What's in a Bill? A Model of Imperfect Moral Hazard in Healthcare*

David M. Anderson† Alex Hoagland‡ Ed Zhu§

November 2, 2022

Abstract

Models of consumer responsiveness to medical care prices are central in the optimal design of health insurance. However, consumers are rarely given spending information at the time they consume care, a delay which many models do not account for and may create distortions in the choice of future care consumption. We study household responses to scheduled medical services before and after pricing information arrives, leveraging quasi-experimental variation in the time it takes for an insurer to process a bill. Immediately after scheduled services, households increase their spending by roughly 46%; however, a bill's arrival causes households to reduce their spending by 7 percentage points (nearly 20% of the initial increase). These corrections are concentrated among households who are under-informed about their deductible or whose expenditures fall just short of meeting the deductible, suggesting that a bill provides useful pricing information. These corrections occur across many types of health services, including high- and low-value care. We model household beliefs and learning about expenditures and find that households overestimate their expenditures by 10% prior to a bill's arrival. This leads to an over-consumption of \$842.80 (\$480.59) for the average (median) affected household member. There is evidence that households learn from repeated use of medical services; however, household priors are estimated to be as high as 80% larger than true prices, leaving low-spending households especially under-informed. Our results suggest that long deductible periods coupled with price non-transparency may lead to an over-consumption of medical care.

Keywords: *Ex-post* moral hazard, price transparency, learning, low-value care
JEL codes: I12, I13, D01, D90

*The authors are grateful to participants in the APPAM 2022 Research Conference, as well as seminar participants at Brigham Young University, Duke University, and the University of Toronto for useful feedback.

†Duke University, Duke-Margolis Center for Health Policy, 100 Fuqua Drive, Durham, North Carolina, USA. Email: dma34@duke.edu.

‡Corresponding author. University of Toronto, Institute of Health Policy, Management, and Evaluation, 155 College St, Toronto, Ontario, Canada. Email: alexander.hoagland@utoronto.ca. Website: alex-hoagland.github.io.

§Boston University, Department of Economics, 270 Bay State Road, Boston, MA 02215. Email: edzhu1@bu.edu.

1 Introduction

The provision of health insurance plays a vital role in protecting consumers against the risk of volatile, unpredictable health shocks. However, incomplete information plagues health insurance markets, ultimately leading both public institutions (e.g., governments) and private organizations (e.g., insurers) to provide sub-optimal coverage (Einav and Finkelstein, 2018; Dave and Kaestner, 2009). One of the main sources of this incomplete information is "*ex-post* moral hazard," which captures how an individual's healthcare consumption is responsive to changes in the price they face for care.[1]

Consumer price responsiveness ultimately justifies some consumer cost-sharing out-of-pocket (OOP) (Chandra et al., 2010; Goldman and Philipson, 2007; Johnson et al., 1997). Many studies have found evidence that consumers are price-responsive in their demand for health care, with health consumption increasing as individuals move into regions of their insurance contract with lower marginal costs (Kowalski, 2016; Duarte, 2012; Dunn, 2016). Subsequently, the rate of consumer exposure to cost-sharing has been increasing steadily over the last two decades in the United States, nominally to limit the consumption of excessive or low-value services (Geyman, 2012). However, the optimal level of cost-sharing depends in large part on how consumers respond to changes in the nonlinear pricing scheme of their insurance plan, including dynamic incentives such as their expectations of their end-of-year price (Aron-Dine et al., 2015; Baicker et al., 2015).

However, while increasing exposure to cost-sharing—for example, through increasing enrollment in high-deductible health plans (HDHPs)—may reduce overall levels of consumption, individuals report being confused about the structure of their insurance contract and reducing consumption of even high-value preventive health services which should be free to them, ultimately making the plans a blunt instrument (Reed et al., 2009; Agarwal et al., 2017; Mazurenko et al., 2019; Brot-Goldberg et al., 2017).[2] In addition, previous work has failed to find large discontinuities in consumption around the transition points of a contract, such

[1]Ultimately, *ex-post* moral hazard captures the extent to which one's demand curve for healthcare slopes downward. Note that the use of the term "moral hazard" to refer to elastic demand responses for medical care utilization is a slight abuse of notation that is now widely used in this literature, beginning with Arrow (1963). Although consistent with the notion of "hidden action" in the sense that individuals may consumer more care as they exert less effort in health production, we will refer to moral hazard more as the extent to which, conditional on health status, individuals adapt their care consumption decisions to the spot prices of care (Pauly and Blavin, 2008; Cutler and Zeckhauser, 2000)

[2]During the study's time frame, value-based insurance designs—where certain high-value services are carved out of cost-sharing obligations for consumers—have become more prevalent (Chernew et al., 2007). This includes certain cost-sharing exemptions promoted by the Affordable Care Act (ACA). These designs have been shown to increase quality of care without changing spending decisions (Lee et al., 2013); however, recent work has found that lack of clear guidelines in these policies may result in confusion about cost-sharing exemptions, ultimately harming consumers (Hoagland and Shafer, 2021; Shafer et al., 2021).

as the deductible (Dalton, 2014). This lack of a sharp response—which would be predicted in a standard model of moral hazard where only spot prices influence decision-making—has been previously explained either as evidence of forward-looking consumers (Aron-Dine et al., 2015) or by noting various frictions in the consumption of health care, including poor understanding of health plan design (Loewenstein et al., 2013). Health insurance literacy is low among many Americans (Edward et al., 2019), and the increasing complexity of insurance contracts, coupled with price intransparency, may lead to worsening consumer predictions of incurred but not yet received (IBNR) spending and OOP cost-sharing.

In this paper, we highlight an overlooked feature of many healthcare systems with significant implications for models of *ex-post* moral hazard and cost-sharing: lack of timely pricing information. Consumers are rarely, if ever, given information about the total costs of their health services at the point of service, and are even less likely to be told their expected OOP contribution then. Although patients value price transparency and would like to know OOP costs before agreeing to a service (Henrikson et al., 2017), consumers in our sample wait an average of 4.1 weeks to be given detailed pricing information in a medical bill, including both the portion their insurer has elected to pay and the portion that falls to them. During the waiting period, consumers must form expectations about their already realized expenses when making future care decisions, a nontrivial task given that there is substantial variation in the price of even basic health services (Gruber, 2022; Cooper et al., 2019).[3] We argue that this mistimed pricing information limits the effectiveness of cost-sharing in mitigating over-consumption, particularly in the case of HDHPs.

We study how households with employer-sponsored insurance (ESI) in the US make medical spending decisions after a significant health expenditure for one of their household members. Specifically, we assess household responses following the use of a health service classified as "shoppable" by the Centers for Medicare & Medicaid Services (CMS), services which are expected to constitute a nontrivial but unknown amount of OOP spending for the household (CMS, 2019). We identify the causal effect of a bill's arrival in changing those decisions, and further explore the mechanisms generating observed responses. Before the bill arrives, households must make decisions based on their expectations of OOP spending, conditional on their underlying health needs and risk aversion. It may be that risk averse households avoid increasing their consumption even after meeting a deductible, as they are unsure that they have actually done so and hesitant to incur additional costs. On the other hand, households may over-predict OOP spending, leading them to respond in the short run by over-consuming care. Hence, our work is related to previous work identifying the strategic delay of services such as dental care (Cabral, 2017). In particular, we highlight

[3]Appendix Figure A.7 illustrates some of the variation in prices for common services in our sample.

the importance of corerct household beliefs about spending when making strategic decisions, and show that this type of behavior applies to a far broader and potentially more expensive set of hard to predict events across the healthcare system.

We identify the causal parameter of interest—the impact of pricing information on consumption decisions—based on exogenous variation in the time it takes insurers to receive and process bills, which affects the length of the household's "interim period". Appendix Figure A.8 illustrates the substantial variation in how long it takes an insurer to receive and process a claim, both within and across years. The average household waits 4.1 weeks for a bill from a shoppable service; however, waiting times tend to be higher at the beginning of a calendar year and the first month of each quarter, when insurers have billing changes and new policies to incorporate into their processing algorithms.[4] Waiting times are also affected by other time-varying features of the healthcare system that are exogenous to the household, including the rate at which physicians submit claims to insurers for reimbursement. As this variation is insurer-specific and unpredictable by a household, the length of time a household will wait for a bill is exogenous at the consumer level.

Using a triple-differences regression design, we estimate how scheduled healthcare consumption generates distinct household consumption spillovers before and after the bill's arrival. We find that, in the interim period between the service and its bill, household members increase their total health spending by about 46.4% ($55.91 per person per week). However, once the bill arrives, consumption drops significantly by about one-fifth of that increase, a reduction of $9.28 per person per week. These results are largest for households with more limited information before the bill's arrival, including households who have consumed none or few services prior to the shoppable service, as well as those for whom the shoppable service was expected to be particularly expensive. This suggests that the effects of a bill are largely the result of eliminating information frictions regarding the price of care. Finally, we examine heterogeneity in household responses across a spectrum of services, and find that households are most likely to over-respond in their use of elastic care, such as preventable hospitalizations and general practitioner visits or lab services.

These findings are consistent with a model of imperfect moral hazard, in which households learn to form expectations of their OOP spending as bills arrive following a delay. We write and estimate a model of consumer learning in the face of delayed information, where consumers experience a lag between consuming a good and learning how consumption affected future spot prices. We use the exogenous variation in our data—now widening the

[4] Waiting times are also affected by more general health policies, such as the national transition to the International Classification of Diseases, 10th Revision, Clinical Modification (ICD-10-CM), in October 2015. This transition increased billing complexity by roughly five times and, subsequently, the rate of administrative frictions in processing billing information (Caskey et al., 2014).

scope of analysis to include all health services consumed by the household within a year—to separately identify average beliefs about OOP spending as well as how households learn from health events over time.

Consistent with the reduced-form regression results, our model predicts that in the absence of learning, households over-inflate their OOP spending by 11% without pricing information. As a result, 10.5% of households in our sample over-consume health services specifically because they believe erroneously that they have already met their deductible, with the average (median) affected household spending \$842.80 (\$480.59) more per household member than they would have selected under full information. We do find evidence of consumer learning: at the beginning of the year, households over-inflate their expectations of OOP spending by an average of 80%, an over-estimate which is quickly corrected as bills for medical services provide new information.

Taken together, our findings make several important contributions to a discussion of moral hazard in health care. First, we present the first model which incorporates price uncertainty into demand for health care and highlights its implications for discussions of *ex-post* moral hazard. Our model is related to literature on learning about prices in microeconomic theory more broadly, including uncertain prices of financial assets and agricultural goods (Ngangoue, 2021; Boyd and Bellemare, 2020). Our work is also related to a broader literature on learning models with delays in belief updating (Karlsson et al., 2009; Peng, 2005). However, in these models delays typically arise endogenously, either by consumer decisions to delay learning or information processing limits which slow the usefulness of information. In contrast, our model exploits exogenous variation in the delayed *arrival* of information that lies outside the consumer's control, but which still affects the marginal utility and costs associated with choices retroactively. That is, consumers must commit to expenditures weeks or months before knowing what those expenditures actually are. Finally, our model allows for household learning over time across a heterogeneous set of health services.

Second, our findings fit into a larger discussion of the usefulness of price transparency policies in mitigating large levels of healthcare consumption in the United States (Muir et al., 2012; Zhang et al., 2020). In contrast to previous work—which highlighted how the availability of price information may change the strategic decisions of patients shopping for a service (Gondi et al., 2021; Reed et al., 2005)—we highlight a new mechanism through which price transparency may affect *future* care decisions across entire households. Our findings provide strong evidence that reducing price uncertainty in the weeks or months after a service may have snowball effects reducing the utilization of unnecessary care for a greater number of people over a longer duration of time. Policies which shorten the length of a delay for medical cost estimates would reduce variance between expected and actual cost-sharing, as would real time claims adjudication, similar to prescription drug claims adjudication (Hartzema et al., 2011).

Finally, we provide the first policy-relevant estimate of household beliefs about health expenditures, which also allows us to estimate the effect of under-information on over-consumption of care. Our model also allows us to compare alternative plan designs that may limit over-consumption, including those currently proposed by medical regulators (Korenstein et al., 2012; Elshaug et al., 2017).[5] Our findings are therefore related to a broader discussion of how consumers respond to nonlinear health insurance contracts (Brot-Goldberg et al., 2017; Stockley, 2016) and imperfect belief formation in healthcare Hoagland (2022). Similar to these papers, we find that consumers are responsive to cost-sharing across a broad spectrum of services. However, we also find that these fluctuations could be smoothed by targeted health policies, such as plans with shorter time increments for deductibles (Shafer et al., 2022).

