2022

Essays on information and innovation in health economics

BOSTON UNIVERSITY

GRADUATE SCHOOL OF ARTS AND SCIENCES

Dissertation

ESSAYS ON INFORMATION AND INNOVATION IN HEALTH ECONOMICS

by

ALEXANDER HOAGLAND

B.S., B.A., Brigham Young University, 2017

Submitted in partial fulfillment of the

requirements for the degree of

Doctor of Philosophy

2022

Approved by

First Reader

Randall P. Ellis, PhD
Professor of Economics

Second Reader

Tal Gross, PhD
Associate Professor of Markets, Public Policy & Law

Third Reader

Marc Rysman, PhD
Professor of Economics

Fourth Reader

Jihye Jeon, PhD
Assistant Professor of Economics

Acknowledgments

I am enormously indebted to the support of wonderful advisors who assisted me in my PhD. I would like to thank Randy Ellis for his unfailing enthusiasm for my work and attention to detail; Tal Gross for providing invaluable feedback on my raw ideas and even more invaluable encouragement; Marc Rysman for giving a keen eye to the intricacies of structural estimation while simultaneously helping them to come alive as tools of research; and Jihye Jeon for her careful consideration of my ideas and thoughtful, heartening advice. I would also like to thank other faculty in the Department of Economics and the Questrom School of Business for generously giving of their time for both research and general advice. These include (but are not limited to): Keith Ericson, James Feigenbaum, Raymond Fisman, Jetson Leder-Luis, and Daniele Paserman.

I am grateful for the members of my dissertation writing group for providing a network of support through our collective dissertation writing process: Corinne Andriola, Marshall Drake, Shomik Ghosh, Laurie Hakes, and Patrick Power. I also thank the participants of many internal seminars, including the health economics reading group, the industrial organization reading group, and the empirical microeconomics reading groups, who provided useful feedback on many aspects of these essays.

Finally, and most importantly, I am eternally indebted to my perfect family: my wife Kristen and daughter Gwendolyn. Without them, I would not have finished this degree (or had a reason to do so). I am grateful to them for talking me into taking days off to go on summer bike rides, taking in stride the stresses of graduate school, and providing the best companionship I could ask for.

Alex Hoagland

Doctoral Candidate

Department of Economics

ESSAYS ON INFORMATION AND INNOVATION IN HEALTH ECONOMICS

ALEXANDER HOAGLAND

Boston University, Graduate School of Arts and Sciences, 2022

Major Professor: Randall P. Ellis, PhD
Professor of Economics

ABSTRACT

This dissertation consists of three essays that study the role of information acquisition and processing in health decision-making. Each chapter underscores the ways in which new information shapes the choices of health providers and consumers. Understanding these responses sheds light on critical health policy problems, including the potential overuse of low-value health services, gaps between medical evidence and practice, and inequitable access to high-value health services.

The first essay studies the role of a consumer's family network in the formation of their risk perceptions. I assess whether people correctly interpret new risk information communicated through household health events and analyze how these responses impact household welfare. Individuals respond to new diagnoses in ways most consistent with individual reevaluations of health risk rather than other possible explanations. To assess welfare implications, I estimate a structural model of health choices in which individuals learn about risk after health events reveal information. I find that consumers over-respond to recent, salient health events by over-weighting their risks *ex-post*. This leads to individual and social welfare losses, and suggests

that aiding consumers in interpreting health risk information should be an important aim of health literacy policies.

The second essay explores how health providers respond to information about innovations in mental health treatments, paying particular attention to the heterogeneous adoption costs of different practices. I compare the impact of continuing education on takeup across innovations that incur learning costs (psychotherapy) and those that do not (psychopharmacology). I use a novel extension of an estimator proposed by Calvi et al. (2021) to estimate a dynamic treatment effect in the presence of classification error. Therapists respond more to education when learning costs are negligent, being about three percentage points more likely to write new prescriptions following a conference.

The third essay assesses the tradeoff between adopting novel medical technologies and achieving health equity. I study the adoption of transcatheter valve replacement surgeries in Medicare patients; these surgeries disrupted the supply of medical interventions from cardiothoracic surgeons to interventional cardiologists. This transition led providers to adjust practice styles along two margins: medium-risk patients became more likely to receive surgery, and low-risk patients received fewer medical interventions overall. I incorporate these findings into a model of physician decision-making, showing that both the expansion of high-intensity intervention and the crowd-out of low-intensity treatment can be rationalized by the presence of technological spillovers. The model further highlights that crowd-out may be inequitably distributed across the patient population when treatment appropriateness is not directly observed. I validate these predictions in my setting, showing that technology adoption resulted in disproportionately high barriers to care for low-income patients.

