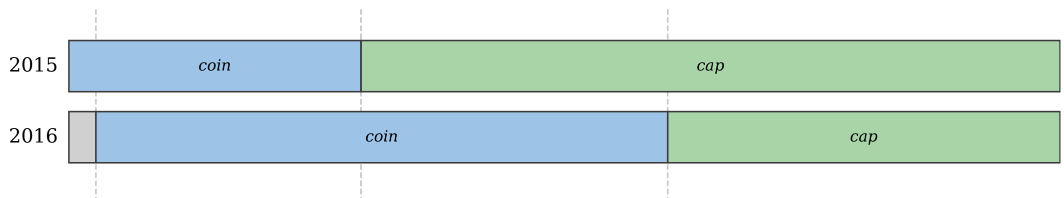
Figure 2: Simulated Spending Response Under Perfect Foresight by Baseline Risk



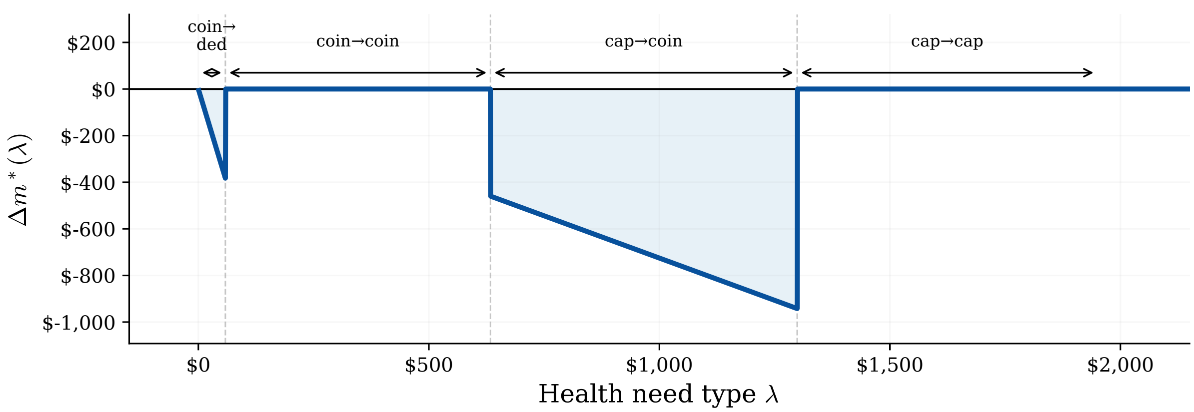
Notes: The figure plots the simulated change in annual individual spending from 2015 to 2016 under the homogeneous- and heterogeneous-ω perfect-foresight specifications, by percentile of the January 2015 prospective DCG risk score. Negative values indicate lower simulated spending under the 2016 contract relative to the 2015 contract.

in marginal prices. Panel a shows the cost-sharing regions households occupy under the 2015 and 2016 contracts at each level of latent annual health need λ. Panel b shows the corresponding change in optimal spending. The predicted response depends on whether the reform changes the marginal price a household faces. Spending falls for households who move into a higher-price region but remains unchanged for those who face the same marginal price under both contracts. For the highest-need households, the marginal price remains at zero because they reach the out-of-pocket maximum under both contracts. This feature of the perfect-foresight model limits its ability to reproduce the spending response among the highest-risk households, even when ω_i varies flexibly across households. This limitation motivates a myopic model in which households respond to the current marginal price rather than anticipating how current spending affects prices later in the year.

Figure 3: Contract Structure and Simulated Spending Response Under Perfect Foresight



(a) Contract Structure by λ Percentile



(b) Perfect-Foresight Spending Response by λ Percentile

Notes: Panel A shows the cost-sharing structure of the 2015 and 2016 contracts by percentile of latent annual health need λ. Shading indicates the deductible (gray), coinsurance (blue), and out-of-pocket maximum (green) regions. Panel B plots the simulated change in annual spending from 2015 to 2016 under the estimated perfect-foresight model, by percentile of λ. Negative values indicate lower simulated spending under the 2016 contract relative to the 2015 contract.

5 A Myopic Model of Utilization

The results in the preceding section suggest that the perfect-foresight benchmark places too much weight on future prices in current utilization decisions. I therefore consider a myopic model in which households respond to the price they perceive at the time of care but do not internalize how current spending affects prices later in the year. This shifts the decision problem from a forward-looking annual choice to a sequence of weekly utilization decisions. The weekly structure can allow high-need households to remain responsive to cost sharing when they use care, even when a forward-looking household would anticipate reaching lower marginal prices later in the year.

Formally, I divide the coverage year into $T = 52$ weekly decision periods and let λ_i denote household i's annual latent health need. To capture the lumpiness of medical care, annual health need is realized through discrete claim arrivals rather than as a smooth weekly flow. As in equation 7, weekly claim counts follow $n_{it} \sim \text{Poisson}(\mu_i/T)$, where μ_i is the household's expected annual number of claim arrivals and is higher for households with greater health need. Let ℓ_{it} denote realized weekly health need. Weeks with no claim arrival have $\ell_{it} = 0$; conditional on at least one annual claim, λ_i is allocated across positive-claim weeks using random claim-size shares, so that $\sum_{t=1}^{T} \ell_{it} = \lambda_i$.

Let $\tilde{p}_{it}$ denote the price household i responds to in week t. Given realized weekly health need ℓ_{it}, the household takes $\tilde{p}_{it}$ as the price of current care. In weeks with positive realized need, the household solves

$$\max_{m_{it} \geq 0} \; h(m_{it}; \ell_{it}, \omega_i) - \tilde{p}_{it} m_{it}. \tag{12}$$

The resulting interior choice is

$$m_{it} = \ell_{it}[1 + \omega_i(1 - \tilde{p}_{it})]. \tag{13}$$

In weeks with no realized health need, $\ell_{it} = 0$ and $m_{it} = 0$.

Let

$$C_{i,t-1} = \sum_{\tau=1}^{t-1} \Delta OOP_{i\tau} \tag{14}$$

denote cumulative out-of-pocket spending under the true contract at the beginning of week t, where $\Delta OOP_{i\tau}$ is the out-of-pocket payment generated in week τ. After utilization is chosen, ΔOOP_{it} is calculated under the true nonlinear contract, and cumulative out-of-pocket spending is updated to $C_{it} = C_{i,t-1} + \Delta OOP_{it}$.

A natural benchmark sets $\tilde{p}_{it} = p_{it}$, where p_{it} is the true spot price implied by the current claim and the household's cost-sharing position at the beginning of week t, $C_{i,t-1}$.[9] This benchmark assumes that households know where they are on the cost-sharing schedule and can anticipate how the current episode of care will move them across the deductible, coinsurance, and out-of-pocket-maximum regions. It is not clear, however, that households have the information needed to identify this price at the time of care.

The myopic framework therefore raises a further question about which price households actually respond to at the time of care. The weekly decision problem specifies how utilization responds to a given price but does not determine how households form that price. Answering this question requires understanding what households know about their current contract position and how they use that knowledge to assess the cost of additional care. The next section provides evidence on how the timing and interpretation of medical bills shape the spot price households perceive.

6 Empirical Evidence on Perceived Spot Prices

6.1 Billing Delays

Households may not know the price of care when making utilization decisions. The services ultimately provided and the billing codes assigned to them may not be determined until after care is received.[10] Even when the underlying price is available, it does not by itself reveal the relevant out-of-pocket price, which also depends on the household's current position on the cost-sharing schedule. When recent claims remain unprocessed, the position households can infer from the bill they have received may lag behind their true position under the contract. As a result, households may be unable to assess whether additional care will remain within the current cost-sharing region or move them across the deductible or out-of-pocket maximum. A bill or explanation of benefits (EOB) is therefore often the

[9] Appendix C reports the SMM estimates and model fit for this true-spot-price benchmark. Relative to the perfect-foresight model, it generates larger spending reductions among higher-risk households but continues to understate the response at the top of the risk distribution.

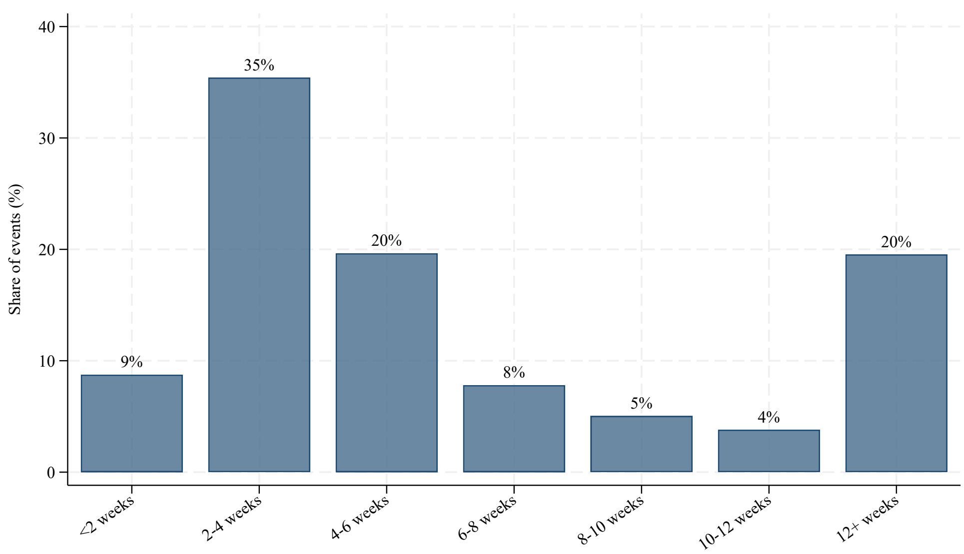
[10] Current federal transparency rules, introduced after the period studied here, have expanded access to pricing information. To obtain a relevant cost-sharing estimate, however, consumers must identify the item or service by a billing code or descriptive term, together with the provider and other relevant details. The estimate also need not incorporate outstanding claims that have not yet been processed and may differ from final liability. More broadly, prior work documents difficulties locating and interpreting disclosed prices and shows that posted hospital prices may not capture the total cost of an episode of care (Haque et al., 2021; Horný et al., 2021a). See 85 Fed. Reg. 72158 (Nov. 12, 2020).

first reliable indication of how an episode was priced and applied under the contract. In this section, I document delays in the arrival of this information, show that these delays are longest among high-risk enrollees, and examine how utilization changes following bill receipt.

6.1.1 Timing of Cost-Sharing Information

Following Hoagland et al. (2022), I define billing delay as the number of days between the date of service and the date on which the insurer pays its portion of the claim. At this point, the claim has been processed and an EOB can be generated, making the paid date an observable proxy for when information about the household's cost-sharing obligation becomes available. Consistent with this interpretation, Hoagland et al. find a spending response around the actual insurer-paid date, while estimated responses around placebo dates constructed from randomly assigned billing delays are centered near zero. Figure 4 shows the distribution of these delays across all claims in the 2016 sample. Thirty-five percent of claims are paid within two to four weeks after service, but delays are highly dispersed, with 20 percent not paid until more than 12 weeks later.

Figure 4: Distribution of Billing Delays



Note: The figure plots the distribution of billing delays across all claims in the sample. Each bar reports the share of claims for which the number of days between the medical service date and the paid date falls within the indicated range.

The long right tail in Figure 4 is not evenly distributed across types of care. Table 7 compares conditions with the longest and shortest billing delays. Long delays are concentrated in types of care that likely involve more complex billing and claims processing, including chronic disease management and cancer-related treatment. Among the longest-delay conditions, median delays range from roughly two to four months, and a large share of events are not paid until at least 12 weeks after the service date. The shortest-delay conditions, by contrast, are more routine or acute categories, such as eye disorders and common infections, with median delays of about three to four weeks and much smaller long-delay shares.