We discuss the setting of shoppable services and the data in Section 2. We then present our methods and identifying assumptions in Section 3, followed by our empirical results in Section 4. We incorporate these findings into a model of imperfect moral hazard in Section 5, with estimated results and insights in Section 6. Finally, Section 7 highlights the relevance of these findings for optimal design of insurance contracts.

2 Setting and Data

2.1 Data

Our primary data on household health utilization come from the IBM/Truven Marketscan *Commercial Claims and Encounters* Data, spanning from 2006 to 2018. These data contain detailed inpatient, outpatient, and pharmaceutical claims for a sample of households enrolled in ESI through large U.S. firms. Each observation includes diagnostic, procedural, and payment information, including the date of service and the corresponding date on which the insurer paid their portion of the claim. In addition, the data includes household, firm, and insurance plan identifiers.[6]

[5] Although the policy relevance of price uncertainty in healthcare for developing countries has been noted briefly (Knowles, 1995), our work is the first to formalize this in a model of learning with direct policy implications. Our model is useful both in the context of the United States and other countries where consumers face demand-side cost-sharing for using health care; this includes a broad range of countries, including those with universal health care systems that still incorporate some degree of cost-sharing (e.g., Australia, Germany, and the Netherlands, among others) (Globerman, 2016). For a more in-depth discussion of models of optimal contract design under moral hazard, see Winter (2013) and Blomqvist (1997).

[6] Note that insurance plan identifiers are only available through 2013, which will affect the analytical sample used in the structural exercise.

We limit our analytical sample to enrollees in one of eight large firms with plan identifiers available. Our final sample includes 386,240 households with two or more members, full eligibility, and continuous enrollment across their window of observation. Throughout, spending data has been normalized to 2022 USD using the Consumer Price Index for All Urban Consumers series.

	Full Sample	Plan-Identified Sample
Panel A: Demographics		
Age (individual)	31.67 (0.000)	31.15 (0.000)
% female (individual)	0.51 (0.000)	0.51 (0.000)
Risk Score	0.29 (0.000)	0.29 (0.000)
Family Size	3.08 (0.000)	3.10 (0.000)
Panel B: Medical Utilization		
Total medical spending (indvidiual)	$4,764 [$975] (0.002)	$4,406 [$887] (0.002)
% of individuals with no spending	0.17 (0.000)	0.20 (0.000)
OOP medical spending (individual)	$650 [$198] (0.000)	$562 [$167] (0.000)
Household deductible | deductible > 0	—	$1,040.24 (0.001)
% Households with zero deductible	—	0.26 (0.000)
Household coinsurance rate	—	0.29 (0.000)
% individuals with shoppable services	0.06 (0.000)	0.06 (0.000)
Total cost, shoppable service	$5,572 [$3,721] (0.011)	$5,645 [$3,814] (0.015)
OOP, shoppable service	$691 [$388] (0.002)	$574 [$290] (0.002)
Years	2006–2018	2006–2013
N_{families}	368,237	367,445
$N_{\text{individuals}}$	1,357,392	1,311,554

Notes: Enrollees include employees and their covered dependents. Risk scores are calculated using the CMS-HCC 2014 community model. Household plan characteristics are calculated as discussed in Section 2. Spending values are reported in 2022 USD. Standard errors are reported in parentheses; medians (when reported) are in brackets.

Table 1. Household Summary Statistics

Table 1 presents summary statistics for the full sample as well as the subset of the sample with insurance plan identifiers. Households tend to be young and relatively low-risk, with an average age of 31.7 years and between 3 and 4 household members. Insurance coverage is more generous than average, although the conditional average deductible is over $1,000, and household members who select into shoppable services typically spend close to a full year's OOP costs on that service alone. Note that the sub-sample with plan identifiers does not appear substantially different from the full sample, an important fact given that we use the plan-identified sample in our structural approach (Section 6). Households in the plan-identified sample incur slightly lower OOP costs than the full sample; however, this is likely indicative of decreasing insurance coverage generosity over time, given that the latest 5 years of data are excluded in this sub-sample.

2.2 CMS Shoppable Services

Our goal is to assess how pricing information contained in a medical bill alters household utilization patterns. In our primary specifications, we analyzed the impact of medical bills for individual health services that are expected to generate a significant—but unknown—amount of OOP spending for the household. We identified the utilization of 30 CMS "shoppable services," which correspond to non-urgent services that patients can schedule in advance (CMS, 2019). We identified these services in the claims data using Current Procedural Technology (CPT) codes for outpatient and inpatient services and Diagnostic Related Groups (DRGs) for inpatient hospitalizations; the complete list of services is available in Appendix Table A.1.[7]

2.3 Bill dates & Plan characteristics.

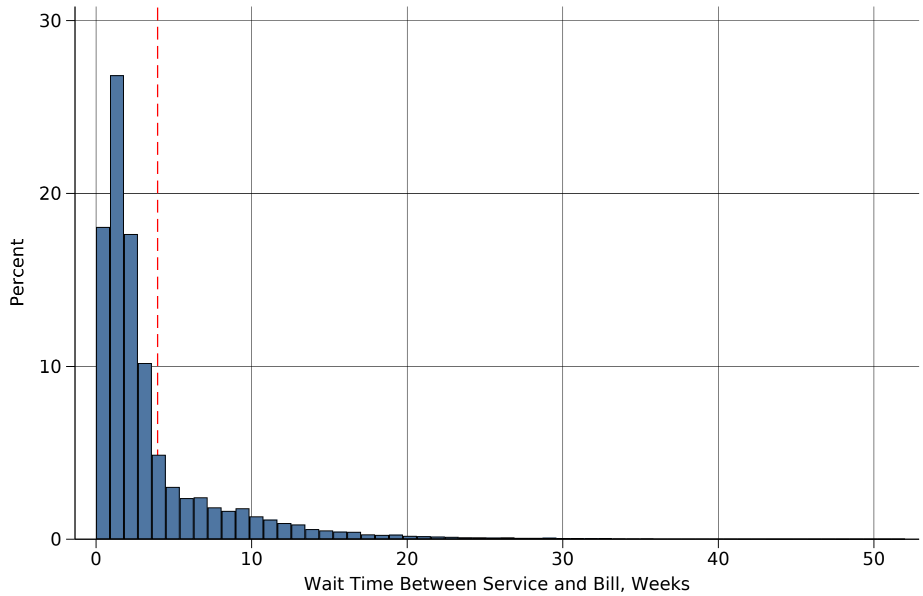
One limitation of our data is that we are not able to view the exact date on which consumers first received a bill from their provider for the services rendered. Instead, we observe the date the insurance plan paid the provider their portion of the pill. As this is the first possible date at which a patient will receive their Explanation of Benefits (EOB), it is the earliest date that definitive OOP cost information becomes available to a patient. Hence, we use this date as a proxy for patient bill information.

Figure 1 presents the distribution of wait times (in weeks) between the date a shoppable service was received and the date the insurance plan paid their portion of the service bill. Note that there is substantial variation in this wait time, with roughly 60% of bills being paid by insurers within the first four weeks, and the rest taking longer than a month for payment to be settled. As discussed in Section 1, the variability in the length of this period is exogenous to the household and allows us to identify the causal impact of receiving information about OOP expectations on household health utilization.

In addition to data on individual health services, we utilize data on insurance plan characteristics to estimate how households respond during the period when OOP prices remain

[7]Note that effective January 1, 2021, hospitals must publish standard charges for these services online, including negotiated rates. However, this does not affect our analytical sample which only goes through 2018; instead, we selected these as services which patients typically schedule or plan for ahead of time, meaning there is high potential for household strategic response to the service. Additionally, patients may engage in some form of price shopping for the procedure ahead of time, although there is little evidence of shopping in healthcare (Mehrotra et al., 2017).

Figure 1. Variation in Wait Times Between Service Date and Bills' Arrival



Notes: Figure depicts the distribution of wait times between the date the service was provided and the date the insurance company paid their portion of the service bill to the provider, measured in weeks. Only services included as shoppable health events in our analytical sample are shown here. Vertical dashed red line indicates the average duration of the waiting period, approximately 4.5 weeks.

unknown to them. In constructing measures for these characteristics, we follow previous literature (Hoagland, 2022; Marone and Sabety, 2022).[8]

3 Methods

Household consumption of shoppable health services—particularly those with large expected OOP costs relative to a deductible—may generate strategic responses in household health spending (Cabral, 2017). When exact OOP spending information is not immediately transmitted, responses take place in two stages: first, households respond to the event itself, based on their expectations of spending; second, and only after the bill arrives, households make decisions with full information in hand. Hence, we leverage these two distinct response periods in a triple-differences regression framework to estimate spillover responses to scheduled healthcare consumption *separately* for the periods before and after a bill's arrival.

[8] For tractability, our model allows for plans to have a piecewise-linear cost-sharing contract comprised of: a family deductible, a simplified non-specialist coinsurance rate, and a family OOP maximum. Measures are constructed based on the empirical distribution of payments in the claims data and cost-sharing parameters which minimize the sum of squared residuals between predicted and observed OOP household spending (Zhang et al., 2018; Marone and Sabety, 2022). See Hoagland (2022) Appendix A for a detailed description of this methodology and an evaluation of the quality of these inferences.

We estimate the causal impact of a bill on spillover spending (e.g., for all household members excluding the original consumer of the shoppable service). The estimating equation for total spillover health spending (measured per week per household-member) in household i at week t of year y is given by Equation 1:

$$\mathbb{E}[\text{spend}_{ity}] = \exp\left\{\beta_1 \mathbb{1}(\text{post\_service}_{ity}) + \beta_2 \mathbb{1}(\text{post\_bill}_{ity}) + \alpha_{\mathcal{I}} + \tau_t + \gamma_y + \xi_{\text{MD}}\right\}, \quad (1)$$

where the two main regressors are dummy variables indicating whether the shoppable service had already been performed and if the bill for the service had arrived by week t, respectively. In addition, we consider the robustness of our estimation approach to controlling for various time-invariant fixed effects, including those for individual households, years, relative week of the year (to account for within-year seasonality in health spending), and provider fixed effects (for the providers offering the shoppable service).

We use Poisson regression to estimate multiplicative effects on spending. A Poisson regression model is advantageous as it allows us to deal with the skewed nature of our (non-negative) spending data while appropriately including weeks with zero spending and avoiding a complete specification of the dependent variable's distribution (Manning and Mullahy, 2001). Our estimator will be consistent as long as the conditional mean of the dependent variable is correctly specified, as is the case in ordinary least squares (OLS) regression Gourieroux et al. (1984). Additionally, Poisson regression allows us to avoid the inconsistency of regression coefficients induced by heteroskedasticity in a log-linear transformed model (Santos Silva and Tenreyro, 2006).[9]

A critical assumption for the identification of our parameter of interest ($\beta_{\text{post\_bill}}$ in Equation 1) is that the arrival of spending information through the bill is exogenous to the affected households' strategic spending decisions. Previous work has highlighted the endogeneity inherent in estimating demand elasticities to major health events, especially when those events are planned or scheduled in advance (Duarte, 2012). In our estimation, as we are not attributing spending responses to changes in price (e.g., $\beta_{\text{post\_service}}$ is not a demand elasticity), there are no potential endogeneity concerns in estimation, as we are including both strategic and non-strategic responses collectively in $\beta_{\text{post\_service}}$. As long as bill arrival times are exogenous to the household, $\beta_{\text{post\_bill}}$ will represent a causal estimate of household "corrections" in response to pricing information. However, should bills take systematically longer for higher-risk patients (who might have families who are more likely to respond to health spending in the first place), our estimates would be inconsistent.

There is strong evidence that variation in the time households wait for a bill to arrive

[9] Poisson regressions were estimated in Stata using the "ppmlhdfe" command to handle high-dimensional fixed-effects (Correia et al., 2020).