Contents

List of Tables

List of Figures

List of Abbreviations

2SLS	.....	Two-stage least squares
AR	.....	Autoregressive model
ATE	.....	average treatment effect
BLS	.....	Bureau of Labor Statistics
CABG	.....	Coronary artery bypass graft surgery
CAD	.....	Coronary artery disease
CME	.....	Continuing medical education
CPT-4	.....	Current Procedural Terminology, fourth edition
D-LATE	.....	Dynamic local average treatment effects
DDIV	.....	Difference-in-differences with instrumental variables
DX	.....	Diagnosis
ED	.....	Eating disorder
ESI	.....	Employer-sponsored insurance
ESIV	.....	Event study with instrumental variables
FBT	.....	Family-based therapy
FDA	.....	United States Food and Drug Administration
FFS	.....	Fee-for-service
HCC	.....	Hierarchical Condition Category
HHI	.....	Herfindahl-Hirschman Index
HSA	.....	Health Services Area
ICD-9-CM	.....	International Classification of Diseases, Ninth Revision,
	.....	Clinical Modification
ICD-10-CM	.....	International Classification of Diseases, Tenth Revision,
	.....	Clinical Modification
ICED	.....	International Conference on Eating Disorders
KFF	.....	Kaiser Family Foundation
LATE	.....	Local average treatment effect
MBSF	.....	Master Beneficiary Summary File
MR-LATE	.....	Mismeasurement-Robust Local average treatment effect
MSA	.....	Metropolitan Statistical Area

OOP	.....	Out-of-pocket
PCI	.....	Percutaneous coronary intervention
RCT	.....	Randomized control trial
RPG	.....	Research-to-practice gap
SAVR	.....	Surgical aortic valve replacement
SSRI	.....	Selective serotonin reuptake inhibitors
STS	.....	Society of Thoracic Surgeons
STS-PROM	.....	Society of Thoracic Surgeons Predicted Risk of Mortality
TAVR	.....	Transcatheter aortic valve replacement
TWFE	.....	Two-way fixed effects
USD	.....	United States Dollars
WTP	.....	Willingness to pay

Chapter 1

An Ounce of Prevention or a Pound of Cure? The Value of Health Risk Information

1.1 Introduction

Social networks provide important information for consumers making health care choices. Through connections with family, friends, and neighbors, individuals form expectations of their own health risks, learn about the value of specific medical practices, and identify how or from whom to receive care. Family relationships provide particularly influential sources of health information due to their close proximity and the high relevance of their health experiences, as both shared genetic profiles and lifestyle choices influence expected health care consumption. Understanding how individual health experiences shape family health behaviors is essential for policies aiming to improve public health, such as those attempting to address high levels of health care spending or the takeup of high-value health services.

One especially salient dimension of health information individuals may learn from family health experiences is knowledge about health risks, including both current and expected future health care needs. Individuals may choose to seek out high-value, life-saving care after witnessing a family member's health experience, particularly if that experience reveals their own risk. For instance, some may choose to become

vaccinated against COVID–19 once a family member becomes infected Chen (2021); Giardinelli (2021); Salcedo (2021).

There exists evidence that family members react to the health events of their loved ones (Fadlon and Nielsen, 2019; Hodor, 2021), but it remains unclear what information drives these reactions. Health events may lead individuals to reassess their specific health risks, but individuals may also respond to other features of an event. These include changes to the expected price of medical care, preferences for health consumption, or knowledge about the availability of health services. Understanding the role that social connections play in both the utilization of high-return medical care and the propagation of low-value services relies on separating these competing effects. In particular, assessing the welfare effects of transmitting new health information requires both understanding whether individuals respond to information itself and the extent to which they update their beliefs correctly.

In this paper, I examine how consumers who receive health risk signals through witnessing a major health event within their household—such as a diagnosis with a new chronic condition—modify their assessments of their own risks and alter their choices accordingly. I study households with employer-sponsored insurance (ESI) obtained through large employers between 2006 and 2018. Highly-detailed claims data provides insight into how individuals respond to quasi-random health events in their family, including overall responses in plan choices and health spending as well as decisions about the use of specific services. Additionally, these data include rich variation in coverage generosity and plan characteristics among enrollees, an important fact I leverage to separate changes in household beliefs about risk from other, potentially confounding, effects of health events.