Table 7: Conditions with the Longest and Shortest Billing Delays

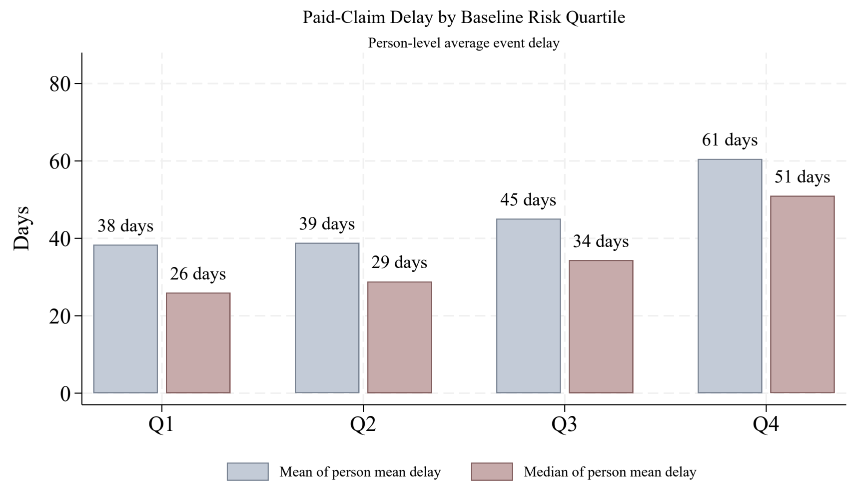
Condition	N	Median delay	Share 12+ weeks	Mean allowed amount
Panel A. Longest-delay conditions				
Crohn's Disease	3,473	120.0	58.5	1,038.8
Cancer – Leukemia	1,506	93.0	53.5	682.9
Cancer – Colon	1,180	95.0	53.8	717.8
Diabetes	26,412	81.0	49.5	179.8
Chemotherapy Encounters	1,969	83.0	49.6	3,275.4
Ulcerative Colitis	2,089	69.0	45.8	949.0
Epilepsy	1,708	52.5	43.1	420.4
Cancer – Breast	7,038	80.0	48.8	820.3
Panel B. Shortest-delay conditions				
Eye Disorders, NEC	49,603	22.0	3.8	95.7
Infections – Eye	5,440	22.0	3.6	170.7
Infections – Body Sites, NEC	1,417	23.0	3.7	139.4
Otitis Media	8,399	23.0	3.3	215.1
Infections – Multisystem	1,719	25.0	3.0	374.6
Eye Disorders, Congenital	1,948	22.0	4.7	260.3
Infections – ENT Ex Otitis Med	42,744	23.0	4.7	177.4
Pneumonia, Bacterial	3,730	27.0	3.0	309.0

Notes: Delay is measured in days. Share 12+ weeks is the percentage of events paid at least 12 weeks after the event date. Allowed amounts are measured at the event level.

The concentration of long delays in complex care suggests that billing delays may be especially important for high-risk enrollees. Consistent with this, Figure 5 shows that mean delays increase from 38 days in the lowest-risk quartile to 61 days in the highest-risk quartile. This implies that high-risk households face longer periods in which their perceived cost-sharing position may lag behind their true position.

These patterns show that delay timing is not random with respect to the type of care. But even when the type of care is held fixed, bill arrival remains difficult to predict precisely. Prior work documents substantial administrative complexity in medical billing, including claim denial and resubmission that vary across insurers and claim types (Gottlieb et al., 2018; Dunn et al., 2024). Consistent with this, bills arrive across a wide range of delay

Figure 5: Mean and Median Lag by Risk Quartile



Note: The figure shows mean and median person-level billing delays by baseline risk quartile. Billing delay is measured as the number of days between the medical service date and the paid date.

windows even within identical provider-procedure-insurer-diagnosis combinations (Appendix Figure D.1).

To quantify how much of the variation in delay timing is predictable from claim characteristics, Table 8 reports regressions of paid-claim delay on increasingly rich sets of fixed effects, starting from a baseline RMSE of 72.64 days. Absorbing separate fixed effects for provider, procedure, insurer, and diagnosis reduces the RMSE modestly to 65.17 days and explains only 20 percent of the total variation in delay timing. Column 3 absorbs fully interacted provider-procedure-insurer-diagnosis cell fixed effects, yet the RMSE falls only marginally to 64.9 days and R-squared rises to just 0.217. These results show that observable claim characteristics explain some systematic variation in billing delays, but that most of the exact timing remains difficult to predict even within detailed claim cells.

6.1.2 Utilization Responses to Bill Receipt

I next examine whether utilization changes when cost-sharing information becomes available following these delays. Building on Hoagland et al. (2022), I center the event study on the insurer-paid date of the first encounter in each person-year that generates a deductible payment. To distinguish changes around the insurer-paid date from the natural trajectory of care following the encounter, the specification controls flexibly for weeks since the encounter and restricts the sample to claims paid at least five weeks after service. The estimates show that the probability of any claim declines following the insurer-paid date, while the allowed-

Table 8: Predictability of Paid-Claim Delays from Claim Characteristics

	(1) No fixed effects	(2) Provider, procedure, insurer, and diagnosis fixed effects	(3) Provider × procedure × insurer × diagnosis cell fixed effects
Observations	149,257	149,257	149,257
R-squared	0.000	0.202	0.217
Adjusted R-squared	0.000	0.195	0.202
Residual RMSE, days	72.64	65.17	64.90
RMSE / raw SD	1.000	0.897	0.893

Notes: This table reports regressions of event-level paid-claim delay on increasingly rich sets of fixed effects. The outcome is paid-claim delay in days, defined as the number of days from the incurred date to the first paid date within the event. Events are defined by person, incurred date, provider, procedure, insurer, and principal diagnosis. The sample is restricted to events incurred in 2016 and first paid in 2016. Column 2 absorbs provider, procedure, insurer, and principal diagnosis fixed effects separately. Column 3 absorbs provider × procedure × insurer × principal diagnosis cells. Cells with fewer than 20 events are excluded. Residual RMSE measures the remaining unexplained variation in delay days after absorbing the fixed effects.

spending estimates are generally negative but less precise (see Appendix Section D.2 for the full specification and event-study estimates). Consistent with the prior evidence, these patterns suggest that households adjust subsequent utilization when cost-sharing information becomes available.

To examine whether the response depends on the cost-sharing position revealed by the bill, I compare enrollees whose first bill leaves them in the deductible region with those whose first bill exhausts the deductible. I estimate a pooled specification comparing weeks 1-3 after the insurer-paid date with weeks -3 to -1 before it, excluding the week of insurer payment.

Table 9 shows that, among enrollees who remain in the deductible region, the probability of any claim declines by 1.7 percentage points and allowed spending declines by $28 following insurer payment. For enrollees whose first bill exhausts the deductible, the post-payment change is 0.8 percentage point larger for the probability of any claim and $49 larger for allowed spending. The difference across groups suggests that households respond not simply to bill receipt, but to what the bill reveals about their position on the cost-sharing schedule.

6.2 Inferring Prices from Medical Bills

Beyond billing delays, a second potential source of price misperception arises from how households interpret the cost-sharing information they receive. The standard model predicts

Table 9: Pooled Post-Bill Response

	(1) Any claim	(2) Allowed amount
$PostBill^{1\text{-}3}_{iyt}$	-0.017^{***}	-28.01^{**}
	(0.002)	(13.05)
$PostBill^{1\text{-}3}_{iyt} \times HitDed_{iy}$	0.008^{***}	48.67^{**}
	(0.003)	(23.37)
Observations	181,488	181,488
Person FE	Yes	Yes
Week FE	Yes	Yes
Encounter-time controls	Yes	Yes

Notes: The sample consists of person-weeks falling in the symmetric window of weeks -3 to -1 and 1 to 3 relative to bill arrival, excluding the bill week. $PostBill^{1\text{-}3}_{iyt}$ equals one in weeks 1 to 3 and zero in weeks -3 to -1. $HitDed_{iy}$ indicates whether the first bill exhausts the deductible. Column (1) uses an indicator for having any claim in the person-week as the outcome. Column (2) uses the dollar amount of allowed spending in the person-week as the outcome. Encounter-time controls consist of a full set of week-by-week indicators from the encounter date through the end of the coverage year. Standard errors are clustered at the person level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

that households respond to the true marginal price of care, which changes discontinuously at the deductible and the out-of-pocket maximum. However, evidence from other nonlinear pricing settings suggests that consumers often rely on simpler price measures, including average rather than marginal prices. Rees-Jones and Taubinsky (2020) document that many taxpayers respond to average rather than marginal tax rates, and Ito (2014) finds similar patterns in electricity pricing.

The same pattern may arise in health insurance. An Explanation of Benefits (EOB) statement, the document insurers send to enrollees after each claim is processed, displays total charges and total patient responsibility, making the average cost share salient while the true marginal price is not directly displayed (Appendix Figure E.1). For claims that fall entirely within a single cost-sharing region, the average and marginal prices coincide. When a claim crosses the deductible or out-of-pocket maximum, however, the average price blends rates from two regions and can diverge from the true marginal price.

For example, if an enrollee has \$50 remaining before the deductible and then incurs a \$200 claim, a 10 percent coinsurance rate implies \$65 in out-of-pocket spending. The displayed totals could make a 32.5 percent average cost share salient, even though the true marginal price is only 10 percent. This distinction is especially relevant for high-risk enrollees, who are more likely to have claims near cost-sharing thresholds. If they rely on the average cost share implied by recent EOBs, they may continue to perceive care as costly even when the true marginal price is low or zero.

The survey experiment below is designed to test this mechanism directly. It presents respondents with EOBs in which marginal and average prices sometimes coincide and sometimes diverge, so that their answers reveal whether they rely on the true marginal price or the average price made salient by the EOB. The resulting estimates provide direct evidence on the price-perception channel and help calibrate the share of average-price responders in the structural model.

6.2.1 Experimental Design

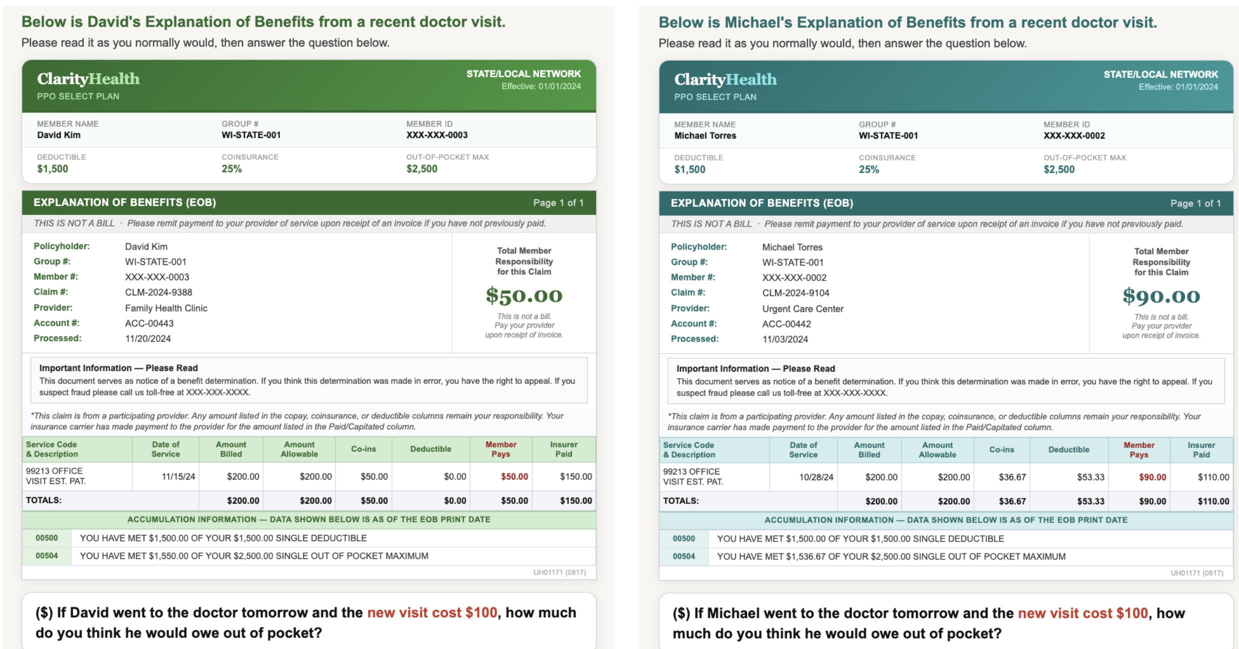
To estimate the share of households that respond to average price rather than marginal price of care, I conducted a survey experiment in which respondents report their expected out-of-pocket cost across different cost-sharing scenarios.[11] In the experiment, each respondent was assigned a health insurance plan with a $1,500 deductible, a randomly assigned coinsurance rate, and a $2,500 out-of-pocket maximum. They then answered four questions, each presenting a hypothetical EOB statement with a different spending history under that plan and asking how much the enrollee would owe out of pocket for a $100 medical visit the following day.