	Average Spending		**Difference**			**Sample Size**	
Procedure	$d \leq 30$	$d > 30$	Unadjusted	Adjusted	p-value	$d \leq 30$	$d > 30$
Removal, prostate	$21,834	$25,362	$3,528	$1,260	0.41	917	403
Removal, knee cartilage	$7,619	$8,021	$402	$697	0.00	46,937	15,606
Removal, breast growth	$4,887	$5,173	$286	$674	0.00	10,550	3,916
Injection, anesthetic	$3,258	$3,537	$279	$484	0.00	49,604	16,667
Biopsy, esophagus/stomach	$3,317	$3,238	-$79	$406	0.00	245,411	65,603
Removal, tonsils (age < 12)	$4,578	$4,871	$293	$342	0.00	21,503	4,962
Shaving, shoulder bone	$12,262	$12,040	-$222	$233	0.07	27,952	11,410
Biopsy, prostate	$2,653	$2,377	-$276	$124	0.01	23,172	6397
Removal, gallbladder	$9,217	$9,794	$577	$96	0.38	36,756	13,252
Hernia repair	$6,753	$6,724	-$28	$28	0.83	14,314	5,215
Removal, cataract (no insertion)	$1,408	$1,198	-$210	-$179	0.05	11,776	2,388
Vaginal delivery	$7,789	$7,927	$139	-$344	0.00	82,968	36,068
Removal, cataract (lens insertion)	$6,114	$5,958	-$156	-$346	0.00	43,129	9,266
Vaginal delivery, prior C-section	$8,429	$8,634	$205	-$912	0.01	1298	503

Notes: Table shows differences in means for total spending (patient + insurer payments) by procedure category for the shoppable services included in our analytical sample. Services are divided into groups if (1) the bill took 30 days or fewer to arrive following the service ($d \leq 30$) or (2) the bill took more than 30 days to arrive ($d > 30$). Differences in means are presented both in raw, unadjusted terms, as well as adjusted for provider-specific fixed-effects. p-values are the results of two-way difference in means testing on the adjusted differences.

Table 2. Bill Balance Table (Unadjusted and Adjusted for Provider Fixed Effects)

is exogenous to the household and uncorrelated with underlying patient risk or procedure severity, as discussed in Section 1. Bill wait times are highly seasonal within a year for a given payer, depending on both the total volume of claims they are processing (Appendix Figure A.8) and administrative frictions of initiating new enrollees and groups to new benefits. In addition, factors such as new risk-adjustment policies, transitions in billing systems (such as the 2014-2015 shift to ICD-10-CM), or even the COVID-19 pandemic can overwhelm payer processing of claims, occasionally even drastically increasing the wait time before patients learn with certainty their ultimate OOP costs (Snowbeck, 2022).

Still, bill wait times may be associated with underlying patient risk, potentially introducing selection concerns affecting causal inference. If payers have an incentive to slow down payments for highly-expensive procedures, or if a higher-complexity patient takes longer to process, bill times could be systematically longer for the most at-risk patients in our sample. We test these claims directly in our sample, by comparing differences in the average total cost of shoppable services based on the length of the bill wait time in days.

Table 2 presents the results. For each potential service, the hypothesis that bills that took longer to arrive ($d \geq 30$ days) are associated with more (or less) expensive procedures is tested. We test these hypotheses for both unadjusted means and averages adjusted for provider-specific trends. We find that the large volume of procedures in each group lends itself to statistically significant differences, but not economically meaningful ones. The average difference across services constitutes only \$220, 6.1% of total payments. In addition, the estimated value of the differences varies widely, with almost a quarter of the included procedures estimated to have *shorter* wait times for more expensive instances of the procedure. Taken together, we find little evidence that bill wait times may be endogenous at the household level.

4 Empirical Results

4.1 Effect of Bills on Spending

Table 3 presents the regression results from estimating Equation 1. We find robust evidence that although spillover household spending increases following the utilization of a shoppable service, the bills arrival causally affects these responses Without conditioning on the bill's arrival, the overall estimated spending increase is roughly 40.2% of average weekly per-person household spending (roughly \$120). However, this conflates a period prior to the plan's payment of the claim where spending is estimated to increase by 46.4%, only to decline by 7.7% once the bill arrives. This decline (roughly 17% of the overall change in spending) is consistently estimated even after controlling for time-invariant household and year trends as well as for within-year seasonality and provider fixed-effects. The correction amounts to \$9.28 per person per week for the average household in the sample, or roughly \$482.44 per person annually.

	Main Models		Alternative Specifications		
Post Service	0.402***	0.464***	0.597***	0.472***	0.486***
	(0.0022)	(0.0032)	(0.0032)	(0.0032)	(0.0033)
Post Bill		-0.077***	-0.080***	-0.096***	-0.076***
		(0.0030)	(0.0030)	(0.0030)	(0.0031)
$\overline{\text{spend}}_{it}$	\$120.49	\$120.49	\$120.49	\$120.49	\$120.49
Household FEs	X	X	X	X	X
Year FEs	X	X		X	X
Week of Year FEs	X	X			X
Provider FEs	X	X			
Observations	61,860,735	61,860,735	61,860,735	61,860,735	61,860,735

Notes: Table presents results from triple-difference Poisson regressions highlighting the role of a bill's arrival on health spending of affected household members. Each column in this table estimates the impact of a single household member's shoppable health service—and accompanying bill—on health spending for all other household members. Regression coefficients displayed illustrate the expected change in log household spending (measured per person-week) associated with the service date and bill arrival (both measured as dummy variables). Throughout, standard errors were clustered at the household level.
$^{*}p < 0.05, ^{**}p < 0.01, ^{***}p < 0.001$

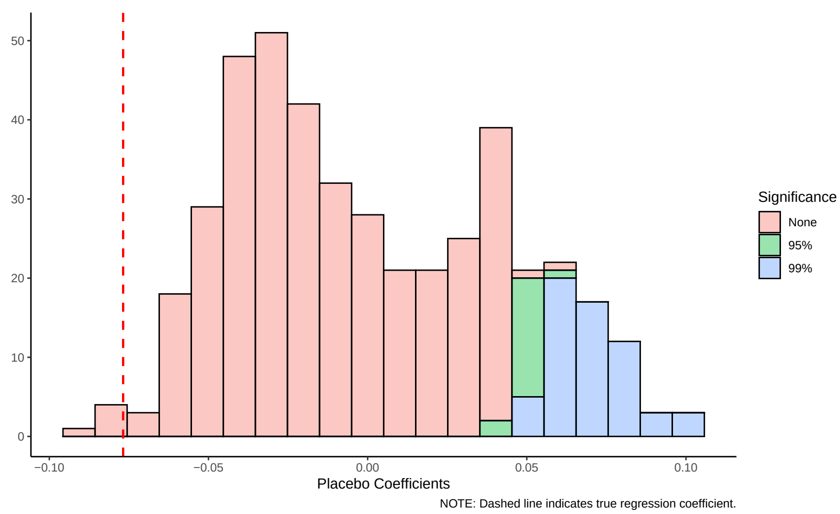
Table 3. Estimated Impact of Bill Arrival on Household Health Spending

These findings are consistent with a model in which consumers over-estimate their actual OOP contributions prior to receiving the definitive pricing information of a medical bill. Due to this over-estimation, consumers enrolled in insurance plans with piecewise-linear cost-sharing insurance plan designs (e.g., a nonzero deductible) may incorrectly assume that the marginal cost associated with additional services has declined (e.g., by meeting a deductible). Once the bill arrives correcting any errors in perception, however, individuals curtail their spending increases in response.

Given that we are using the dates insurance plans pay providers as a proxy for bill arrival, it is possible that we are artificially splitting the post-service period into two essentially random periods and attaching significance to a spurious difference between the two periods. To test this possibility, we conducted placebo tests, running regressions on artificial data that randomly assigned consumers new wait times for their bill based on the empirical distribution of wait times.[10]

Figure 2 presents the results for 250 placebo simulations. The distribution of all estimated regression coefficients $\beta_{\text{Post\_Bill}}$ is shown, compared against the estimated regression coefficient of our preferred specification (-0.077) depicted as the vertical dashed red line. Placebo coefficients are color-coded by their associated p-value. The distribution of placebo coefficients is centered around zero, with most of the coefficients statistically indistinguishable from zero. In fact, the only statistically significant coefficients in the placebo regressions are those that would indicate a *positive* correction in spending following the arrival of the bill. Taken with the results from Table 3, this suggests that it is unlikely that our results are spurious correlations from a convenient semi-random splitting of the post-service period.

Figure 2. Distribution of Placebo Regression Coefficients for $\beta_{\text{post\_bill}}$



Notes: Figure shows the distribution of placebo regression coefficients for the dummy variable Post_Bill_{it} in Equation 1 ($n = 250$). Each placebo is constructed by artificially varying wait times for bills based on the empirical distribution of wait times in the analytical sample. Standard errors are clustered at the household level. Coefficients are color-coded based on statistical significance. The vertical dashed red line indicates the estimated coefficient of the preferred specification (Table 3).

4.2 Understanding Mechanisms: Heterogeneity by Deductible Spending

There are several ways households might find bills informative enough to alter their health-care utilization patterns. First, households may learn about the overall prices of the services

[10] That is, for each individual consuming a shoppable health service, we kept fixed the service date and artificially constructed a new bill arrival date as the service date + a random draw of a bill wait time. Wait times were drawn from the empirical distribution of our sample (e.g., Figure 1). All regressions were then estimated on the new sample with an updated Post_Bill_{it} coefficients in each iteration.

they received, particularly their own OOP burden for their care. This may include more detailed information about the percentage of a household deductible that has now been met as a result of the service. In particular, a bill that informs consumers that they are still short of meeting a deductible may generate the corrective action observed in Section 4.1. Second, households may learn about the extent to which their insurance does or does not cover certain procedures. In this sense, bills inform households not about the overall prices of services, but correct a misunderstanding of the fraction of services they will have to cover OOP. Finally, a bill may reveal discrepancies between a patient's understanding of a service and the provider's billing, including up-coding practices. This information may also alter future healthcare spending to the extent that it erodes household trust in the healthcare system (Webb Hooper et al., 2019).

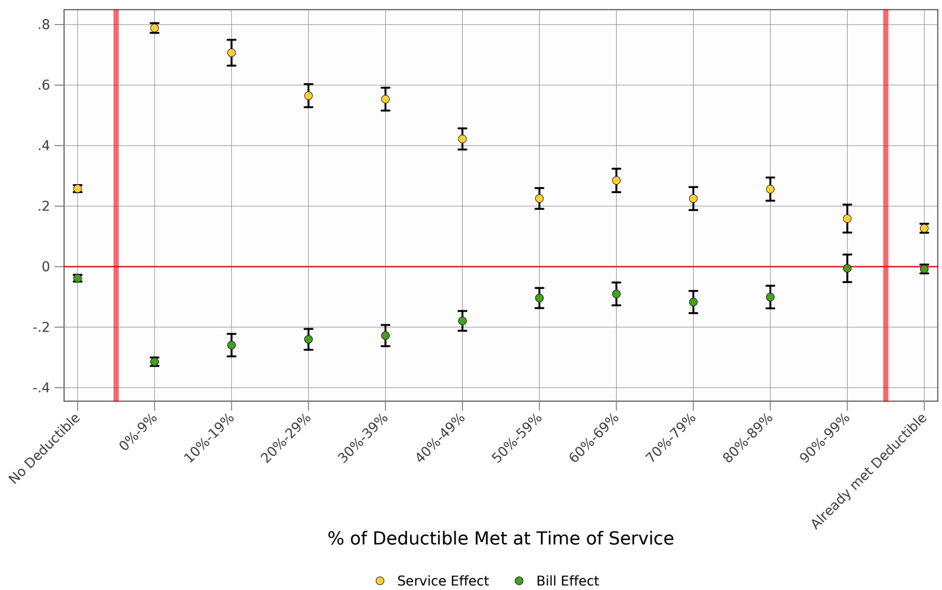
In order to understand the mechanisms driving household responses to bills, we assessed responses across the distribution of spending patterns prior to the shoppable service. The intuition for this exercise is that variation in pre-event spending (specifically, the fraction of a deductible met prior to a service) is informative of the amount of engagement with the health system in a particular year. Hence, the extent to which responses to bills vary across the distribution of pre-event spending indicates the rate at which households are learning about their own prices of care, rather than learning about other features of the health care system.