I show that major health events generate strong informational spillovers among non-diagnosed household members. Those exposed to new health information signif-

icantly and persistently increase both their overall health care utilization and their investments in preventive care, particularly for services that are specific to the condition just diagnosed in their household. I show that these spillover effects are more consistent with individual learning than other potential mechanisms. The magnitude of these increases is constant across insurance plan designs—including plans without deductibles—suggesting that moral hazard concerns are not driving changes.[1] Additionally, chronic events induce stronger and more persistent behavior changes than acute health events, suggesting that salience effects arising from a traumatic health experience do not fully explain observed results (Dalton et al., 2020; Fadlon and Nielsen, 2019). Finally, I show that even individuals who are most familiar with the health care system—such as those taking regular preventive medications for cardiovascular health—are responsive to major health events, implying that learning about health *systems*, rather than health *risk*, is not the main driver of observed results.[2]

In general, one would expect receiving new information about one's risk to lead to improvements in decision-making and welfare. Surprisingly, however, I demonstrate that the welfare effects of new information are not obvious from reduced-form analysis alone. Affected household members increase their use of "low-value" health services, procedures that are generally agreed to be cost ineffective due either to their reach (e.g., benefiting few patients) or their average returns (e.g., low levels of benefits relative to costs) (Colla et al., 2015). Households responding to chronic diagnoses are most likely to increase their utilization of low-value services that appear, from their perspective, closely related to preventive care, including cardiac screenings before low-risk surgeries or imaging services for lower back pain. In addition, households

[1] As is common in the health economics literature, I use the phrase "moral hazard" to denote induced-demand effects arising from changes in the price an individual faces for care. For a more in-depth discussion of this abuse of notation, see Einav et al. (2013).

[2] This general learning may include systematic learning about health care organizations, the process of receiving insurance coverage for care, or building physician relationships (Sabety, 2020).

do not alter their insurance plan choices even after large expected increases in health costs from managing chronic conditions. Both of these findings cast doubt on the extent to which health information improves choice quality.

These findings motivate a structural approach to model the evolution of household decisions following health events and quantify the associated welfare effects of receiving health information. I write and estimate a model in which households form beliefs about their health risks over time. In my model, households first make decisions about their insurance coverage prior to receiving information about their health state in a period; once this information is realized, households choose health spending (Cardon and Hendel, 2001; Einav et al., 2013; Marone and Sabety, 2022). Novel to my model, health shocks take two forms: major health events and non-chronic health fluctuations. Major health events occurring in a household induce other members to update beliefs about their health risks, but also affect consumer choices by potentially lowering the conditional cost of non-chronic care and increasing risk aversion. A structural approach allows me to separately identify these competing effects, yielding clear estimation of the welfare effects from receiving health information.

A key challenge in my model is identifying changes in an individual's beliefs about their health risks separate from these alternative explanations. I use multiple sources of variation in the data to decompose the effects of household health events. First, I use a broad set of health events which vary in their expected treatment costs to identify the effects of price changes on spending decisions. More expensive conditions (e.g., cancers) are associated with stronger price effects than cheaper ones (e.g., asthma) and therefore are expected to induce stronger moral hazard responses. Second, I exploit variation in the availability and generosity of plans offered to households to separately identify changes in household risk aversion at the time of plan choice. Here, the intuition is that individual beliefs about health determine optimal medi-

cal spending and coverage levels, while household risk aversion also determines the gradient of preferred coverage as the price or generosity of plans vary (Ericson et al., 2020). I complement this approach with additional information about the circumstances of a diagnosis (e.g., whether a hospitalization occurred) to further model risk preferences and risk beliefs separately. Finally, I use both acute and chronic health events to assess the extent to which individuals learn more generally about the health care system, rather than the causal effect of new information about health risks.

Counter to expected thought, the new information gained from health events is not welfare-improving for many households. In fact, new health risk information lowers expected household utility by an average of $2,788 per year. The central insight of the model is that there is a tension between the seriousness of a major health event and the appropriate level to which individuals should update their beliefs: new diagnoses in a household spur overly large changes in an individual's assessment of their health risks, resulting in average posterior beliefs that are well above the average in-sample risk of diagnosis. Counterfactual simulations suggest that bounding these changes in risk beliefs substantially increases consumer welfare: 86% of the households in my sample would find health information welfare-improving were their responses mitigated. Finally, I demonstrate that the societal value of communicating health information can be improved by selectively revealing it to specific groups, such as those with higher *ex-ante* risk. This suggests that population health information campaigns—including genetic testing programs and screening practices for important conditions such as COVID-19—can benefit from targeting specific groups.