Across the four questions, the EOB reflected different spending histories, placing the enrollee at different points along the cost-sharing schedule. Two scenarios placed the enrollee within a single cost-sharing region, where the average and marginal prices coincide, while the other two involved a spending history that crosses a threshold, generating a divergence between the two. The coinsurance rate varied randomly across respondents, and within the crossing scenarios, the amount by which spending crosses each threshold was also randomized, creating variation in the gap between the average and marginal prices. Figure 6 shows two example questions from the survey. In Panel (a), the recent claim remains within the coinsurance region, so the average price implied by the claim and the marginal price of of next visit both equal 25 percent. In Panel (b), the recent claim crosses the deductible, so its implied average price is 45 percent, while the marginal price of the next visit is 25 percent.

To incentivize accurate responses, one of the four questions was selected at random after completion, and respondents received a $0.10 bonus payment if their stated cost fell within $5 of the true marginal price for that scenario. The survey was administered to 1,250 respondents; two were excluded from the analysis sample after being flagged for AI-generated responses, yielding a final sample of 1,248.

[11]The preregistration is available at AsPredicted #297749.

Figure 6: Examples of Survey Questions



(a) Recent Claim Within Coinsurance

(b) Recent Claim Crossing the Deductible

Notes: The figure presents example survey questions with a 25% coinsurance rate. Panel (a) shows a recent claim occurring entirely within the coinsurance region, while Panel (b) shows a recent claim crossing the deductible.

6.2.2 Marginal- and Average-Price Responses

Table 10 reports summary statistics for the analysis sample of 1,248 respondents. Ninety percent reported currently have health insurance, and 83 percent reported prior experience with medical bills or EOBs.[12] Respondents performed well on both insurance and financial literacy measures, with 72 percent answering the insurance literacy question correctly and 69 percent answering all of the Big Three financial literacy questions correctly, more than double the 29 percent rate in a nationally representative U.S. sample (Lusardi and Mitchell, 2011; Lusardi and Streeter, 2023)[13]

Figure 7 shows the share of respondents whose stated expected out-of-pocket cost matches the marginal-price or average-price benchmark in each scenario. In the two scenarios where marginal and average prices coincide, a large share of respondents correctly identified the benchmark price, with 84 percent matching in the below-deductible scenario and 67 percent in the within-coinsurance scenario. When prior spending crosses a cost-sharing threshold, however, marginal and average prices diverge, and the marginal-price response rate falls to 47 percent in the crossing-deductible scenario and 57 percent in the crossing-out-of-pocket-maximum scenario. A sizable share of respondents instead report answers consistent with the average-price benchmark, at 14 and 15 percent in the two crossing scenarios, respectively. (See Appendix Figure E.2 for the full distribution). The regression results in Appendix Table E.1 provides additional evidence of average-price reasoning. The association between the average-price benchmark and stated costs remains after accounting for respondent fixed effects and when focusing only on the crossing scenarios.

To examine whether average-price reasoning is concentrated among respondents with lower insurance literacy, Figure 8 compares marginal- and average-price response rates across literacy groups, pooling the two crossing scenarios. Marginal-price responses were more common among respondents who answered correctly (59.2 versus 32.7 percent), whereas average-price responses were less common (12.3 versus 21.3 percent). These patterns suggest that insurance literacy accounts for some, but not all, of average-price reasoning.

Following Banki and Simonsohn (2026), I also collected respondents' explanations of how they arrived at their answers. After completing all four scenarios, respondents were shown their stated answer and the EOB for the crossing-out-of-pocket-maximum scenario and asked to explain in their own words how they calculated their expected out-of-pocket

[12] Results are robust to restricting the sample to currently insured respondents.

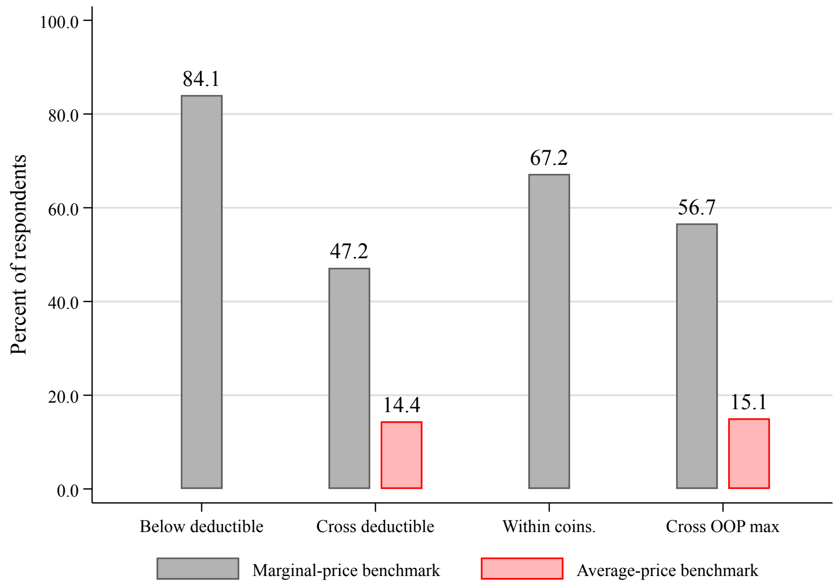
[13] The insurance literacy question asks respondents how much they would pay out of pocket if they had $2,000 in total medical expenses under a plan with a $1,000 deductible, 20 percent coinsurance, and a $4,000 out-of-pocket maximum. The correct answer is "between $1,000 and $4,000."

Table 10: Summary Statistics

	Analysis Sample
Number of Respondents	1,248
Female	55.9%
Bachelor's degree or above	57.1%
Employed	73.6%
Good, very good, or excellent health	80.3%
Currently has health insurance	90.1%
Any prior bill or EOB experience	83.3%
Age	
18–24	12.4%
25–34	24.7%
35–44	28.5%
45–54	17.8%
55–64	10.9%
65 or older	5.7%
Insurance Literacy Measure	
Insurance literacy question correct	72.5%
Big 3 Financial Literacy Measure	
Inflation question correct	88.1%
Interest question correct	94.9%
Stock-risk question correct	75.6%
All three financial literacy questions correct	69.4%
Prior Experience with Medical Bills or EOBs	
Never	16.6%
Once or twice	29.7%
A few times	33.2%
Many times	19.7%
Not sure	0.8%
Survey Duration	
Mean minutes	10.62
SD minutes	5.38

Notes: The table reports summary statistics for respondents in the analysis sample. The sample includes completed non-preview survey responses that pass the attention and authenticity checks and are successfully matched to Prolific demographic information. The insurance-literacy question asks respondents to identify the price range implied by a deductible and out-of-pocket maximum. The three financial-literacy questions cover inflation, compound interest, and risk diversification. Prior bill or EOB experience is based on respondents' self-reported experience seeing medical bills or explanations of benefits. Survey duration is measured in minutes.

Figure 7: Marginal-Price and Average-Price Benchmark Responses



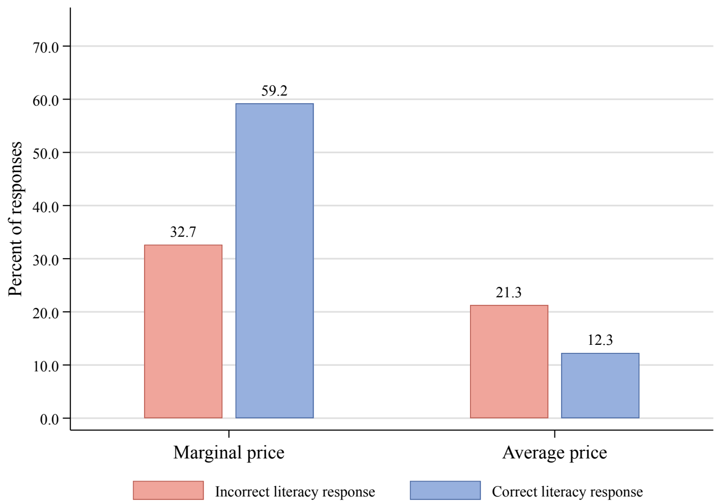
Note: The figure shows the share of respondents whose stated expected out-of-pocket cost matches the marginal-price or average-price benchmark in each insurance-cost scenario. Responses are classified as matching a benchmark if they fall within $1 of the benchmark value. Average-price bars are shown only in scenarios where the marginal and average prices differ. In the below-deductible and within-coinsurance scenarios, the marginal and average prices are identical.

cost. The responses were classified into non-mutually-exclusive categories using the OpenAI API, with the full classification prompt provided in Appendix E.3.

The examples provide direct qualitative evidence of both marginal- and average-price reasoning. Fifty-four percent recognized that reaching the out-of-pocket maximum made the subsequent visit free, with one respondent explaining, "Robert has met both his deductible and maximum out of pocket expenses, so he would owe nothing." Twelve percent instead described scaling the member responsibility from the prior EOB to the new visit as illustrated by the response, "40 is 1/5 of 200 and 20 is 1/5 of 100." Another 9 percent carried forward the member responsibility displayed on the EOB.

The 12 percent who explicitly described average-price reasoning is slightly smaller than the 15 percent whose numeric responses matched the average-price benchmark in this scenario. Two features of the elicitation may contribute to this difference. First, respondents were shown the EOB again before providing their explanation, which may have prompted some to reconsider the reasoning behind their original answer. Second, some explanations were too vague to reveal the calculation rule respondents had used. The free responses therefore provide qualitative evidence that respondents use both marginal- and average-price reasoning, but they do not provide an exact estimate of the share belonging to each type. The

Figure 8: Marginal- and Average-Price Responses in Crossing Scenarios, by Insurance Literacy



Note: The figure pools responses from the crossing-deductible and crossing-out-of-pocket-maximum scenarios and reports the shares matching the marginal-price and average-price benchmarks, separately by whether respondents answered the insurance-literacy question correctly. Each respondent contributes one response from each scenario. Responses are classified as benchmark-consistent if they fall within $1 of the relevant benchmark. Other responses are omitted from the figure but remain in the denominator.

next section estimates these shares using respondents' complete pattern of numeric answers across all four scenarios.

Table 11: Pricing Concepts Expressed in Free-Response Explanations

Pricing concept	Count	Share
Marginal price	676	54.2%
Average price	144	11.5%
Displayed EOB amount	113	9.1%
Other or no stated reasoning	369	29.6%

Notes: The table reports pricing concepts expressed in respondents' open-ended explanations. Marginal-price reasoning recognizes that reaching the out-of-pocket maximum makes the next covered visit free, while average-price reasoning scales the member responsibility from the prior EOB to the new visit. Displayed-EOB reasoning carries forward the responsibility shown on the prior EOB. Other includes coinsurance or deductible reasoning and responses with no identifiable calculation rule. Categories are not mutually exclusive, so shares need not sum to 100 percent.