Figure 3 presents results stratified by decile of household deductible spending prior to the event. The spending responses for both the interim period between the service and the bill (yellow) and the post-bill correction (green) are shown in panel (a). Both spending responses and corrections are largest for households who have spent little towards their deductible before the event: post-service spending increases are estimated to be as high as 80% for households with less than 10% of their deductible met, and fall to under 20% for households close to their deductible. Correspondingly, post-bill corrections are estimated to be as high as 30% (roughly 37.5% of the post-spending increase) for the low-spending group, and converging to zero for the high-spending group. In both cases, spending responses for the group closest to meeting their deductible are statistically indistinguishable from households who met their deductible prior to the shoppable service. Finally, households who don't face changes to their marginal cost of care from the bill (e.g., those in zero-deductible households and those which have already met their deductible) exhibit near-zero spending responses to the bill.[11]

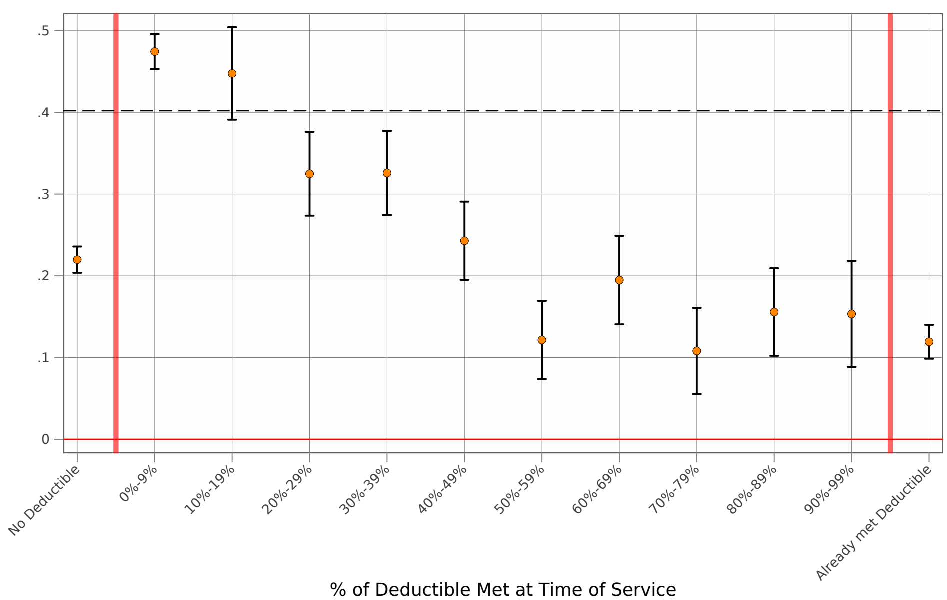
[11] The coefficient for bill response among the group with zero deductible is statistically different from zero, but with a coefficient smaller than 5%. This is likely due to some measurement error in the classification of individuals as being enrolled in a plan without a deductible, which may still include some cost-sharing for specific services (e.g., emergency department use) (Zhang et al., 2018).

Figure 3. Heterogeneous Bill Effects Across Household Deductible Status at Time of Service

(a) Estimated Regression Coefficients



(b) Total Post-Bill Change (Sum of Coefficients)



Notes: Panel (a) shows estimated coefficients and 95% confidence intervals for $\mathbb{1}\{\text{Post\_Service}_{it}\}$ and $\mathbb{1}\{\text{Post\_Bill}_{it}\}$ in Equation 1 by decile of household deductible spending prior to the event. Standard errors are clustered at the household level. Deductibles are imputed based on previous literature (Hoagland, 2022; Zhang et al., 2018). Panel (b) depicts the estimated post-bill spending increase (e.g., the sum of the coefficients). The horizontal dashed black line indicates the estimated pooled spending increase in column 1 of Table 3.

Panel (b) shows the estimated post-bill spending increase, measured as the sum of the two estimated regression coefficients. Spending increases following a bill's arrival tend to be significantly larger for previously lower-spending households, including above the average 40.2% spending increase estimated in Table 3. This may be because households with previously low levels of spending coordinate their care more, exhibiting a greater tendency for strategic decision-making (alternatively, this could be due simply to seasonality in spending patterns). As the level of pre-event deductible spending increases, however, post-bill responses fall to between 10% and 20%, statistically indistinguishable from the group of households which have already met their deductible prior to an event.

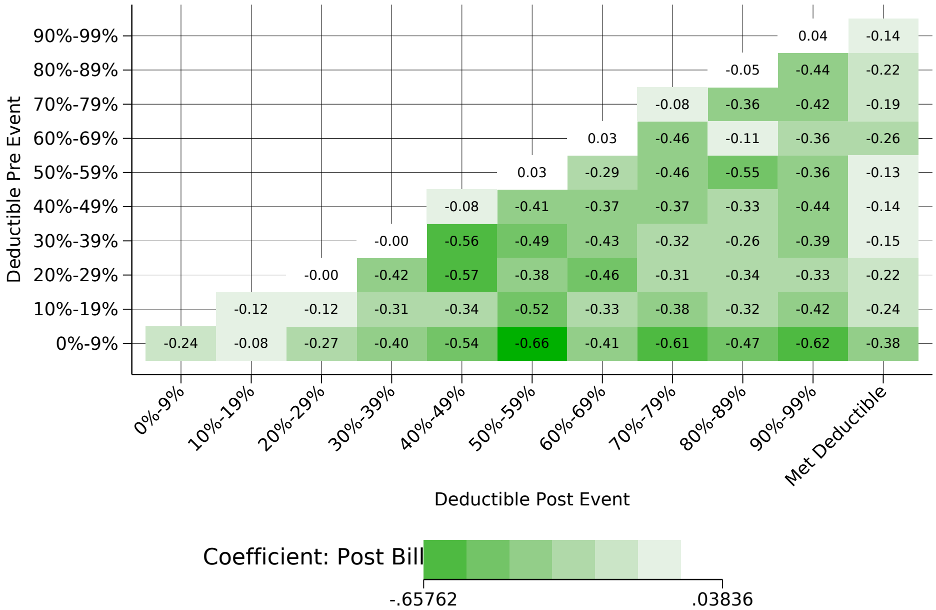
Taken together, these results suggest that households respond, at least in part, to a bill's information about OOP expenditures. This information appears to be especially relevant to households who have yet to contribute much to their deductible, while households without a deductible or who have already met it exhibit negligible responses to the bill.

Additionally, we can gain further insight as to the mechanisms associated with bill responses by using additional variation in the relative cost of a shoppable service. The intuition for this exercise is that high-cost events are more likely to ultimately alter a household's marginal cost of services, while lower-cost events may have similar price uncertainty without any economically meaningful costs to that uncertainty. Given that price information is most valuable to households when it communicates whether or not households have crossed the threshold of their deductible (e.g., marginal costs are reduced), we can identify the extent to which households learn about price from their bills—separate from other forms of learning—by comparing household responses to high- and low-cost events. To do this, we explore two-way heterogeneity in the effects of household responses considering both pre-event deductible contributions *and* the resulting change in deductible spending after the scheduled health consumption.

Figure 4 presents the results. We restrict our attention to households enrolled in plans with a non-zero, unmet deductible at the time of service. We then separately estimate Equation 1 across cells of households who have similar deductible spending both before and because of the major health event. The figure depicts a two-way heatmap of estimated bill responses across cells. Consistent with Figure 3, we find that households starting at lower levels of their deductible exhibit greater sensitivity to their bill. In addition, we find that households appear considerably less responsive to low-cost services; coefficients are estimated to be much closer to zero when households do not move across deciles of spending.

Most notably, however, Figure 4 highlights a stark difference in household responsiveness to events that left households just shy of missing their deductible, relative to events that pushed households across the threshold. Across all levels of pre-event spending, esti-

Figure 4. Heterogeneous Bill Effects By Household Deductibles and Service Cost



Notes: Figure depicts estimated coefficients for $\mathbb{1}\{\text{Post\_Bill}_{it}\}$ in Equation 1 across deciles of household deductible spending prior to *and* following an event. Here, sample is restricted to individuals in a non-zero deductible plan who have not yet met their deductible at the time of service. Each row indicates a different decile of deductible spending prior to the event, while each column indicates deciles following the event. Standard errors are clustered at the household level. Deductibles are imputed based on previous literature (Hoagland, 2022; Zhang et al., 2018).

mated coefficients are largest when a bill arrives for an event that just missed the household deductible. On the other hand, coefficients are smaller and more likely to be statistically indistinguishable from zero if the bill reveals that the household has already met their deductible. This is consistent with a model where households over-estimate OOP spending—or in other words, the probability of having met a deductible—thus exhibiting greater spending corrections closer to the threshold at which marginal costs would change.

Taken together, accounting for heterogeneity in household responsiveness to bills—including both spending histories and the relative cost of services—suggests that households are most responsive to bills when the pricing information they reveal are particularly salient. When households have little information about their deductible, or when the bill indicates that a household's marginal cost of additional care has not changed, the arrival of pricing information curbs overly large responses in spillover spending following a shoppable health service.

4.3 What Services Are Affected?

Finally, we assess whether household responsiveness to price information varies across broad categories of medical services, including hospital care, specific types of outpatient services, and pharmaceutical spending. This decomposition allows us to examine whether household spending responses—or the extent to responses are corrected after a bill arrives—varies with any measure of perceived or real quality of care. Particularly, we examine how bills affect future utilization of typically high-value health services (e.g., preventive screenings, behavioral health services) as well as typically low-value care (e.g., unnecessary pre-operative screenings, imaging services, or surgeries).

Table 4 presents the results. We separately estimated the coefficients for Equation 1 with each sub-category of spending as its own dependent variable.[12] Overall, we find that households respond to shoppable services across a spectrum of services, increasing their consumption in the short-run after a service is performed, and then reducing that consumption significantly once the bill arrives.

After an individual household member receives a significant health services, other household members are more likely to seek hospital care, including a 10.8% increase in emergency department visits and a 37.1% increase in visits for potentially preventable hospitalizations (e.g., admissions to treat dehydration) (Agency for Healthcare Research and Quality, 2007). Following the receipt of the bill, however, use of inpatient care for preventable hospitalizations falls by roughly 52.3% of the increase (a 19.4% decrease).

Some outpatient services, including those for behavioral health (e.g., psychotherapy) and chiropractic care (e.g., physical therapy), are affected neither by the consumption of a shoppable service nor its accompanying bill; this is presumably because these services have more inelastic demand and lower rates of cost-sharing generally. However, we find that when households increase demand for a type of outpatient service following a major health event, they tend to over-increase spending. Bill arrivals cause households to correct spending increases by between 32% and 90%. Households increase their utilization of general practice visits (e.g., E&M visits, lab work, and preventive screenings) the most, followed by specialist visits (e.g., dermatology). While household demand for prescription drugs increases slightly following a health event in the home (by 1.8%), we do not observe a corresponding reduction in demand following the bill's arrival. This could be because of the already high levels of pharmaceutical spending relative to other medical consumption.

Somewhat surprisingly, we do not observe that households reduce their utilization of low-value care following a bill's arrival. These services, which include services such as imaging

[12] Appendix Table XXX includes detailed descriptions of the construction of each of these variables.

	Regression Coefficients		Pre-Treatment Averages	
	Post Service	Post Bill	% ≥ 0	Conditional Mean
Hospital Care				
Emergency Department	0.108***	-0.021	0.67%	$929.98
	(0.0129)	(0.0133)		
Preventable Hospitalizations	0.371***	-0.194*	0.04%	$19,979.89
	(0.0829)	(0.0848)		
Outpatient Care				
Behavioral Health	-0.020	0.018	1.19%	$119.47
	(0.0132)	(0.0134)		
Chiropractic Care	-0.005	0.027	1.86%	$133.39
	(0.0147)	(0.0151)		
Evaluation & Management	1.440***	-0.518***	1.05%	$121.45
	(0.0066)	(0.0062)		
Imaging	0.098***	-0.037***	2.55%	$265.52
	(0.0108)	(0.0111)		
Lab Services	0.198***	-0.178***	3.96%	$62.14
	(0.0113)	(0.0119)		
Low-Value Services	0.084***	0.028**	6.58%	$148.61
	(0.0094)	(0.0097)		
Preventive Care	0.345***	-0.249***	11.68%	$120.47
	(0.0036)	(0.0037)		
Specialist Care	0.550***	-0.181***	0.57%	$114.70
	(0.0192)	(0.0198)		
Prescriptions	0.018***	-0.006	18.30%	$147.14
	(0.0047)	(0.0048)		

Notes: Table shows coefficients from triple-difference regressions capturing service-specific effects of pricing information ($N = 59,177,995$). Columns (1) and (2) present regression coefficients; column (3) indicates the fraction of pre-treatment weeks when spending was positive; and column (4) presents pre-treatment weekly averages, conditional on positive spending. See Appendix Table XXX for a complete list of the CPT codes for each of the outpatient categories. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$

Table 4. Estimated Impact of Bill Arrival on Service-Specific Spending

for lower-back pain, misuse of prescription medications to manage migraines and bacterial infections, or unnecessary pre-operative screenings, are determined based on the recommendations of the Choosing Wisely campaign (Colla et al., 2015). We find that households increase their use of low-value care by 8.4% following a major service, and then further by another 2.8% once the bill arrives. This may be a result of a "cascade of care" effect associated with increased consumption of general medical care, which in turn prompts downstream increases in physician ordering of low-value services (Ganguli et al., 2020). Physicians typically retain control over when low-value services are performed, in order to reduce their own uncertainty, liability, or "just to be safe" (Colla and Mainor, 2017). Our finding of increased consumption of low-value services following major health events is consistent with prior work, but warrants future exploration as to whether it is physician- or patient-driven (Hoagland, 2022).