6.2.3 Estimating Response-Type Shares

To classify respondents into stable response types, I estimate a finite mixture model using each respondent's full pattern of answers across all four scenarios. Let i index respondents and q index survey questions. Marginal price responders anchor on the true marginal price m_{iq}, average price responders on the displayed average ratio r_{iq}, coinsurance-based responders on their assigned coinsurance rate c_i, and random responders show no systematic pattern. Each respondent's behavior is characterized by the vector $(\beta_{MP,i}, \beta_{AP,i}, \beta_{C,i})$, with the latent-type model placing probability mass on four discrete response rules:

$$\text{MP type:} \quad (\beta_{\text{MP},i}, \beta_{\text{AP},i}, \beta_{\text{C},i}) = (1, 0, 0), \tag{15}$$

$$\text{AP type:} \quad (\beta_{\text{MP},i}, \beta_{\text{AP},i}, \beta_{\text{C},i}) = (0, 1, 0), \tag{16}$$

$$\text{C type:} \quad (\beta_{\text{MP},i}, \beta_{\text{AP},i}, \beta_{\text{C},i}) = (0, 0, 1). \tag{17}$$

The population shares $\pi_{MP}, \pi_{AP}, \pi_C$, and π_{RAND} are estimated by maximizing a finite mixture likelihood over each respondent's full pattern of answers, where $y_i = (\text{belief}_{i1}, \ldots, \text{belief}_{i4})$ is respondent i's answer vector. The type shares are recovered by maximizing $\ell = \sum_i \log L_i$. The model is estimated under two fixed error standard deviations, $\sigma = 0.075$ and $\sigma = 0.10$, with results reported in Table 12.

Table 12: Estimated Response-Type Shares

Type	Fixed error SD = 0.075 Share	Fixed error SD = 0.10 Share
Marginal price	0.420	0.446
Average price	0.156	0.176
Coinsurance-only	0.046	0.060
Other / unclassified	0.378	0.318
Observations	1,248	1,248

Notes: The table reports modal classification shares from the maximum-likelihood model under two fixed values of the error standard deviation. Each respondent is assigned to the type with the highest posterior probability. Other/unclassified captures response patterns not well explained by a stable marginal-price, average-price, or coinsurance-only rule.

Table 12 reports the estimated type shares, with 42 to 45 percent of respondents classified as marginal-price types and 16 to 18 percent as average-price types across the two specifications. This average-price share is economically meaningful, suggesting that a substantial minority of respondents relies on average rather than marginal cost-sharing information even in a simplified survey setting. The other/unclassified group accounts for 32 to 38 percent of respondents across the two specifications, indicating that many respondents do not apply any stable pricing rule when interpreting EOB cost-sharing information. Appendix Table E.2 reports an extension of the mixture model in which response-type probabilities are allowed to depend on insurance literacy. Average-price reasoning is less common among respondents who answer the literacy question correctly but remains present, suggesting that insurance literacy explains some, but not all, of this behavior.

These results connect directly to the risk gradient documented in Section 3. High-risk enrollees are more likely to face claims near both cost-sharing thresholds, where average and marginal prices diverge most sharply, and average-price reasoning can therefore generate larger spending reductions than the benchmark model predicts.

7 The Myopic Perceived-Price Model

The myopic model shifts attention from future prices to the current price of care, but it leaves open which current price households perceive. Section 6.1 shows that bills often arrive after care is received, so recent claims may not yet be incorporated into the household's perceived position on the cost-sharing schedule. Section 6.2 then shows that even when cost-sharing information is observed, households do not always translate it into the marginal price implied by the contract. The model below formalizes this distinction by allowing perceived cumulative spending to differ from true cumulative spending. Perceived spending reflects only the bills that have arrived by the start of the period, and perception determines how households map that observed billing history into a spot price.

7.1 Model Setup

Let $C_{i,t-1}$ denote household i's true cumulative out-of-pocket spending at the start of period t, and let $\tilde{C}_{i,t-1}$ denote its perceived cumulative spending. True cumulative spending reflects all prior care received before period t and determines how the contract is actually applied. Perceived cumulative spending instead reflects only prior claims whose bills have arrived by the start of period t. If claim τ arrives after billing delay $L_{i\tau}$, then

$$\tilde{C}_{i,t-1} = \sum_{\tau<t:\ \tau+L_{i\tau}\leq t-1} \Delta OOP_{i\tau}. \tag{18}$$

When all bills arrive immediately, perceived and true cumulative spending coincide. With delayed billing, however, $\tilde{C}_{i,t-1}$ may lag behind $C_{i,t-1}$, so the household may believe it is in a different region of the cost-sharing schedule than its true contract position implies. At the start of each period, the perceived price is therefore based on the contract position implied by bills that have already arrived, not on the out-of-pocket payment that will be determined for the current claim.[14]

The perceived balance $\tilde{C}_{i,t-1}$ captures which prior claims have been reflected in arrived bills, but it does not by itself determine the price households perceive. I allow households to differ in their price-perception rule, denoted by $\theta_i \in \{MP, AP, COINS\}$. For marginal-price types, $\theta_i = MP$, the perceived spot price is the marginal price implied by the household's perceived position on the cost-sharing schedule. For average-price types, $\theta_i = AP$, the perceived spot price is based on the most recently arrived positive-OOP bill, measured as the ratio of out-of-pocket payment to allowed spending on that bill. Let τ^*_{it} denote that bill,

[14] Before any bill has arrived, perceived cumulative spending is zero. Under the 2016 contract, this implies a perceived price of one, even if the first claim would cross the deductible once processed, because the allowed amount and resulting out-of-pocket payment are not yet known at the time of care. Coinsurance-price types use the plan coinsurance rate by construction.

if a prior bill has arrived. Then

$$\bar{p}_{it} = \begin{cases} \dfrac{\Delta OOP_{i\tau^*_{it}}}{m_{i\tau^*_{it}}}, & \text{if such a bill has arrived,} \\ p_{i0}, & \text{otherwise.} \end{cases} \tag{19}$$

The initial price p_{i0} is the price implied by the household's initial perceived contract position before any bill has arrived. For coinsurance-price types, $\theta_i = COINS$, the perceived spot price is the plan coinsurance rate s, regardless of the household's perceived contract position.

Given these price-perception rules, the perceived spot price in period t is

$$\tilde{p}_{it} = \begin{cases} p(\tilde{C}_{i,t-1}), & \text{if } \theta_i = MP, \\ \bar{p}_{it}, & \text{if } \theta_i = AP, \\ s, & \text{if } \theta_i = COINS. \end{cases} \tag{20}$$

where $p(\tilde{C}_{i,t-1})$ is the marginal price implied by the household's perceived position on the cost-sharing schedule. Given current health need ℓ_{it}, the household chooses utilization using the same myopic decision rule as above, replacing the true spot price with the perceived spot price:

$$m_{it} = \arg\max_{m \geq 0} \left\{ h(m; \ell_{it}, \omega) - \tilde{p}_{it} m \right\}. \tag{21}$$

For the benefit function used above, this gives $m_{it} = \ell_{it}[1 + \omega(1 - \tilde{p}_{it})]$. After utilization is chosen, the household's true out-of-pocket payment ΔOOP_{it} is computed under the nonlinear contract. This true payment updates C_{it} immediately, but it affects perceived cumulative spending only when the bill arrives.

7.2 Estimation and Simulation Details

I estimate the myopic perceived-price model using the same SMM framework as the forward-looking benchmark. The latent health-need distribution, claim-arrival process, claim-size allocation, and targeted moments are unchanged from the myopic model introduced in Section 5. The perceived-price model differs only in how the price entering each utilization choice is constructed. Rather than responding to the true spot price, households respond to the price they infer from the bills that have arrived, according to their price-perception type.

Billing delay is implemented by assigning each simulated positive claim week a delay drawn from the empirical distribution of observed bill-arrival lags. To match delays to simulated claims, I group enrollees into 100 bins by observed baseline risk and group positive claim weeks into 100 bins by simulated latent claim size. Each simulated claim is then assigned a delay drawn from the empirical delay distribution for the corresponding baseline-risk and claim-size cell. By matching on risk and claim size, the simulation captures systematic differences in billing delays across enrollees and claims, but still preserves residual variation in bill timing. The claim is applied to the true contract when care occurs, but it enters the household's perceived balance only when the simulated bill arrives.

Price-perception types are assigned using the estimated shares in Table 12. Because the structural model includes only marginal-price, average-price, and coinsurance-price responders, I exclude the noise category and renormalize the remaining shares to sum to one. These shares are held fixed during estimation and used to randomly assign each simulated enrollee a price-perception type that remains fixed across years. Thus, the SMM estimates the same continuous structural parameters as in the myopic model, with price-perception heterogeneity affecting behavior only through the perceived spot price.

7.3 Model Fit with Perceived Prices

I next present the parameter estimates from the perceived-price model. Table 13 reports estimates under the homogeneous- and heterogeneous-ω specifications. In the homogeneous-ω specification, the estimate of ω is 0.087, compared with 8.007 under the perfect-foresight model. The smaller estimate reflects the larger and more persistent price differences implied by the perceived-price model, as billing delays and price-perception rules can keep perceived prices different across contracts even after true marginal prices have converged. The heterogeneous-ω estimates imply that mean price responsiveness remains low across most of the baseline risk distribution and rises to about 0.42 at the 99th percentile, still far below the estimate under perfect foresight. Appendix Figure F.1 plots this relationship across the baseline risk distribution.

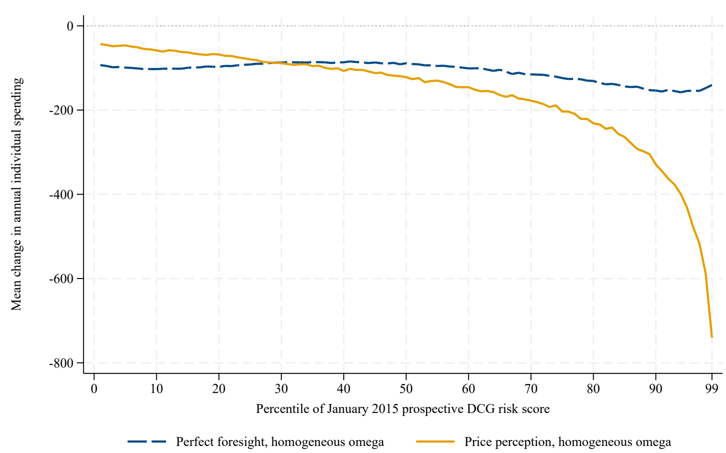
Table 14 reports the fit of the perceived-price model to the targeted moments. Both specifications closely reproduce the pre-reform spending distribution, including its mean, zero-spending share, and upper-tail moments. The perceived-price model also predicts an aggregate spending decline of about \$174, compared with \$170 in the data. The perfect-foresight model, by comparison, predicts an aggregate decline of approximately \$108. The perceived-price model therefore more closely matches the aggregate response while continuing to reproduce the pre-reform spending distribution.

Table 13: Estimated Myopic Perceived-Price Model Parameters

	Estimate	
Parameter	Homogeneous-ω	Heterogeneous-ω
ω	0.087	–
ρ_0	–	-7.281
ρ_1	–	4.133
ρ_2	–	-0.723
σ_ω	–	0.152
β_0	7.212	7.209
$\beta_{\ln(DCG)}$	0.790	0.815
$\beta_{\ln(DCG)^2}$	0.069	-0.007
$\bar{\mu}$	2.423	2.425
σ_λ	1.331	1.346

Notes: The table reports estimated parameters for the homogeneous- and heterogeneous-ω perceived-price models. In the heterogeneous-ω specification, the coefficients ρ_0, ρ_1, and ρ_2 govern how mean price responsiveness varies with standardized log baseline risk, while σ_ω captures residual heterogeneity in responsiveness among households with the same baseline risk. Price-perception type shares are held fixed at the estimates reported in Table 12. Estimation uses non-preventive spending because preventive care faces zero cost sharing under both contracts.

Figure 9: Predicted Spending Responses under Perfect Foresight and Myopic Perceived Prices



Notes: The figure plots the simulated change in annual spending from 2015 to 2016 by percentile of the January 2015 prospective DCG risk score. It compares the homogeneous-ω perfect-foresight model with the homogeneous-ω myopic perceived-price model. Negative values indicate lower simulated spending under the 2016 contract relative to the 2015 contract. The figure is shown through the 99th percentile. All spending measures exclude preventive care.