5 Model

Based on the empirical findings from our reduced-form analysis, we propose a model of imperfect moral hazard, in which consumers make medical care choices based on beliefs about realized spending. Central to the model is the delayed nature of pricing information, which may lag behind consumption by weeks or months while still affecting the spot prices of care in ways that are unknown to the consumer before the bill arrives. As a result, consumers must form expectations about realized OOP spending and the implied marginal cost of care in each period. We first consider a case where consumer beliefs are static (e.g., where there is no learning) before introducing a learning component to the model in Section 5.2.

A patient i chooses medical spending in each period t, with the choice variable m_{it} measured in the dollar value of the services.[13] The patient selects m_{it} as a utility-maximizing response to a health shock λ_{it}. Following Einav et al. (2013), we calibrate individual patient utility as a quadratic loss function in the distance between selected health spending and the unobserved health shock:

$$u_{it} = (m_{it} - \lambda_{it}) - \frac{1}{2\omega_i}(m_{it} - \lambda_{it})^2 - c_{ijt}(m_{it}; M_{\mathcal{I}t}). \tag{2}$$

Here, ω_i is an individual time-invariant "moral hazard" parameter capturing individual heterogeneity in demand responsiveness to the price of services.[14] In addition, $c_{ijt}(m_{it}; M_{\mathcal{I}t})$ denotes the OOP costs associated with m_{it}, which depends on the piecewise-linear cost-

[13]To be consistent with the reduced-form evidence presented above, we model spending choices at the weekly level and measure spending normalized by household size.

[14]For more discussion on this heterogeneity, see Einav et al. (2013).

sharing contract of individual i's chosen insurance plan, j, as well as the OOP spending to date at the household level, $M_{\mathcal{I}t} = \sum_{i \in \mathcal{I}} \sum_{s=1}^{t-1} m_{is}$. Note that an individual's OOP costs for services are weakly decreasing in $M_{\mathcal{I},t}$.

Under full information, patients know both the value of $M_{\mathcal{I}_t}$ and how it affects $c_{ij}(\cdot)$. Furthermore, in the case where cost-sharing is linear at all stages of the contract, a patient's marginal OOP cost is given by $c_{ijt} \in [0, 1]$, where $c = 1$ applies to all services before a deductible has been met and $c = 0$ applies for all services after an OOP-max has been met. Between the deductible and the OOP-max, c is typically in the open interval $(0, 1)$. With full information about prices, the static choice of m_{it} in each period is simply the solution to the first order condition:

$$1 - \frac{1}{\omega_i}(m_{it} - \lambda_{it}) - c_{ijt} = 0 \Rightarrow m_{it}^* = \max\left[0, \lambda_{it} + \omega_i(1 - c_{ijt})\right]. \tag{3}$$

That is, medical expenses in each period are chosen so that the marginal utility of those services is equal to the marginal (known) OOP cost. In particular, as c changes from 1 to $c < 1$ as households meet their deductible, household members will have a discontinuous increase in their medical consumption in future periods.

However, based on the discussion in Section 4, we suppose that $M_{\mathcal{I},t}$ is not known with certainty at the time a service is performed. Rather, household spending can be divided into two components: spending for services whose bills have already arrived (e.g., where prices are known), and spending for services without pricing information yet available. For ease of notation, suppose that each bill takes τ weeks to arrive, so that a bill for a service procured in week t would arrive in week $t + \tau$.[15] Based on these components, households respond to a signal of their spending θ:

$$\theta_{it} = \underbrace{\sum_{s=0}^{t-\tau} \sum_{i \in \mathcal{I}} c_{ij}(m_{is})}_{\text{known spending}} + \underbrace{\sum_{s=t-\tau+1}^{t} \sum_{i \in \mathcal{I}} s_i(m_{is}|x_{is})}_{\tilde{\theta}_{it} = \text{unknown spending}}, \tag{4}$$

where $s_i(m_{is}|x_{is})$ represents service-specific signals of spending, which may depend on individual, household, and service level characteristics.

[15]Note that in the empirical estimation of the model, the length of time between a service and bill's arrival is allowed to vary across services; this assumption is only made in this section for ease of exposition.

5.1 Simple Case: Constant Under-information

In the simplest case, we suppose that signals do not vary across services, but rather assume that cost signals are a constant multiple of true costs:

$$s_i(m_{is}|x_{is}) = \beta \cdot c_{ijs}(m_{is}). \tag{5}$$

That is, before a bill arrives, patients inflate (or deflate) their true OOP spending by a constant parameter β, which does not vary across services or individuals.[16] In the simplest version of the model, we also assume that there is no learning about β over time; we introduce learning in Section 5.2. Based on these assumptions, a household's signal of their OOP spending (and hence, of their marginal cost for additional spending) can be simplified as

$$\theta_{it} = \sum_{i\in\mathcal{I}} \sum_{s=0}^{t} (1 - D_{is})\beta c_{ijs}(m_{is}) + D_{is}c_{ijs}(m_{is}), \tag{6}$$

where D_{is} is a binary variable indicating if the bill for services performed in week s has arrived ($D_{is} = 1$) or not ($D_{is} = 0$). Based on the household's value of θ_{it} in each period, the signal for the marginal cost of future expenditures in the simple piecewise-linear insurance contract setting is given by

$$\hat{c}_{it} = \begin{cases} 1 & \theta_{it} < d \\ c & \theta_{it} \geq d, \end{cases} \tag{7}$$

where $c < 1$ in general.[17]

In this simplified case, the central parameter of interest is β, the rate at which households systematically over- or under-inflate their true levels of OOP spending prior to the arrival of the pricing information contained in a bill. Additional unobservable parameters in the model, which threaten identification, include heterogeneity in moral hazard ω_i and individual health shocks λ_{it}. Separate identification of β relies on being able to credibly identify the hyper-parameters governing the distributions of these characteristics.

When estimating the model, we calibrate these nuisance hyper-parameters to match moments predicted by (a) previous research and (b) training data not used in the structural estimation. We use the estimated regression coefficients predicted by Einav et al. (2013)

[16] Note that allowing β to be a random coefficient varying across individuals is a simple extension of the model; for the present purposes, however, we focus on an average of β across the population of interest.

[17] Note that in practice, we estimate the model on the sample of individuals enrolled in plans with non-zero deductibles. This is to cleanly capture the ways in which misperception of OOP spending may affect discontinuous changes in the marginal cost of spending across thresholds of the linear insurance contract.

in order to capture variation in moral hazard parameters across households.[18] We model individual-level health shocks as draws from an individual-specific shifted lognormal distribution; this distribution captures both the skewed nature of the observed spending data and the nonzero probability of an individual choosing zero spending in a period. That is, each individual in each period draws λ_{it} from a distribution $F(\mu_i, \sigma_i, \kappa_i)$ such that

$$\log(\lambda_{it} - \kappa_i) \sim \mathcal{N}(\mu_i, \sigma_i^2). \tag{8}$$

We calibrate the three hyper-parameters $(\mu_i, \sigma_i, \kappa_i)$ to match the empirical distribution of weekly spending using the individuals in our analytical sample who are *not* included in the structural estimation. These include individuals enrolled in plans with no deductible, as well as patients enrolled in any type of plan between 2014 to 2018. Individuals in this sample are grouped into cells based on patient demographics—including age, sex, risk score quartile, and relationship to the main employee—and the empirical distribution in each cell is matched to the shifted lognormal moments.[19] Once these parameters are identified, individual-period shocks are drawn for each member of the model sample and then summed to the household-period level.[20]

Given these calibrations, identification of the main parameter of interest β comes centrally from exogenous variation (at the household level) in the length of time required for a bill to arrive after different health services. This variation may exist across services as well as across households; importantly, underlying variation in $\tilde{\theta}_{it}$ which artificially moves households above or below their deductible is central to identifying how β changes household estimates of $\hat{c}$ in ways that most closely fit the observed choice data.

5.2 Learning

Once beliefs about OOP costs can be reasonably calibrated in the model, a natural question is whether consumers correct their beliefs with repeated exposure to health information. Households with frequent interactions with the health system, particularly within a plan year, may have beliefs about their bills which converge to the truth over time.

To assess this question, we incorporate household learning about the calibration parameter β. We model each bill's arrival as a signal from which consumers can learn.[21] Households

[18]Note that these regression models result in individual-level predictions for ω_i; in estimation, we aggregate these to the household level by taking the mean of $\log(\omega_i)$ across all members $i \in \mathcal{I}$.

[19]This is done using three properties of a shifted lognormal distribution: $\overline{\lambda} = \exp(\mu + \frac{1}{2}\sigma^2) + \kappa$, $\lambda^M = \exp(\mu) + \kappa$, and $\frac{\text{sd}(\lambda)}{\overline{\lambda}} = \sqrt{\exp(\sigma^2) - 1}$, where λ^M denotes the median. The solution to this system of equations given the moments of the empirical distribution of λ identifies the three hyperparameters μ, σ, κ.

[20]In order for shocks to be meaningful, we restrict $\lambda_{\mathcal{I}t} < m_{\mathcal{I}t}$ in each period.

[21]For now, we model each signal as having equal impact; future extensions of this model could flexibly

are assumed to have prior beliefs about β which follow a normal distribution with a mean $\mu_{\beta,0}$ and variance $\sigma^2_{\beta,0}$:

$$\hat{\beta}_{i0} \sim \mathcal{N}(\mu_{\beta,0}, \sigma^2_{\beta,0}). \tag{9}$$

When a bill arrives for a household conveying information about the prices of medical services, it in essence communicates that $\beta = 1$. Hence, we model each signal s_{it} as being drawn from a normal distribution centered at 1 and with a signal variance σ^2_s:

$$s_{it} \sim \mathcal{N}(1, \sigma^2_s). \tag{10}$$

We assume that households update their prior beliefs in accordance with Bayes' Rule, conditional on their observed signal. Assuming normal distributions for both the prior and posterior distributions allows for closed-form solutions for household beliefs about β at each period, and is consistent with previous learning models (Crawford and Shum, 2005). This expanded version of the model therefore has three parameters of interest. First, the average prior mean $\mu_{\beta,0}$ captures the extent to which households are uninformed about the relative costs of their medical care at the start of an enrollment period. The dispersion of this lack of information across households is captured in the variance of the average prior, $\sigma^2_{\beta,0}$. Finally, the variance of the signal, σ^2_s, reflects how precise the information communicated by each medical bill is, and subsequently how rapidly household beliefs converge to the true parameter of 1.

Estimating household learning allows deeper insight into the spread of household beliefs about their expenses both across households in the sample and over the relative course of an enrollment period. In particular, the speed with which beliefs converge informs the rate of over-consumption of medical care relative to fully-informed households. As in the simpler case of the model, identification of the three learning parameters $(\mu_{\beta,0}, \sigma^2_{\beta,0}, \sigma^2_s)$ stems from exogenous variation in bill timing. When the parameter space is expanded, identification comes from various sources. Within-household variation in expenditures relative to pending (hence, unknown) OOP expenditures serves to identify both the starting point of household beliefs (the prior mean) and the rate of convergence (governed, in this case, by the signal variance). Similarly, variation in choices across households identifies the spread of beliefs, summarized in the prior variance; this parameter informs both the spread of households' starting beliefs as well as how that spread evolves over time.