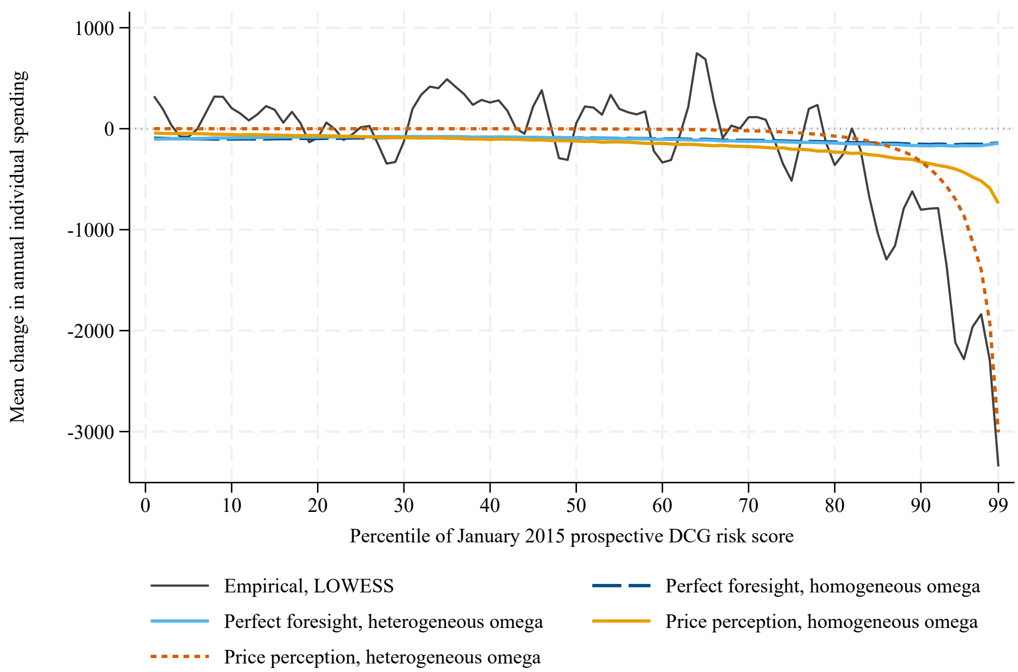
Table 14: Targeted Moment Fit: Myopic Perceived-Price Models

Moment	Data	Homogeneous ω	Heterogeneous ω
Panel A. Pre-reform spending distribution			
Mean spending, 2015	4,805.9	4,757.5	4,741.9
Standard deviation of spending, 2015	13,587.0	13,647.7	13,659.8
Zero-spending share, 2015	0.104	0.104	0.103
Share reaching out-of-pocket maximum, 2015	0.189	0.201	0.199
Mean spending, 2015, 75th–90th risk percentiles	6,974.8	6,983.4	7,013.8
Mean spending, 2015, top 10% of risk	17,345.9	17,376.8	17,362.1
Panel B. Spending response to the reform			
Regression-adjusted spending change, 2016–2015	-170.1	-174.0	-173.9
Regression-adjusted spending change, 75th–90th risk percentiles	-400.6	-251.6	-124.4
Regression-adjusted spending change, top 10% of risk	-1,958.5	-547.2	-1,460.0

Notes: The table reports the empirical moments and their simulated counterparts under the homogeneous- and heterogeneous-ω perceived-price specifications. Spending amounts and spending changes are measured in dollars. Baseline risk groups are defined using predicted spending before the reform. Spending-change moments are regression-adjusted. All moments use non-preventive spending because preventive care faces zero cost sharing under both contracts.

The remaining spending-response moments in Panel B show that the perceived-price model generates a response gradient even when price responsiveness is held constant across households. The predicted spending reduction increases from approximately \$252 among households in the 75th–90th risk percentiles to \$547 among those in the top 10 percent. Figure 9 shows that the same pattern extends across the full baseline risk distribution. Predicted spending reductions become progressively larger with risk under the perceived-price model but remain relatively flat under perfect foresight. The model nevertheless understates the empirical gradient, as the corresponding spending reductions in the data are \$401 and \$1,959. The remaining gap is particularly pronounced at the top of the risk distribution, pointing to a role for heterogeneity in price responsiveness.

Figure 10: Observed and Model-Predicted Spending Responses by Baseline Risk



Notes: The figure plots the empirical and simulated changes in annual spending from 2015 to 2016 by percentile of the January 2015 prospective DCG risk score. The empirical series shows the LOWESS-smoothed relationship between the spending change and baseline risk. The simulated series compare the homogeneous- and heterogeneous-ω specifications of the perfect-foresight and myopic perceived-price models. Negative values indicate lower spending under the 2016 contract relative to the 2015 contract. The figure reports results through the 99th percentile. All spending measures exclude preventive care.

Allowing ω_i to vary across households and with baseline risk shifts the predicted spending response toward the highest-risk households. Heterogeneity in responsiveness has a larger effect in the perceived-price model because billing delays and price-perception rules can preserve differences in perceived prices across contracts for high-risk households. By contrast, under perfect foresight, many high-risk households expect to reach the out-of-pocket maximum under both contracts, so even highly responsive households face little marginal-price difference to respond to. Among households in the top 10 percent of risk, the predicted spending reduction increases from approximately \$547 under the homogeneous-ω perceived-price model to \$1,460 under the heterogeneous-ω model, compared with \$1,959 in the data. Among households in the 75th–90th risk percentiles, however, the predicted spending reduction falls from approximately \$252 to \$124, compared with \$401 in the data. The improvement in fit is therefore concentrated at the very top of the risk distribution rather than extending throughout its upper range, while the predicted aggregate decline remains essentially unchanged. Figure 10 compares predicted spending responses from the forward-looking benchmark and the perceived-price model, each estimated under homogeneous and heterogeneous price responsiveness. Both perfect-foresight specifications produce relatively flat responses, the homogeneous-ω perceived-price model generates a broad gradient, and the heterogeneous-ω perceived-price model produces additional steepening that largely reproduces the empirical pattern.

Taken together, these results suggest that the spending gradient reflects both differences in perceived-price exposure and heterogeneity in responsiveness. Billing delays and price-perception rules allow perceived price differences across contracts to persist among high-risk households, generating a broad increase in spending reductions with risk even when responsiveness is homogeneous across households. Heterogeneity in ω_i then concentrates additional response among the highest-risk households, largely reproducing the pronounced empirical steepening at the very top while continuing to match the aggregate spending decline.

8 Conclusion

The response of medical spending to cost sharing is often interpreted as a measure of moral hazard, the additional care households use when insurance lowers its price. This paper shows why that interpretation is incomplete under a nonlinear contract. Households respond to the price they perceive, and that price can differ systematically from the marginal price implied by their true position on the contract.

The Wisconsin reform demonstrates why this distinction matters. The introduction of a modest increase in cost sharing produced a small average reduction in spending but a steep gradient across health risk, with the largest reductions among the highest-risk households. High-risk households expect to reach the out-of-pocket maximum under either contract, so the reform raises their total out-of-pocket spending but leaves their end-of-year marginal price unchanged at zero. A standard forward-looking model therefore predicts that they should respond least, yet they reduce spending most in the data.

The administrative and experimental evidence points to two mechanisms that can keep perceived prices high even after true marginal prices have fallen. First, delays in claims processing are substantial and increase with health risk. Second, some households infer the price of subsequent care from the average cost share on the bill rather than the lower marginal price that applies after a claim crosses a threshold. Both mechanisms disproportionately affect high-risk households, who face longer processing delays and whose claims more frequently cross cost-sharing thresholds. The forward-looking benchmark requires substantial price responsiveness to generate the modest aggregate reduction, yet it still predicts little response among the high-risk households. By contrast, the perceived-price model matches the aggregate reduction and accounts for much of the steep risk gradient while implying substantially less underlying price responsiveness.

The main implication is that high-risk households behave as though their marginal price is higher than it ultimately is. Because they do not fully recognize that their end-of-year marginal price is likely to be low, they may reduce more care than they would if they correctly understood their coverage. The concentration of spending reductions among households with the greatest predicted medical needs makes the possibility of underuse especially concerning, particularly for those managing chronic conditions and other ongoing care needs. The clinical-category results show that reductions among high-risk households extend to preventive and mental-health spending, although the analysis cannot determine the value of the foregone care or its health consequences. Narrowing the gap between perceived and actual prices could therefore help avoid reductions driven by misunderstanding the contract rather than by its intended incentives.

The relevant transparency challenge is not simply whether prices are available, but whether households receive timely and interpretable information about the marginal price they face. The 2020 Transparency in Coverage Final Rule requires most private health plans to show enrollees their estimated out-of-pocket costs for services and their progress toward the deductible and out-of-pocket limit. These requirements are an important step toward helping households understand where they stand on their cost-sharing schedule. However, because these amounts may not reflect unprocessed claims, the position shown by the tool can still lag behind the household's true position on the cost-sharing schedule. Even when the displayed position is current, obtaining a useful estimate may remain difficult because households may not know which services to search for or which components of an episode will be billed separately. Prior work similarly finds that price-transparency tools are used infrequently and that the information they provide may differ from eventual patient liabilities or omit portions of an episode of care (Gourevitch et al., 2017; Stults et al., 2019; Horný et al., 2021b). These limitations point to faster claims processing to keep benefit accumulators current, episode-level estimates that better capture the full cost of care, and explanations of benefits that clearly distinguish the average cost share on completed care from the marginal price of subsequent care.

More broadly, the findings suggest that comparable spending control may be possible with less cost sharing and greater financial protection than a standard forward-looking model would imply. Because the standard model assumes that households correctly perceive and anticipate contractual prices, it interprets the observed spending response as evidence of price sensitivity. In the perceived-price model, however, the same spending reduction can arise from much weaker responsiveness because households behave as though their marginal price is higher than it ultimately is. A model that equates contractual and perceived prices may therefore overstate the amount of cost sharing needed to achieve a given spending target. Although determining the optimal contract requires additional evidence on the value of care that is reduced, the low estimated responsiveness suggests that more generous coverage would induce less additional spending than the standard model implies.

References

Aron-Dine, A., L. Einav, A. Finkelstein, and M. Cullen (2015). Moral hazard in health insurance: do dynamic incentives matter? *Review of Economics and Statistics 97*(4), 725–741.

Banki, D. and U. Simonsohn (2026). Confounds protocols: Collecting & analyzing participant explanations to identify confounds (and sometimes mechanisms). Working paper.

Brot-Goldberg, Z. C., A. Chandra, B. R. Handel, and J. T. Kolstad (2017). What does a deductible do? the impact of cost sharing on health care prices, quantities, and spending dynamics. *Quarterly Journal of Economics 132*(3), 1261–1318.

Cardon, J. H. and I. Hendel (2001). Asymmetric information in health insurance: evidence from the national medical expenditure survey. *RAND Journal of Economics*, 408–427.

Dalton, C. M., G. Gowrisankaran, and R. J. Town (2020). Salience, myopia, and complex dynamic incentives: Evidence from Medicare Part D. *Review of Economic Studies 87*(2), 822–869.

Diaz-Campo, C. S. (2022). Dynamic moral hazard in nonlinear health insurance contracts. Working paper, SSRN 4651013.

Dunn, A., J. D. Gottlieb, A. H. Shapiro, D. J. Sonnenstuhl, and P. Tebaldi (2024, 02). A denial a day keeps the doctor away. *The Quarterly Journal of Economics 139*(1), 187–233.

Einav, L., A. Finkelstein, S. P. Ryan, P. Schrimpf, and M. R. Cullen (2013). Selection on moral hazard in health insurance. *American Economic Review 103*(1), 178–219.

Einav, L., A. Finkelstein, and P. Schrimpf (2015, 05). The response of drug expenditure to nonlinear contract design: Evidence from medicare part d. *The Quarterly Journal of Economics 130*(2), 841–899.

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

Gelber, A. M., D. Jones, I. Lurie, and D. W. Sacks (2025). Misperceptions of nonlinear budget sets: Evidence from administrative tax data. Technical report, National Bureau of Economic Research.

Gottlieb, J. D., A. H. Shapiro, and A. Dunn (2018). The complexity of billing and paying for physician care. *Health Affairs 37*(4), 619–626. PMID: 29608348.

Gourevitch, R. A., S. Desai, A. L. Hicks, L. A. Hatfield, M. E. Chernew, and A. Mehrotra (2017). Who uses a price transparency tool? implications for increasing consumer engagement. *INQUIRY: The Journal of Health Care Organization, Provision, and Financing 54*, 0046958017709104.

Guo, A. and J. Zhang (2019). What to expect when you are expecting: Are health care consumers forward-looking? *Journal of Health Economics 67*, 102216.

Haque, W., M. Ahmadzada, H. Allahrakha, E. Haque, and D. Hsiehchen (2021, 05). Transparency, accessibility, and variability of us hospital price data. *JAMA Network Open 4*(5), e2110109–e2110109.

Ho, K. and R. S. Lee (2023). Health insurance menu design for large employers. *The RAND Journal of Economics 54*(4), 598–637.

Hoagland, A., D. M. Anderson, and E. Zhu (2022). Medical bill shock and imperfect moral hazard. *arXiv preprint arXiv:2211.01116*.

Horný, M., P. R. Shafer, and S. B. Dusetzina (2021a, 12). Concordance of disclosed hospital prices with total reimbursements for hospital-based care among commercially insured patients in the us. *JAMA Network Open 4*(12), e2137390–e2137390.

Horný, M., P. R. Shafer, and S. B. Dusetzina (2021b, 12). Concordance of disclosed hospital prices with total reimbursements for hospital-based care among commercially insured patients in the us. *JAMA Network Open 4*(12), e2137390–e2137390.

Ito, K. (2014, February). Do consumers respond to marginal or average price? evidence from nonlinear electricity pricing. *American Economic Review 104*(2), 537–63.

Keeler, E. B., J. P. Newhouse, and C. E. Phelps (1977). Deductibles and the demand for medical care services: The theory of a consumer facing a variable price schedule under uncertainty. *Econometrica 45*(3), 641–655.

Kowalski, A. E. (2015). Estimating the tradeoff between risk protection and moral hazard with a nonlinear budget set model of health insurance. *International Journal of Industrial Organization 43*, 122–135.

Lusardi, A. and O. S. Mitchell (2011). Financial literacy around the world: an overview. *Journal of Pension Economics and Finance 10*(4), 497–508.

Lusardi, A. and J. L. Streeter (2023). Financial literacy and financial well-being: Evidence from the us. *Journal of Financial Literacy and Wellbeing 1*(2), 169–198.

Marone, V. R. and A. Sabety (2022). When should there be vertical choice in health insurance markets? *American Economic Review 112*(1), 304–342.

Rees-Jones, A. and D. Taubinsky (2020, 10). Measuring "schmeduling". *The Review of Economic Studies 87*(5), 2399–2438.

Simonsen, M., L. Skipper, N. Skipper, and A. I. Christensen (2021). Spot price biases in non-linear health insurance contracts. *Journal of Public Economics 203*, 104508.

Stults, C. D., J. Li, D. L. Frosch, H. Krishnan, G. Smith-McCurdy, V. G. Jones, and A. S. Chan (2019, 12). Assessment of accuracy and usability of a fee estimator for ambulatory care in an integrated health care delivery network. *JAMA Network Open 2*(12), e1917445–e1917445.

Appendices

Appendix A Additional Details on the Empirical Setting and Spending Response

Table A.1 reports enrollment by plan type from 2015 through 2019. The traditional non-HDHP remained the dominant plan during the reform period, accounting for 98.8 percent of enrollment in 2015 and 95.8 percent in 2016. Although HDHP enrollment increased over time, the primary analysis follows individuals continuously enrolled in the non-HDHP, so the estimated spending changes are not driven by switching into the HDHP.

Table A.1: Enrollment Shares by Plan Type, 2015-2019

Plan Type	**2015**	**2016**	**2017**	**2018**	**2019**
non-HDHP	98.77	95.76	94.54	92.96	90.09
HDHP	0.96	3.94	5.09	6.63	9.34
Access Plan	0.26	0.29	0.34	0.36	0.45
Access HDHP Plan	0.00	0.01	0.03	0.06	0.11
Number of Households	61,774	61,021	60,178	59,577	58,741

Notes: The table reports the share of households enrolled in each plan type by year. The HDHP option was first offered in 2015. HDHP enrollment moderately rises over time, coinciding with an increase in employer HSA contributions in 2016. The final row reports the number of households in the sample in each year.

Table A.2 reports the non-preventive spending changes used as model targets. Spending declined by \$170 overall, with reductions of \$401 among individuals in the 75th–90th percentiles and \$1,959 among those in the top 10 percent.

Table A.2: Non-Preventive Spending Changes Used in Model Estimation

	(1) Overall spending	(2) Spending by baseline risk
Overall change, 2016–2015	-170.1^{***}	
	(49.48)	
Bottom 75%		128.8^{***}
		(43.43)
75th–90th percentiles		-400.6^{***}
		(139.84)
Top 10%		$-1{,}958.5^{***}$
		(296.59)
2015 mean: Overall	4,806	4,806
2015 mean: 75th–90th percentiles		6,975
2015 mean: Top 10%		17,346
Observations	173,974	173,974
Adjusted R-squared	0.467	0.468
Person fixed effects	Yes	Yes
Demographic and plan controls	Yes	Yes

Notes: The table reports OLS estimates of annual non-preventive allowed medical spending at the individual level. The primary sample contains 86,987 individuals observed in 2015 and 2016, yielding 173,974 person-year observations. Column (1) reports the overall 2016–2015 change. Column (2) reports the complete change within three mutually exclusive baseline-risk groups rather than differences relative to an omitted group. Baseline risk is measured using the January 2015 prospective risk score. All specifications include person fixed effects and controls for age, gender, coverage tier, family size, and health plan. Standard errors, clustered by family, are in parentheses. * $p < 0.10$, ** $p < 0.05$, and *** $p < 0.01$.

Appendix B Additional Perfect-Foresight Model Results

Appendix B.1 Claim-Arrival Process

The expected number of claim arrivals is estimated separately using the relationship between projected annual spending and the number of active claim days. Table B.1 reports the regression used to determine the claim-arrival parameters in equation 6.

Table B.1: Relationship Between Predicted Spending and Claim Days

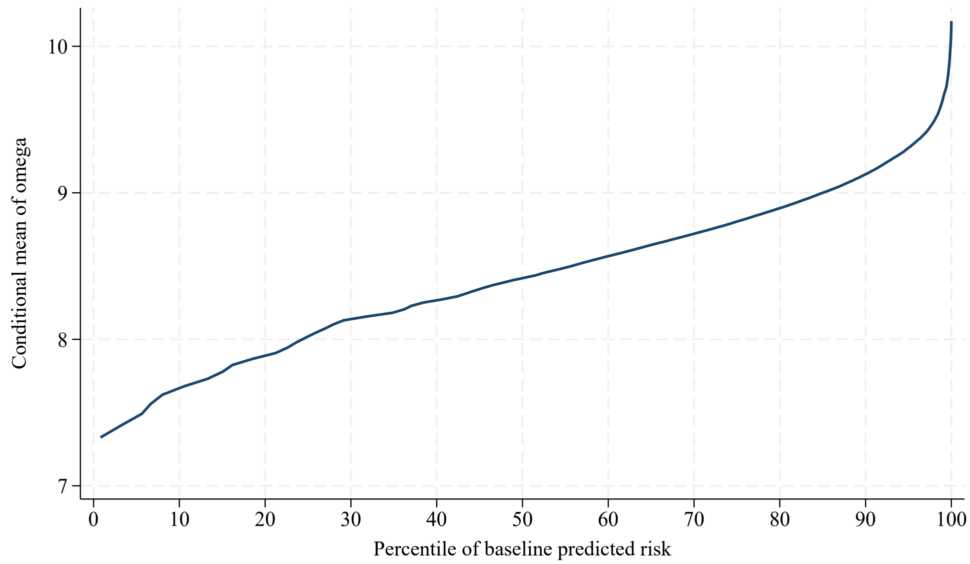
	Number of active claim days
Projected annual cost	0.0001556^{***}
	(0.00000126)
Projected annual cost squared	$-7.55 \times 10^{-10***}$
	(1.07×10^{-11})
Constant	2.6710^{***}
	(0.0087)
Observations	115,267

Notes: The table reports an OLS regression of the number of active claim days on projected annual cost and its square using 2016 claims. Standard errors are reported in parentheses. Projected annual cost is measured in dollars. $^{***}p < 0.01$.

Appendix B.2 Heterogeneity in Price Responsiveness

Figure B.1 plots the estimated conditional mean of ω_i across the baseline risk distribution under the heterogeneous-ω specification.

Figure B.1: Estimated Mean Price Responsiveness under Perfect Foresight



Notes: The figure plots the estimated conditional mean of ω_i by percentile of the January 2015 prospective DCG risk score under the heterogeneous-ω perfect-foresight specification. Higher values of ω_i indicate greater responsiveness of medical spending to the marginal price.

Appendix C Myopic True-Spot-Price Benchmark

This appendix reports estimates for the myopic true-spot-price benchmark introduced in Section 5. Households make a sequence of utilization decisions and do not account for how current spending affects future prices. The model nevertheless assumes that households know their current position on the cost-sharing schedule and respond to the true spot price of care. It therefore isolates the effect of myopia from billing delays and imperfect price perception.

Table C.1 reports the parameter estimates. The homogeneous-ω specification implies a price-responsiveness parameter of 2.002, substantially smaller than the corresponding perfect-foresight estimate. The heterogeneous-ω specification allows mean responsiveness to vary with baseline risk and permits residual heterogeneity among households with the same measured risk.

Table C.1: Estimated Myopic True-Spot-Price Model Parameters

	Estimate	
Parameter	Homogeneous ω	Heterogeneous ω
ω	2.002	–
ρ_0	–	0.537
ρ_1	–	-0.079
ρ_2	–	0.304
σ_ω	–	0.010
β_0	6.305	6.416
$\beta_{\ln(\mathrm{DCG})}$	0.926	0.956
$\beta_{\ln(\mathrm{DCG})^2}$	-0.024	-0.183
$\bar{\mu}$	2.268	2.288
σ_λ	1.164	1.176

Notes: The table reports estimated parameters for the homogeneous- and heterogeneous-ω specifications of the myopic true-spot-price model. In the heterogeneous-ω specification, ρ_0, ρ_1, and ρ_2 govern how mean price responsiveness varies with standardized log baseline risk, while σ_ω captures residual heterogeneity among households with the same baseline risk. The positive scale parameters are estimated in logs internally and reported on their natural scale. Estimation uses non-preventive spending because preventive care faces zero cost sharing under both contracts.

Table C.2 reports the model fit. Both specifications reproduce the main features of the pre-reform spending distribution and generate an aggregate spending reduction similar to that observed in the data. The model nevertheless understates the spending response among higher-risk households. In the top 10 percent of baseline risk, it predicts a reduction of $265 to $340, compared with $1,949 in the data. Thus, myopia generates a stronger risk gradient than perfect foresight but does not fully reproduce the pronounced response at the top of the risk distribution.