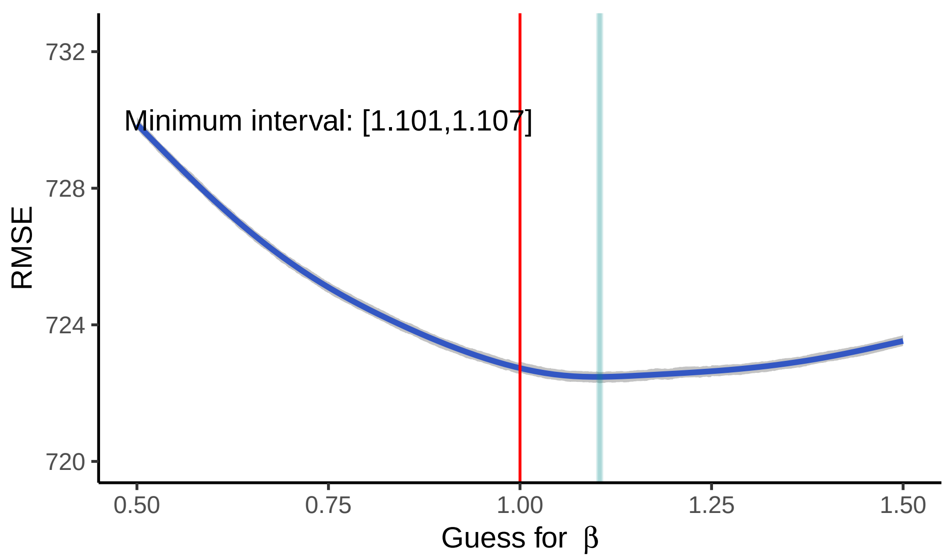
model heterogeneous signals based on the total cost of a service or by different service types.

6 Model Results

6.1 The Case of Constant Over-/Under-Estimation

We estimate the model presented in Section 5 for 240,111 households in our analytical sample enrolled in plans with nonzero deductibles from 2006 to 2013. For each household-week, we simulate household health shocks and draw idiosyncratic moral hazard parameters; then, for different values of β, we estimate household signals of underlying OOP spending and the marginal cost of incremental spending, $\hat{c}_{it}$. Taken together, these estimates produce a prediction of spending $m_{it}(\beta)$, which differs as β changes. Our primary measure of model fit is the root mean squared error (RMSE) between observed and predicted levels of weekly spending at the household level.

Figure 5. Estimating Household Responsiveness β to Spending Before Bills' Arrival



Notes: Figure depicts the relationship between chosen level of household pre-bill discounting parameter β and the mean squared error (MSE) of the model presented in Section 5. MSE is measured as the mean squared error between observed and predicted household spending at the weekly level. For each value of β, the median result of 50 simulations with independently drawn health shocks is shown in the black line; the confidence band illustrates one standard deviation above and below the median. The blue band denotes the full range of observed $\min_\beta MSE(\beta)$.

Figure 5 presents the estimated relationship between β and model fit, based on 50 simulations with different individual health shocks. The median RMSE for each value of β, as well as a confidence band of two standard deviations, are shown. Increasing the guess of β reduces the RMSE until the function reaches a minimum at $\beta = 1.104$ (in the median simulation), after which RMSE increases. The blue band in the figure shows the estimated 95% confidence region for the minimizing value of β, $[1.101, 1.107]$.

The model estimates, as illustrated, suggest that households over-estimate the OOP spending of services prior to the arrival of price information by between 10.1% and 10.7%. This is consistent with the findings of Section 4, which similarly illustrated a "correction" in implicit marginal costs following the arrival of the bill.

We conduct a simple counterfactual analysis to compare how spending predictions differ given this inflation, against a counterfactual world where β is correctly perceived to be 1 for all household-weeks (see Appendix Figure A.9 for details on the simulation). We find that over-estimating the costs of medical services leads to over-spending for 10.5% of households, with the average (median) affected household spending $842.80 ($480.59) more per household member in medical services that they would not have selected had they been correctly informed of their true OOP costs. This corresponds to an over-spending of 35.2% (33.3%) for the average (median) affected household (see Section 6.3 below for more details).[22]

6.2 Learning

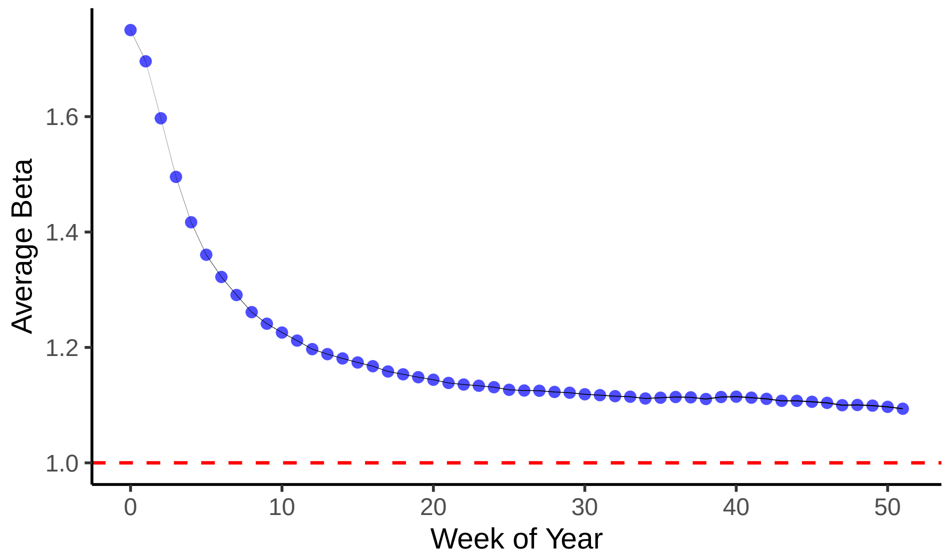
Finally, we incorporate the possibility of household learning into our estimates. We estimate that the median household's prior for β is roughly 1.75, indicating an 75% over-estimate of OOP costs (95% bootstrapped confidence interval: [1.702,1.798]). There is relatively little variation across households, as captured by the estimated prior variance parameter $\sigma^2_{\beta,0} = 0.011$ (95% CI: [0.002, 0.020]). Put into context, we estimate that roughly 95.5% of households (two standard deviations to either side of the mean) have prior beliefs that fall in the interval $(1.54, 1.96)$, indicating high levels of misinformation.[23]

Figure 6 shows how beliefs evolve in response to medical spending. We estimate that bills provide extremely precise information, with the signal variance term estimated to be 0.0002 (95% CI: [-8e-5,.0005]). That is, roughly 95.5% of household signals for β fall in the interval $(0.97, 1.03)$. This leads to rapid convergence of beliefs as the year progresses, as illustrated in the figure in the blue curve, which indicates the average value of $\hat{\beta}$ across the sample by week of year. Within the first quarter of the year, average household inflation for OOP costs has converged to below 20%. Following this rapid convergence, however, belief convergence stalls—it isn't until week 33 that the average household's value of β crosses the upper bound of the 95% confidence interval for β estimated in the non-learning model (1.107), and average beliefs don't dip below 110% until week 45. The average household does not have sufficiently many medical encounters for their beliefs to converge completely;

[22] Note that these percentage changes are measured relative to the counterfactual predicted spending; that is, as $(Actual - Counterfactual)/Counterfactual$.

[23] Note that given the estimated mean and standard deviation of prior beliefs, fewer than 5e-11% of households would be expected to have beliefs of $\beta \leq 1$.

Figure 6. Evolution of Beliefs about β Across Plan Year



Notes: Figure depicts average value of simulated β across the relative week of a plan year for the full sample, with 95% confidence intervals shaded in black. Simulations are performed based on the median equilibrium parameters of the model discussed in Section 5.2.

by the end of the plan-year, the average household value of β is estimated to be about 1.094, just outside of the confidence interval for the β in the non-learning model.

Figure A.10 in the Appendix presents results which further illustrate the heterogeneity in household beliefs across the year. The figure shows the fraction of households in the sample with simulated β greater than or equal to some threshold β_{min} for various thresholds. In general, extreme beliefs are rare after the first quarter of the year (fewer than 5% of households have an estimated $\beta \geq 1.5$ after week 10); however, many households stall in the convergence of their beliefs, with over a quarter of households estimating their OOP costs as at least 25% more expensive than the truth for the entire plan-year.

6.3 Counterfactual Exercises

The full model with household learning permits the same counterfactual simulations as above. We estimate predicted spending differences between the observed data (using the equilibrium model parameters) and three counterfactual states of the world: one where consumers are fully informed about their OOP costs at the time of service (e.g., where $\beta = 1$ for all household-weeks); one where priors are re-centered around the truth (e.g., where $\mu_{\beta,0} = 1$, but the other learning parameters remain unchanged); and one where deductibles reset more frequently than at the yearly level. While the first counterfactual exercise assumes full information, the second exercise allows for idiosyncratic differences across signals, but centered around the truth. That is, in each household-week, beliefs are centered around $\beta = 1$ but drawn randomly, with decreasing uncertainty over time. Finally, the exercises in which deductibles reset more frequently capture changes in both how often household uncertainty affects the estimated marginal costs of services and the length of time for which uncertainty is allowed to persist.

	Spending/Person			Conditional Differences	
	Average	Median	% Diff > 0	Average	Median
Panel A: Non-Learning Model					
Observed Data	\$2,007	\$1,126	–	–	–
Full Information ($\beta = 1$)	\$1,899	\$1,066	12.5%	\$863	\$487
Deductible Resets Quarterly					
Deductible Resets Monthly					
Panel B: Learning Model					
Observed Data	\$2,135	\$1,189	–	–	–
Full Information ($\beta = 1$)	\$1,899	\$1,068	22.4%	\$1,056	\$573
Re-centered Priors ($\mu_{\beta,0} = 1$)	\$1,895	\$1,064	23.4%	\$1,023	\$554
Deductible Resets Quarterly					
Deductible Resets Monthly					

Notes: Table presents average and median spending per household member predicted by the models outlined in Section 5 under different assumptions of the underlying structural parameters. Panel A uses the model described in Section 5.1 without learning, and Panel B uses the model outlined in Section 5.2. The first row in each panel indicates predicted spending using the observed choice data and the estimated equilibrium parameters presented in Section 6. The subsequent rows impose arbitrary assumptions on the parameter space to capture salient features of counterfactual scenarios, including full information without learning (row 2), learning with re-centered priors (row 3, panel B only), and policies shortening the length of a deductible. All currencies are reported in 2022 USD.

Table 5. Comparison of Predicted Spending Across Counterfactual States of the World

Table 5 presents the results.[24] Accounting for greater heterogeneity in household beliefs about OOP spending results in a greater fraction of individuals being affected by changes to the learning parameters. This makes intuitive sense, given that households restricted to no learning in the simpler model may have been estimated to have reasonably correct beliefs for the full plan year, when in fact they experienced a period of rapid learning early in the year. Compared to a state where β is fixed at 1, roughly 22% of households over-consume care, with the average (median) affected household consuming \$1,056 (\$573) more per household member. This corresponds to an over-spending of 45.0% (41.0%) for the average (median)

[24] See also Appendix Figure A.9 for a distribution of estimated spending differences.

household relative to the predicted spending under full information. This over-spending can be nearly entirely attributed to high household priors, as can be seen in row 3 of Panel B of the table. Re-centering household priors—without completely eliminating residual uncertainty around prices for each unique medical event—accounts for more than 95% of the over-spending for both the average and median household.

Our modeling exercise corroborates the findings of the empirical strategy: households over-estimate their actual medical expenditures while they wait for a bill, and these over-estimates are worse in cases where households have little information with which to form expectations. This results in significant over-consumption of care, particularly for households on the margin of meeting their deductible. More frequent signals, shorter waiting periods for pricing information, and more frequent deductible reset periods could all help reduce the effects of incorrect beliefs on over-consumption of care.

7 Discussion & Conclusion

This paper assesses how households respond to pricing information in their strategic decisions for future care consumption. We show that although households increase their spending following health events which may reduce the future (spot) marginal cost of care, they do so based on mis-informed over-estimates of actual spending. When a bill arrives with meaningful and accurate price information, households curtail their spending increases. This delayed pricing information meaningfully contributes to over-consumption of medical care, particularly general medical services which may spark downstream cascades of care.

We encapsulate our findings into a model of imperfect *ex-post* moral hazard with delayed learning from prices. Our model, just as our reduced-form evidence suggests, indicates that consumers over-inflate expectations of OOP spending before they receive bills, particularly early on in a plan year. Our model allows us to consider alternative plan designs—including more frequent deductible resetting—that might curtail the associated over-expenditures of such underinformation.

The analysis we present could be extended in several meaningful ways. First, future work could incorporate observed payment interactions between patients and physicians, rather than relying on claims data alone. Data on physician practices—including how quickly physicians submit claims to payers for medical claims and send bills to patients—may provide insights into both the source of variation in processing times as well as the potential policy benefits of reducing the length of provider billing cycles.

In addition, doing so would prevent any measurement error in the exposure variable arising from our poor proxy. In general, however, the measurement error associated with our proxy for bill arrival is likely to attenuate our estimates. This is not because the measurement error is classic, but instead based on the fact that any measurement error in the actual transmission of price information would result in contamination bias from the interim period, when households still do not know their OOP spending. If consumers over-estimate OOP prices before the bill arrives, any regressions misclassifying $\mathbb{1}\{\text{Post\_Bill}\} = 1$ when it should be 0 will attenuate the correction parameter $\beta_{\text{Post\_Bill}}$ to zero.

More generally, future research could build on the learning model presented here. This could include a more thorough treatment of heterogeneous learning across service types, or allowing the learning parameters to be covariate-dependent in other ways. In particular, exploring the health equity concerns associated with learning about prices could provide valuable insight in the persistence of health disparities in accessing even high-value preventive services (Teutsch et al., 2020; McMorrow et al., 2014). Finally, future work could explore the impact of real-times claim adjudication on consumer spending responses. This could be especially policy-relevant when exploring how heterogeneity across payers and providers (e.g., integrated care practices) could be used to leverage improved price transparency.

Increasing understanding of how consumers form expectations about their health needs and utilization is a vital component of designing optimal insurance contracts and health policies. Economic modeling and health policy alike are well-served from incorporating delayed learning as we assess how consumers make health decisions in real time.

References

Agarwal, R., Mazurenko, O., and Menachemi, N. (2017). High-deductible health plans reduce health care cost and utilization, including use of needed preventive services. *Health Affairs*, 36(10):1762–1768.

Agency for Healthcare Research and Quality (2007). Guide to Prevention Quality Indicators: Hospital Admission for Ambulatory Care Sensitive Conditions. Technical Report AHRQ Pub. No. 02-R0203, Department of Health and Human Services, Washington, D.C.

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

Arrow, K. J. (1963). Uncertainty and the welfare economics of medical care. *American Economic Review*.

Baicker, K., Mullainathan, S., and Schwartzstein, J. (2015). Behavioral hazard in health insurance. *The Quarterly Journal of Economics*, 130(4):1623–1667.

Blomqvist, Å. (1997). Optimal non-linear health insurance. *Journal of Health Economics*, 16(3):303–321.

Boyd, C. M. and Bellemare, M. F. (2020). The microeconomics of agricultural price risk. *Annual Review of Resource Economics*, 12(1):149–169.

Brot-Goldberg, Z. C., Chandra, A., Handel, B. R., and Kolstad, J. T. (2017). What does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics*. *The Quarterly Journal of Economics*, 132(3):1261–1318.

Cabral, M. (2017). Claim Timing and Ex Post Adverse Selection. *The Review of Economic Studies*, 84(1):1–44.

Caskey, R., Zaman, J., Nam, H., Chae, S.-R., Williams, L., Mathew, G., Burton, M., Li, J. J., Lussier, Y. A., and Boyd, A. D. (2014). The Transition to ICD-10-CM: Challenges for Pediatric Practice. *Pediatrics*, 134(1):31–36.

Chandra, A., Gruber, J., and McKnight, R. (2010). Patient Cost-Sharing and Hospitalization Offsets in the Elderly. *The American economic review*, 100(1):193–213.

Chernew, M. E., Rosen, A. B., and Fendrick, A. M. (2007). Value-Based Insurance Design. *Health Affairs*, 26(Supplement 2):w195–w203.

CMS (2019). Medicare and Medicaid Programs: CY 2020 Hospital Outpatient PPS Policy Changes and Payment Rates and Ambulatory Surgical Center Payment System Policy Changes and Payment Rates. Price Transparency Requirements for Hospitals to Make Standard Charges Public.

Colla, C. H. and Mainor, A. J. (2017). Choosing Wisely Campaign: Valuable For Providers Who Knew About It, But Awareness Remained Constant, 2014–17. *Health Affairs*, 36(11):2005–2011.

Colla, C. H., Morden, N. E., Sequist, T. D., Schpero, W. L., and Rosenthal, M. B. (2015). Choosing Wisely: Prevalence and correlates of low-value health care services in the United States. *Journal of General Internal Medicine*, 30(2):221–228.

Cooper, Z., Craig, S. V., Gaynor, M., and Van Reenen, J. (2019). The price ain't right? Hospital prices and health spending on the privately insured. *The Quarterly Journal of Economics*, 134(1):51–107.

Correia, S., Guimarães, P., and Zylkin, T. (2020). Fast Poisson estimation with high-dimensional fixed effects. *The Stata Journal*, 20(1):95–115.

Crawford, G. S. and Shum, M. (2005). Uncertainty and Learning in Pharmaceutical Demand. *Econometrica*, 73(4):1137–1173.

Cutler, D. M. and Zeckhauser, R. J. (2000). The anatomy of health insurance. In *Handbook of Health Economics*, volume 1, pages 563–643. Elsevier.

Dalton, C. M. (2014). Estimating demand elasticities using nonlinear pricing. *International Journal of Industrial Organization*, 37:178–191.

Dave, D. and Kaestner, R. (2009). Health insurance and ex ante moral hazard: Evidence from Medicare. *International journal of health care finance and economics*, 9(4):367–390.

Duarte, F. (2012). Price elasticity of expenditure across health care services. *Journal of Health Economics*, 31(6):824–841.

Dunn, A. (2016). Health insurance and the demand for medical care: Instrumental variable estimates using health insurer claims data. *Journal of Health Economics*, 48:74–88.

Edward, J., Wiggins, A., Young, M. H., and Rayens, M. K. (2019). Significant Disparities Exist in Consumer Health Insurance Literacy: Implications for Health Care Reform. *HLRP: Health Literacy Research and Practice*, 3(4):e250–e258.

Einav, L. and Finkelstein, A. (2018). Moral hazard in health insurance: What we know and how we know it. *Journal of the European Economic Association*, 16(4):957–982.

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

Elshaug, A. G., Rosenthal, M. B., Lavis, J. N., Brownlee, S., Schmidt, H., Nagpal, S., Littlejohns, P., Srivastava, D., Tunis, S., and Saini, V. (2017). Levers for addressing medical underuse and overuse: Achieving high-value health care. *The Lancet*, 390(10090):191–202.

Ganguli, I., Lupo, C., Mainor, A. J., Wang, Q., Orav, E. J., Rosenthal, M. B., Sequist, T. D., and Colla, C. H. (2020). Assessment of prevalence and cost of care cascades after routine testing during the Medicare annual wellness visit. *JAMA network open*, 3(12):e2029891–e2029891.

Geyman, J. P. (2012). Cost-Sharing under Consumer-Driven Health Care Will Not Reform U.S. Health Care. *The Journal of Law, Medicine & Ethics*, 40(3):574–581.

Globerman, S. (2016). Select cost-sharing in universal health countries. Technical report, Fraser Institute.

Goldman, D. and Philipson, T. (2007). Integrated Insurance Design in the Presence of Multiple Medical Technologies. *American Economic Review*, 97(2):427–432.

Gondi, S., Beckman, A. L., Ofoje, A. A., Hinkes, P., and McWilliams, J. M. (2021). Early Hospital Compliance With Federal Requirements for Price Transparency. *JAMA Internal Medicine*, 181(10):1396–1397.

Gourieroux, C., Monfort, A., and Trognon, A. (1984). Pseudo maximum likelihood theory. *Econometrica*, 52(3):681.

Gruber, J. (2022). Financing Health Care Delivery. *Annual Review of Financial Economics*, 14(1).

Hartzema, A. G., Racoosin, J. A., MaCurdy, T. E., Gibbs, J. M., and Kelman, J. A. (2011). Utilizing Medicare claims data for real-time drug safety evaluations: Is it feasible? *Pharmacoepidemiology and Drug Safety*, 20(7):684–688.

Henrikson, N. B., Chang, E., Ulrich, K., King, D., and Anderson, M. L. (2017). Communication with Physicians about Health Care Costs: Survey of an Insured Population. *The Permanente Journal*, 21.

Hoagland, A. (2022). An ounce of prevention or a pound of cure? The value of health risk information. *Working Paper*.

Hoagland, A. and Shafer, P. (2021). Out-of-pocket costs for preventive care persist almost a decade after the Affordable Care Act. *Preventive Medicine*, 150:106690.

Johnson, R. E., Goodman, M. J., Hornbrook, M. C., and Eldredge, M. B. (1997). The Effect of Increased Prescription Drug Cost-Sharing on Medical Care Utilization and Expenses of Elderly Health Maintenance Organization Members. *Medical Care*, 35(11):1119–1131.

Karlsson, N., Loewenstein, G., and Seppi, D. (2009). The ostrich effect: Selective attention to information. *Journal of Risk and uncertainty*, 38(2):95–115.

Knowles, J. C. (1995). Research note: Price uncertainty and the demand for health care. *Health Policy and Planning*, 10(3):301–303.

Korenstein, D., Falk, R., Howell, E. A., Bishop, T., and Keyhani, S. (2012). Overuse of health care services in the United States: An understudied problem. *Archives of Internal Medicine*, 172(2):171–178.

Kowalski, A. (2016). Censored quantile instrumental variable estimates of the price elasticity of expenditure on medical care. *Journal of Business & Economic Statistics*, 34(1):107–117.

Lee, J. L., Maciejewski, M. L., Raju, S. S., Shrank, W. H., and Choudhry, N. K. (2013). Value-Based Insurance Design: Quality Improvement But No Cost Savings. *Health Affairs*, 32(7):1251–1257.

Loewenstein, G., Friedman, J. Y., McGill, B., Ahmad, S., Linck, S., Sinkula, S., Beshears, J., Choi, J. J., Kolstad, J., Laibson, D., Madrian, B. C., List, J. A., and Volpp, K. G. (2013). Consumers' misunderstanding of health insurance. *Journal of Health Economics*, 32(5):850–862.

Manning, W. G. and Mullahy, J. (2001). Estimating log models: To transform or not to transform? *Journal of Health Economics*, 20(4):461–494.

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

Mazurenko, O., Buntin, M. J., and Menachemi, N. (2019). High-deductible health plans and prevention. *Annual Review of Public Health*, 40(1):411–421.

McMorrow, S., Kenney, G. M., and Goin, D. (2014). Determinants of receipt of recommended preventive services: Implications for the Affordable Care Act. *American Journal of Public Health*, 104(12):2392–2399.

Mehrotra, A., Dean, K. M., Sinaiko, A. D., and Sood, N. (2017). Americans Support Price Shopping For Health Care, But Few Actually Seek Out Price Information. *Health Affairs*, 36(8):1392–1400.

Muir, M. A., Alessi, S. A., and King, J. S. (2012). Clarifying costs: Can increased price transparency reduce healthcare spending. *William & Mary Policy Review*, 4:319.

Naqvi, S. A. A. (2021). SCHEMEPACK: Stata module providing ready-to-use graph schemes. Boston College Department of Economics.

Ngangoue, M. K. (2021). Learning From Unrealized versus Realized Prices. *American Economic Journal: Microeconomics*, 13(2):174–201.

Pauly, M. V. and Blavin, F. E. (2008). Moral hazard in insurance, value-based cost sharing, and the benefits of blissful ignorance. *Journal of Health Economics*, 27(6):1407–1417.

Peng, L. (2005). Learning with information capacity constraints. *Journal of Financial and Quantitative Analysis*, 40(2):307–329.

Reed, M., Fung, V., Brand, R., Fireman, B., Newhouse, J. P., Selby, J. V., and Hsu, J. (2005). Care-Seeking Behavior in Response to Emergency Department Copayments. *Medical Care*, 43(8):810–816.