Table C.2: Targeted Moment Fit: Myopic True-Spot-Price Models

Moment	Data	Homogeneous ω	Heterogeneous ω
Panel A. Pre-reform spending distribution			
Mean spending, 2015	4,805.9	4,560.5	4,537.8
Standard deviation of spending, 2015	13,587.0	13,822.3	13,236.0
Zero-spending share, 2015	0.104	0.108	0.106
Share reaching out-of-pocket maximum, 2015	0.189	0.197	0.199
Mean spending, 2015, 75th–90th risk percentiles	6,974.8	7,101.2	7,127.6
Mean spending, 2015, top 10% of risk	17,345.9	17,567.4	16,818.5
Panel B. Spending response to the reform			
Regression-adjusted spending change, 2016–2015	−170.1	−156.9	−145.1
Regression-adjusted spending change, 75th–90th risk percentiles	−400.6	−171.2	−146.5
Regression-adjusted spending change, top 10% of risk	−1,958.5	−264.6	−339.8

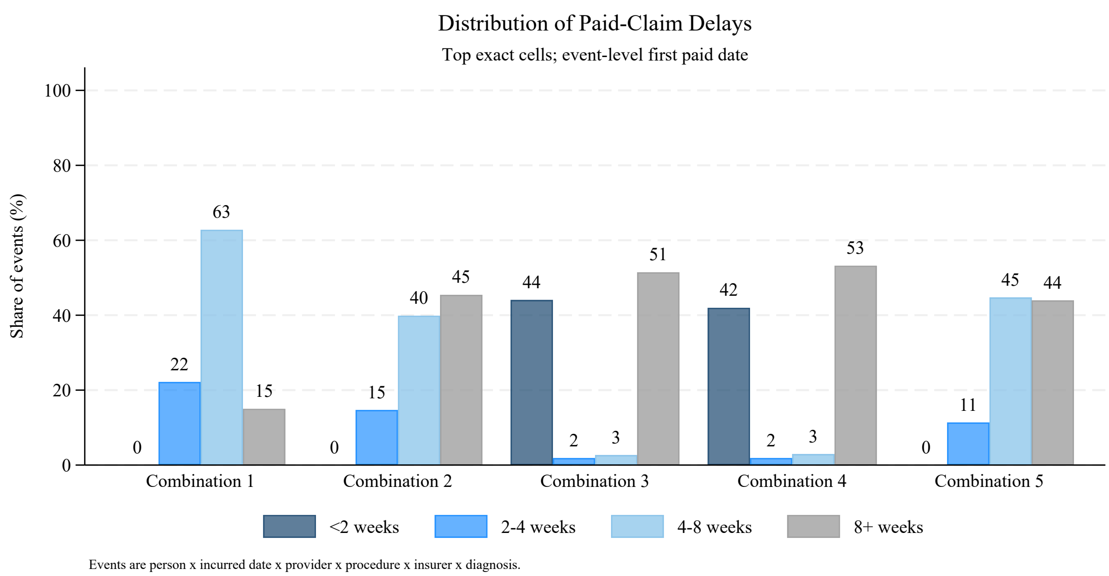
Notes: The table reports the empirical moments and their simulated counterparts under the homogeneous- and heterogeneous-ω specifications of the myopic true-spot-price model. Spending amounts and spending changes are measured in dollars. Baseline risk groups are defined using the January 2015 prospective DCG risk score. Spending-change moments are regression-adjusted. All moments use non-preventive spending because preventive care faces zero cost sharing under both contracts.

Appendix D Additional Evidence on Billing Delays and Bill Arrival

Appendix D.1 Variation in Billing Delays Within Claim Types

Figure D.1 examines whether billing delays remain dispersed after holding observable claim characteristics fixed.

Figure D.1: Variation in Billing Delays Within Claim Types



Note: The figure plots the distribution of billing delays for the five most common provider-procedure-insurer-diagnosis combinations in the 2016 sample. Each bar reports the share of claim lines within the indicated delay window. Delay is measured as the number of days between the medical service date and the paid date. Definitions of the five combinations are reported separately.

Appendix D.2 Bill-Arrival Event Study

Building on Hoagland et al. (2022), I estimate an event study centered on the insurer-paid date of the first encounter in each person-year that generates a deductible payment. Event time is measured in weeks relative to the insurer-paid date, which provides an observable proxy for when cost-sharing information becomes available. To distinguish changes around this date from the natural trajectory of care following the encounter, the specification controls flexibly for weeks since the encounter. I restrict the sample to claims paid at least five weeks after service to provide a clean pre-payment window.[15]

Let Y_{iyt} denote weekly utilization for person i in year y and week t. I estimate equation D.1, where α_i and η_t are person and calendar-week fixed effects, b_{iy} is the insurer-payment week, e_{iy} is the week of the first encounter, and standard errors are clustered at the person level.[16]

$$Y_{iyt} = \alpha_i + \eta_t + \sum_{m \neq -1} \theta_m \mathbf{1}\{t - b_{iy} = m\} + \sum_{k \geq 0} \delta_k \mathbf{1}\{t - e_{iy} = k\} + \varepsilon_{iyt}. \tag{D.1}$$

Table D.1 reports estimates for the probability of any claim and allowed spending.Relative to the week immediately before insurer payment, the probability of any claim is 0.9 to 1.1 percentage points lower in weeks 1 through 3 after payment. The allowed-spending estimates are also negative over these weeks,ranging from approximately \$13 to \$29, although they are less precisely estimated. The coefficients immediately preceding payment do not indicate a steady pre-payment trend. Week -2 is statistically insignificant for both outcomes, while week -3 is positive and significant for the probability of any claim but insignificant for allowed spending. The endpoint bin at $m \leq -4$ is negative and significant but pools all earlier weeks and therefore cannot be attributed to a particular pre-payment trend.

[15]The five-week minimum provides at least four clean pre-payment weeks. With shorter billing delays, the encounter and insurer-payment dates are too close to separate changes associated with the paid date from the natural dynamics of care following the encounter.

[16]The paid-date event-time indicators include endpoint bins at $m \leq -4$ and $m \geq 4$, so all coefficients are identified relative to week -1. The encounter-time controls include a full set of week-by-week indicators from the encounter through the end of the coverage year.

Table D.1: Bill-Arrival Event Study

	(1) Any claim	(2) Allowed amount
Bill event time ≤ -4	-0.020^{***}	-65.50^{***}
	(0.002)	(11.57)
Bill event time -3	0.006^{**}	12.33
	(0.002)	(14.67)
Bill event time -2	0.003	5.806
	(0.002)	(12.98)
Bill event time 0	-0.007^{***}	5.349
	(0.002)	(16.08)
Bill event time 1	-0.009^{***}	-25.81^{*}
	(0.002)	(13.39)
Bill event time 2	-0.011^{***}	-13.28
	(0.002)	(15.97)
Bill event time 3	-0.009^{***}	-28.98^{**}
	(0.002)	(13.56)
Bill event time ≥ 4	-0.015^{***}	-50.12^{***}
	(0.002)	(12.98)
Observations	1,183,298	1,183,298
Person FE	Yes	Yes
Week FE	Yes	Yes
Encounter-time controls	Yes	Yes

Notes: The table reports estimates of equation D.1. The sample consists of person-weeks from 2016 and 2017 for which the first deductible-applying claim was paid at least five weeks after the encounter. The omitted category is paid-date event time = −1. Paid-date event-time indicators include endpoint bins at $m \leq -4$ and $m \geq 4$. Column (1) uses an indicator for having any claim in the person-week as the outcome. Column (2) uses the dollar amount of allowed spending in the person-week as the outcome. Encounter-time controls consist of a full set of week-by-week indicators from the encounter through the end of the coverage year. Standard errors are clustered at the person level. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Appendix E Additional Evidence from the EOB Experiment

Appendix E.1 EOB Example and Response Distributions

Figure E.1 presents an example EOB statement from the State of Wisconsin employee health plan. The EOB reports the claim's allowed amount and the member's responsibility, making the average price paid on the completed claim salient.

Figure E.2 plots the distribution of respondents' expected out-of-pocket costs across the four experimental scenarios. In the below-deductible and within-coinsurance scenarios, the marginal and average prices coincide, so responses are expressed as deviations from the common benchmark. In the two threshold-crossing scenarios, responses are normalized as

$$\frac{\text{stated cost} - \text{marginal-price benchmark}}{\text{average-price benchmark} - \text{marginal-price benchmark}}.$$

Under this normalization, zero corresponds to the marginal-price benchmark and one corresponds to the average-price benchmark. The crossing-scenario distributions show substantial mass at the marginal-price benchmark but also distinct mass at the average-price benchmark. Responses elsewhere in the distribution reflect other pricing rules.

Appendix E.2 Additional Statistical Results

Table E.1 regresses respondents' stated expected out-of-pocket costs on the marginal- and average-price benchmarks. The specifications pool all scenarios, add respondent fixed effects, and restrict the sample to the two threshold-crossing scenarios. The positive average-price coefficient across specifications provided additional evidence that respondents use the average price implied by the prior EOB.

I next extend the finite-mixture model by allowing the probability of belonging to each response type to depend on insurance literacy. Table E.2 reports the resulting literacy-dependent response-type distributions. The model is estimated jointly using each respondent's answers across all four experimental scenarios. The first two columns report the model-implied response-type distribution among respondents who answered the insurance-literacy question incorrectly and correctly, respectively. The Overall column averages these literacy-dependent probabilities using the sample shares of the two literacy groups.

Figure E.1: Example Explanation of Benefits Statement

Page: 1 of 1

EXPLANATION OF BENEFITS (EOB)

Policyholder:
Group #:
Patient:
Member #:
Claim #:
Provider of Service:
Account #:
Processed Date:

Total Member Responsibility for this Claim (see below for details)	$41.84

This is not a bill. Please remit payment to your provider of service upon receipt of an invoice if you have not previously paid.

THIS IS NOT A BILL

Important Information -- Please Read
This document serves as notice of a benefit determination. If we have declined to provide benefits, in whole or in part, for the requested treatment or service described below, and you think this determination was made in error, you have the right to appeal (see the back of this page for information about your appeal rights). If you suspect fraud please call us toll-free at XXX-XXX-XXXX.

*This claim is from a participating provider. The "Amount Not Covered" below *without* any remark codes are *not* your responsibility to pay. Any amount listed in the copay, coinsurance, or deductible columns remain your responsibility. Your insurance carrier has made payment to the provider for the amount listed in the Paid/Capitated column.

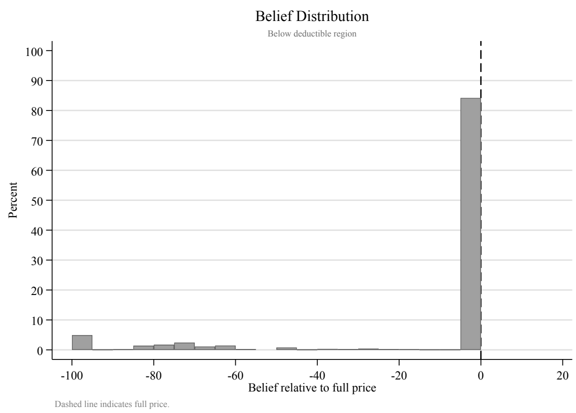
Service Code and Description	Date of Service	Amount Billed	Amount Allowable	Copay	Co-ins	Deductible	Amount COB	*Amount Not Covered	Remark Code	Paid/ Capitated
98940 CHIROPRACTIC MANIPULATIVE TRE		$72.00	$41.84	$0.00	$0.00	$41.84	$0.00	$30.16		$0.00
	TOTALS:	$72.00	$41.84	$0.00	$0.00	$41.84	$0.00	$30.16		$0.00

Accumulation Information - Data shown below is as of the EOB print date	
Benefit Accum Code	**Benefit Accumulation Description**
00500	YOU HAVE MET $103.94 OF YOUR $1,500.00 SINGLE DEDUCTIBLE
00501	YOU HAVE MET $103.94 OF YOUR $1,500.00 FAMILY DEDUCTIBLE
00504	YOU HAVE MET $103.94 OF YOUR $2,500.00 SINGLE OUT OF POCKET MAXIMUM
00505	YOU HAVE MET $103.94 OF YOUR $2,500.00 FAMILY OUT OF POCKET MAXIMUM

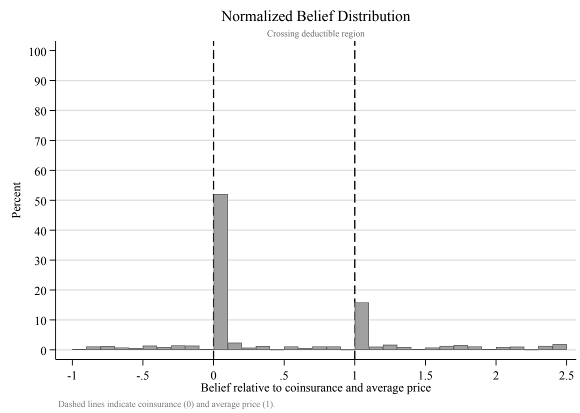
UH01171 (0817)

Note: The figure shows a sample Explanation of Benefits (EOB) statement from the State of Wisconsin employee health plan.