Reed, M., Fung, V., Price, M., Brand, R., Benedetti, N., Derose, S. F., Newhouse, J. P., and Hsu, J. (2009). High-deductible health insurance plans: Efforts to sharpen a blunt instrument. *Health Affairs*, 28(4):1145–1154.

Santos Silva, J. M. C. and Tenreyro, S. (2006). The log of gravity. *The Review of Economics and Statistics*, 88(4):641–658.

Shafer, P. R., Hoagland, A., and Hsu, H. E. (2021). Trends in Well-Child Visits With Out-of-Pocket Costs in the US Before and After the Affordable Care Act. *JAMA Network Open*, 4(3):e211248.

Shafer, P. R., Horný, M., and Dusetzina, S. B. (2022). Monthly Cost-Sharing Limits and Out-of-pocket Costs for Commercially Insured Patients in the US. *JAMA Network Open*, 5(9):e2233006.

Snowbeck, C. (2022). Risk adjustment, claims processing woes still troubling Bright Health Group. *Star Tribune*.

Stockley, K. (2016). Evaluating Rationality in Responses to Health Insurance Cost-Sharing: Comparing Deductibles and Copayments. *Working Paper*.

Teutsch, S., Carey, T. S., and Pignone, M. P. (2020). Health equity in preventive services: The role of primary care. *American Family Physician*, 102(5):264–265.

Webb Hooper, M., Mitchell, C., Marshall, V. J., Cheatham, C., Austin, K., Sanders, K., Krishnamurthi, S., and Grafton, L. L. (2019). Understanding multilevel factors related to urban community trust in healthcare and research. *International Journal of Environmental Research and Public Health*, 16(18):3280.

Winter, R. A. (2013). Optimal Insurance Contracts Under Moral Hazard. In Dionne, G., editor, *Handbook of Insurance*, pages 205–230. Springer, New York, NY.

Zhang, A., Prang, K.-H., Devlin, N., Scott, A., and Kelaher, M. (2020). The impact of price transparency on consumers and providers: A scoping review. *Health Policy*, 124(8):819–825.

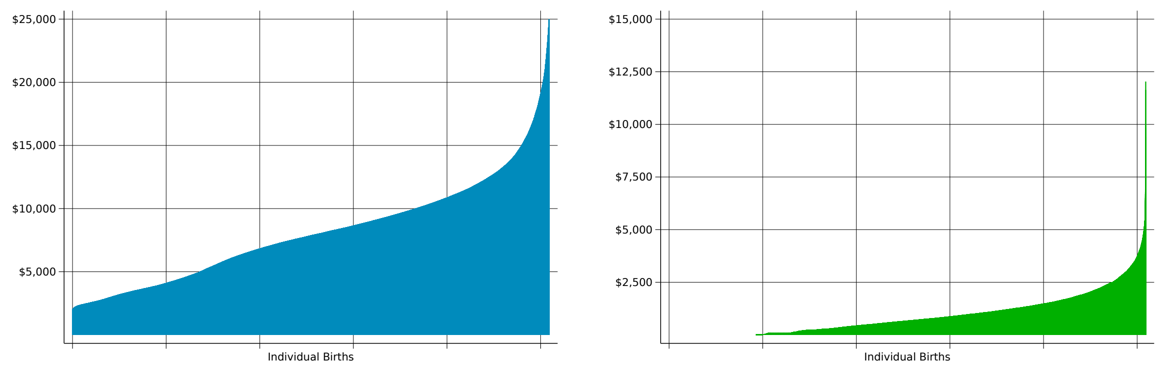
Zhang, X., Haviland, A., Mehrotra, A., Huckfeldt, P., Wagner, Z., and Sood, N. (2018). Does enrollment in high-deductible health plans encourage price shopping? *Health Services Research*, 53(S1):2718–2734.

A Appendix

Figure A.7. Variation in Prices for CPT 59400: Routine Vaginal Delivery

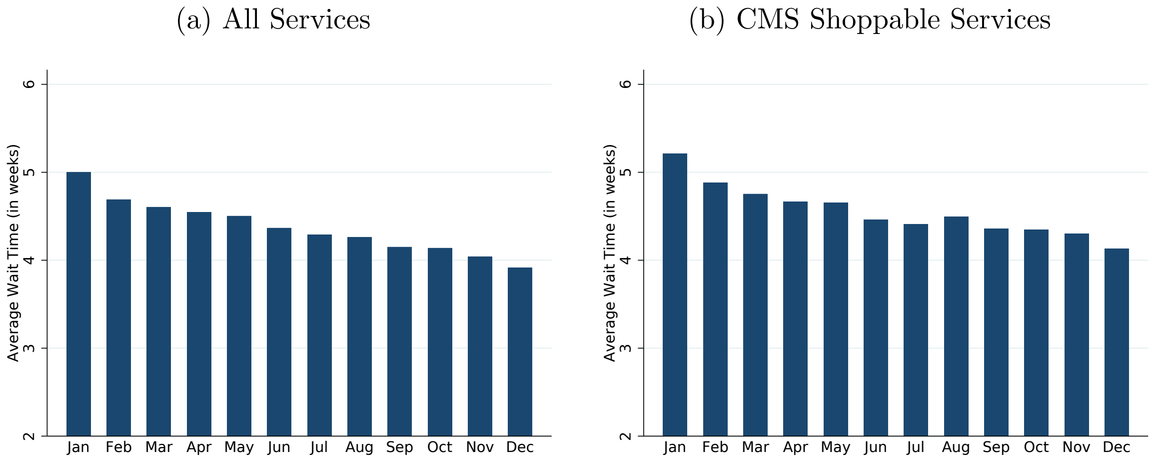
(a) Total Costs

(b) OOP Costs



Notes: Figures show variation in total and OOP costs associated with CPT code 59400: "Routine obstetric care including antepartum care, vaginal delivery (with or without episiotomy, and/or forceps) and postpartum care." Each vertical bar represents a unique encounter in our analytical data set, with the height of the bar corresponding to the price (all measured in 2022 USD).

Figure A.8. Variation in Wait Times for Bills



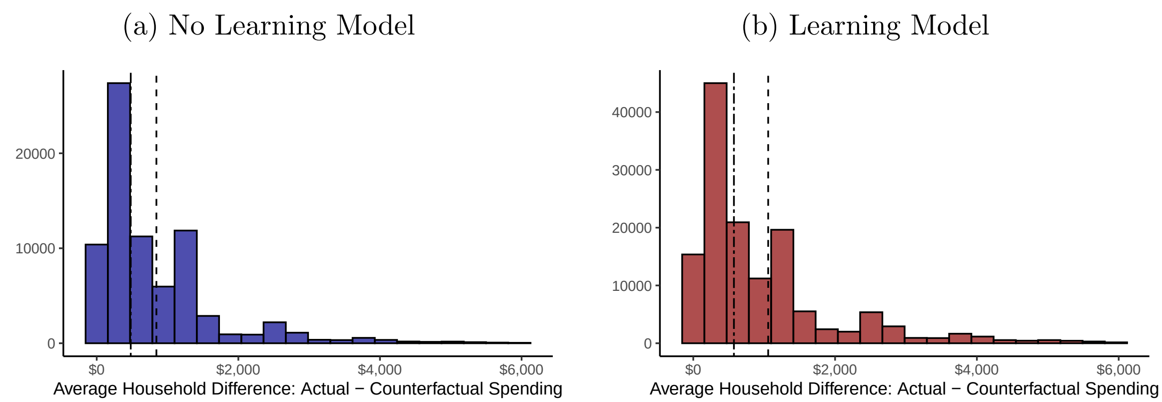
Notes: Indicates average wait time (in weeks) between date of service and date the insurer paid their portion of the claim (the earliest date at which definitive OOP information is known). Panel (a) illustrates variation in average wait times across months of the year (pooled across all years) for all claims in the analytical data; panel (b) limits the sample to only the shoppable services used as major health events in the text.

Type	Code	Service Description
DRG	216	Cardiac valve and other major cardiothoracic procedures w/ cardiac catheterization
DRG	460	Spinal fusion, except cervical
DRG	470	Major joint replacement or reattachment of lower extremity
DRG	473	Cervical spinal fusion
DRG	743	Uterine and adnexa procedures for non-malignancy
CPT	19120	Removal of 1 or more breast growth, open procedure
CPT	29826	Shaving of shoulder bone using an endoscope
CPT	29881	Removal of one knee cartilage using an endoscope
CPT	42820	Removal of tonsils and adenoid glands (patient younger than age 12)
CPT	43235	Diagnostic examination of esophagus, stomach, and/or upper small bowel
CPT	43239	Biopsy of the esophagus, stomach, and/or upper small bowel using an endoscope
CPT	45378	Diagnostic examination of large bowel using an endoscope
CPT	45380	Biopsy of large bowel using an endoscope
CPT	45385	Removal of polyps or growths of large bowel using an endoscope
CPT	45391	Ultrasound examination of lower large bowel using an endoscope
CPT	47562	Removal of gallbladder using an endoscope
CPT	49505	Repair of groin hernia (patient age 5 years or older)
CPT	55700	Biopsy of prostate gland
CPT	55866	Surgical removal of prostate and surrounding lymph nodes using an endoscope
CPT	59400	Routine obstetric care for vaginal delivery
CPT	59510	Routine obstetric care for cesarean delivery
CPT	59610	Routine obstetric care for vaginal delivery after prior cesarean delivery
CPT	64483	Injections of anesthetic and/or steroid drug into lower or sacral spine nerve root
CPT	66821	Removal of recurring cataract in lens capsule using laser
CPT	66984	Removal of cataract with insertion of lens
CPT	93000	Electrocardiogram, routine, with interpretation and report
CPT	93452	Insertion of catheter into left heart for diagnosis
CPT	62322	Injection of substance into spinal canal of lower back or sacrum
CPT	62323	Injection of substance into spinal canal of lower back or sacrum

Notes: Table shows list of procedures used to identify non-urgent "shoppable services," which are the exposure of interest in the primary reduced-form specifications. Services are identified based on lists provided by the Center for Medicare and Medicaid Services (CMS), using the relevant Diagnostic Related Groups (DRGs) or Current Procedural Terminology (CPT) codes to identify procedures.

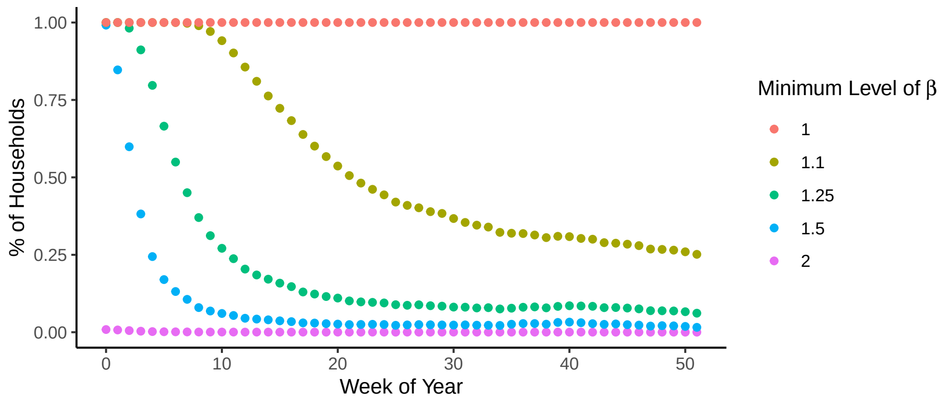
Table A.1. Shoppable Services Used in Analytical Sample

Figure A.9. Counterfactual Analysis: Change in Predicted Spending from Correcting $\beta = 1$



Notes: Figure shows estimated the differences in predicted (total) per-person health spending that arise from requiring that $\beta = 1$ in Equation 6, rather than the parameters estimated in the models (see Figure 5). Panel (a) shows results for the model without learning while panel (b) shows results for the generalized learning model. Histogram displays distribution of household-year average differences per person, conditional on a difference greater than 0. Note that for 87.5% of households in panel (a) and 77.6% of households in panel (b), no differences in spending are predicted. The dashed line indicates the average conditional difference in per-person spending while the dot-dashed line indicates the median in both groups.

Figure A.10. Evolution of Beliefs about β Across Plan Year



Notes: Figure depicts the fraction of households in the sample with simulated β greater than or equal to some threshold $\beta_{\min}$ for various thresholds. Simulations are performed based on the median equilibrium parameters of the model discussed in Section 5.2.