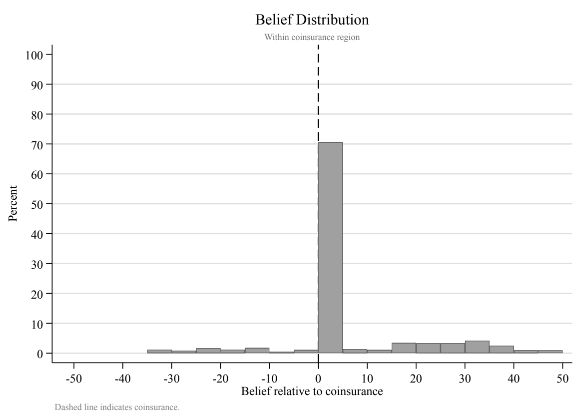
Figure E.2: Distributions of Respondents' Expected Out-of-Pocket Costs



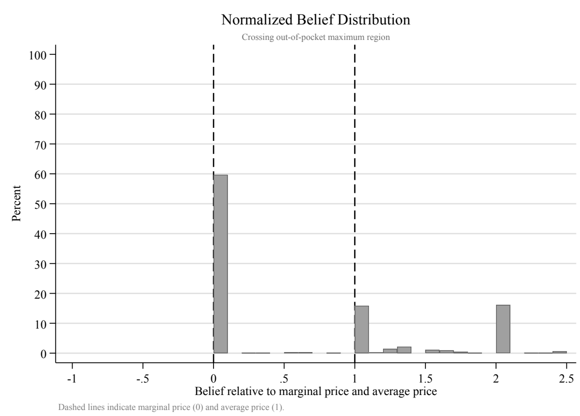
(a) Below deductible



(b) Crossing deductible



(c) Within coinsurance



(d) Crossing out-of-pocket maximum

Note: The figure plots the distribution of respondents' stated expected out-of-pocket cost for a future \$100 visit across the four insurance-cost scenarios. In the below-deductible and within-coinsurance scenarios, the x-axis reports the deviation of the response from the marginal-price benchmark. In the crossing-deductible and crossing out-of-pocket-maximum scenarios, responses are normalized so that the marginal-price and average-price benchmarks can be compared on a common scale. Vertical reference lines mark the benchmark values used for classification in the main analysis.

Table E.1: Belief Regressions

	(1) Pooled	(2) Respondent FE	(3) Crossing only
Marginal price	0.426***	0.417***	0.181
	(0.064)	(0.070)	(0.479)
Average price	0.367***	0.380***	0.617*
	(0.059)	(0.070)	(0.365)
Observations	4,992	4,992	2,496
R-squared	0.228	0.543	0.565
Respondent FE	No	Yes	Yes
Sample	All scenarios	All scenarios	Crossing scenarios only

Notes: The outcome is the respondent's expected out-of-pocket cost in dollars for a future $100 visit. Standard errors are clustered at the respondent level and reported in parentheses. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

Table E.2: Estimated Response-Type Probabilities by Insurance Literacy

Response type	Incorrect	Correct	Overall
Panel A. Fixed error standard deviation = 0.075			
Marginal price	0.202	0.501	0.419
Average price	0.208	0.129	0.151
Coinsurance-only	0.053	0.038	0.042
Other / unclassified	0.536	0.332	0.388
Panel B. Fixed error standard deviation = 0.10			
Marginal price	0.225	0.529	0.445
Average price	0.242	0.146	0.172
Coinsurance-only	0.072	0.051	0.057
Other / unclassified	0.461	0.274	0.325
Observations	343	905	1,248

Notes: The table reports estimated response-type probabilities from finite-mixture models using each respondent's answers across all four survey scenarios. Respondents are classified as insurance literate if they correctly answer that a consumer with $2,000 in total medical spending under a plan with a $1,000 deductible, 20 percent coinsurance rate, and $4,000 out-of-pocket maximum would pay between $1,000 and $4,000 out of pocket. All other answers, including "I don't know," are classified as incorrect. The Overall column averages the literacy-dependent probabilities over the sample distribution of insurance literacy. Panels A and B report results under fixed error standard deviations of 0.075 and 0.10, respectively.

Appendix E.3 Free-Response Classification Prompt

Respondents' open-ended explanations were classified using the OpenAI API with GPT-5.6. The classification prompt was developed iteratively with assistance from ChatGPT and revised by the author. All responses were classified using the same prompt and model settings. The full prompt is reproduced below.

Prompt Used for Free-Response Classification

```
You are coding participants' open-ended explanations of how they calculated
Robert's expected out-of-pocket cost for a future $100 doctor visit.

PURPOSE

Identify which pricing concepts participants express in their explanations.
Do not classify respondents into stable types, judge whether their original
numerical answers were correct, or create categories beyond the five
indicators defined below.

INPUT

Each row contains:

- prolific_pid: respondent identifier.
- reasonforchoice: the respondent's written explanation. Some respondents
  entered only a number.
- coinsurance: the coinsurance percentage assigned to the respondent. Because
  the future visit costs $100, mechanically applying the coinsurance percentage
  produces the same dollar amount.
- moop_gap_pp: one-half of the member responsibility displayed on the prior
  EOB. The prior EOB is for a $200 visit. Thus the displayed member
  responsibility equals 2 * moop_gap_pp, while proportionally scaling that
  responsibility to a $100 visit produces moop_gap_pp.

SCENARIO

Robert's health insurance plan has a $1,500 deductible, the respondent's
assigned coinsurance rate, and a $2,500 out-of-pocket maximum.

The displayed EOB is for a $200 medical visit. It indicates that, after this
visit, Robert has met both his $1,500 deductible and his $2,500 out-of-pocket
maximum.
```

```
Respondents were asked how much Robert would expect to pay out of pocket for
another covered $100 doctor visit the following day, within the same plan year.
Because Robert has met the out-of-pocket maximum, the marginal out-of-pocket
cost of the next covered visit is $0.

GENERAL RULES

1. Code concepts expressed in reasonforchoice.

2. The first four indicators are not mutually exclusive. If an explanation
   clearly expresses more than one pricing concept, code every applicable
   indicator as 1.

3. Do not infer a pricing concept solely because a numerical response matches
   a benchmark. Numeric-only responses are coded as unclear or no stated
   reasoning.

4. If a respondent describes an initial calculation and then a different
   calculation after reconsidering the EOB, code every pricing concept
   explicitly described.

5. Do not create categories such as confused, self-corrected, mixed, other,
   correct, or incorrect.

INDICATOR 1: oop_max_awareness

Set oop_max_awareness = 1 when the explanation recognizes that meeting the
out-of-pocket maximum makes the next covered visit free, fully covered, or
subject to no further out-of-pocket payment.

Examples:
- "He met his out-of-pocket maximum, so insurance covers 100%."
- "He has already paid the maximum and owes nothing for the next visit."
- "The next visit should be free because he reached $2,500."
- "He has met his out-of-pocket maximum."
- "Already maxed."

A concise statement that the out-of-pocket maximum has been met can count when
it is clearly offered as the reason Robert will not owe additional money.

Set oop_max_awareness = 0 when the respondent:
```

```
- provides only the number 0;
- mentions the out-of-pocket maximum but concludes that Robert still owes
  coinsurance, a copay, or another positive amount;
- treats the $2,500 maximum as an amount paid by the insurer; or
- mentions the maximum without showing or clearly implying that it eliminates
  the cost of the next visit.

INDICATOR 2: average_price_scaling

Set average_price_scaling = 1 when the explanation scales the member
responsibility from the prior $200 EOB to the future $100 visit.

This includes:
- dividing the previous member responsibility by two;
- saying that the new visit costs half as much, so Robert pays half as much;
- calculating the cost share shown on the previous EOB and applying it to $100;
- assuming Robert pays the same effective rate as on the previous claim.

Examples:
- "He paid $30 for a $200 visit, so he would pay $15 for a $100 visit."
- "The new visit costs half as much, so I divided the prior amount by two."
- "I calculated the percentage he paid previously and applied it to $100."
- "I used the same rate as the prior claim."

Do not set average_price_scaling = 1 merely because the response equals
moop_gap_pp. The explanation must express proportional scaling or reliance on
the average cost share from the prior EOB.

INDICATOR 3: displayed_eob_anchoring

Set displayed_eob_anchoring = 1 when the explanation carries forward the member
responsibility displayed on the prior EOB without proportionally adjusting it
for the new visit.

This includes:
- copying the displayed member responsibility;
- treating the displayed amount as the amount owed for every visit;
- interpreting it as a fixed copay or recurring fee;
- using the amount because it appears under "member pays," "member
  responsibility," or "patient responsibility."

Examples:
```

```
- "That is what it says on the bill."
- "The EOB says member pays $20, so I used $20."
- "He paid $40 last time, so I assumed he would pay $40 again."
- "I assumed the displayed amount was a copay."
- "The EOB still charged $10, so I assumed it was a per-visit fee."

If the respondent proportionally scales the displayed amount, code
average_price_scaling rather than displayed_eob_anchoring. Code both only if
the explanation clearly expresses both approaches.

INDICATOR 4: coinsurance_or_deductible_rule

Set coinsurance_or_deductible_rule = 1 when the explanation applies coinsurance
or deductible logic without correctly applying the out-of-pocket maximum.

This includes:
- multiplying the assigned coinsurance percentage by $100;
- saying Robert must pay the assigned coinsurance rate;
- saying the deductible has been met and therefore coinsurance applies;
- treating coinsurance as continuing after the out-of-pocket maximum is met.

Examples:
- "Twenty-five percent of $100 is $25."
- "He pays 30% coinsurance."
- "The deductible has been met, so now he pays coinsurance."
- "After meeting the deductible, he is responsible for 20% of the visit."
- "He met the out-of-pocket maximum but still has to pay coinsurance."

Do not set this indicator to 1 when the respondent mentions coinsurance only to
explain that it no longer applies after the out-of-pocket maximum has been met.

INDICATOR 5: unclear_or_no_stated_reasoning

Set unclear_or_no_stated_reasoning = 1 only when none of the first four pricing
concepts can be identified.

This includes:
- numeric-only responses, even when the number matches a benchmark;
- blank or effectively blank responses;
- random characters;
- "I guessed" without further explanation;
- "I don't know" or "I was confused" without substantive reasoning;
```

```
- vague statements such as "I looked at the document" that do not identify a
  calculation rule;
- explanations whose meaning cannot be determined.

Set unclear_or_no_stated_reasoning = 0 whenever at least one of the first four
pricing concepts is identifiable, even if the reasoning is incomplete or
produces an incorrect answer.
```

Appendix F Additional Perceived-Price Model Results

Figure F.1 plots the estimated conditional mean of ω_i across the baseline risk distribution under the heterogeneous-ω myopic perceived-price specification. Estimated mean responsiveness remains low across most of the risk distribution but rises sharply near the top, reaching approximately 0.42 at the 99th percentile. This pattern helps the model generate the additional spending response observed among the highest-risk households. Even at the top of the risk distribution, however, estimated responsiveness remains substantially below the corresponding perfect-foresight estimate.

Figure F.1: Estimated Mean Price Responsiveness by Baseline Risk