My analysis contributes to a burgeoning discussion on the causal spillover effects of health information within social networks, particularly the family. The importance of family relationships in economic decision-making has been well-documented in labor supply and education choices (Browning et al., 2014; Altmejd et al., 2021), but the role

of these relationships in forming health behaviors is not as well understood. Previous work has suggested that an individual's social network informs their decision-making following acute health events (Bouckaert et al., 2020; Hodor, 2021; Song, 2021), health trials (Archibong and Annan, 2021), and infectious disease outbreaks (Agüero and Beleche, 2017).[3]

I contribute to this discussion in three ways. First, I highlight a new type of health information to which individuals are highly responsive: household chronic conditions. I provide evidence that individuals are even more responsive to these chronic diagnoses than to household acute health events. Second, I explore the mechanisms behind these responses, showing that changes to how individuals assess their health risks appear to drive observed spending changes. Finally, I provide evidence that while health events increase investments in high-value care, they are also associated with large errors in risk assessments and the takeup of low-value care, resulting in welfare losses for households on average.

I also contribute to a growing literature that incorporates learning and preferences in structural models of health behavior (Barseghyan et al., 2018; Bundorf et al., 2021a). I incorporate the findings of this literature into the first structural model addressing the value of health information spillovers, and highlight the particular behaviors—such as information misinterpretation—that dampen potential welfare gains. My model incorporates a fully flexible specification for misinterpretation of information (Hauser and Bohren, 2021), and encompasses previous work highlighting the role major health events play in inducing demand responses by changing spot prices for other care (Eichner, 1997; Kowalski, 2016). Additionally, I make use of

[3]A rich literature has highlighted how individuals respond to information about their own health risks, including their own diagnosis. For an in-depth review of this literature, see Alalouf et al. (2019). Some previous work has demonstrated that certain diagnoses can have dramatic impacts (Almond et al., 2010); however, examinations of other diagnoses revealed a lack of noticeable responses (Dupas, 2011; Kim et al., 2019).

previous identification results to simultaneously estimate weighted probabilities and standard risk aversion parameters in a nonlinear framework (Ericson et al., 2020).[4]

Related to this, I also contribute to a literature on non-Bayesian learning, which emphasizes the disproportionate weight put on recent, and particularly salient, events (Kahneman and Tversky, 1973). This literature emphasizes the role of individual over- and under-reactions to new signals, and how this affects the ultimate convergence of individual beliefs (Epstein et al., 2010). Models that incorporate such ideas include Holt and Smith (2009), who find in an experimental setting that individuals significantly overweight new evidence (relative to typical Bayesian predictions) when it had a lower *ex-ante* probability of occurring. Other important models draw attention to biased beliefs in models of consumer choices, including their role in rationalizing choices that would otherwise require unreasonably high levels of risk aversion (Ortoleva, 2012; Paserman, 2008; Spinnewijn, 2015).

My model highlights that over- or under-reactions can be accommodated *ex-ante* in a quasi-Bayesian framework by varying the timing of belief updating. In addition, I simultaneously estimate biased beliefs and risk preferences, providing a microfoundation of how individuals form beliefs in a setting of largely small-probability events. My model provides additional insight into the development of subjective health beliefs; in particular, I provide new evidence that explains why consumers may be better at predicting their relative risk rather than their absolute risk (Bundorf et al., 2021b), and how biases in assessing their own health risks may arise (Arni et al., 2021).

Finally, my work is relevant to the well-established literature exploring suboptimal health decisions made by most consumers (Abaluck and Gruber, 2011, 2016a; Abaluck and Compiani, 2020; Baicker et al., 2015; Handel, 2013; Handel and Kolstad,

[4]See Barseghyan et al. (2013) and their later review paper Barseghyan et al. (2018) for a more thorough discussion of the literature estimating models of probability weighting in other settings in economics.

2015; Iizuka et al., 2021; Ketcham et al., 2012). This literature includes an ongoing discussion about the extent to which improving health information generally may improve decision-making (Abaluck and Gruber, 2016b; Cutler and Zeckhauser, 2004; Gruber et al., 2020). My analysis reveals that some health signals—such as major health events—do little to align household choices with the value of medical care, and may instead lead to an increase in the over-utilization of services that provide little or no benefit to households. Hence, simply improving access to health information may shift consumers only from one type of poor decision-making to another, while increasing total health spending. Additionally, my paper underscores the role of behavioral economics in structural models assessing the quality of consumer choices. I show that including factors such as belief discounting may help to explain why overcoming information frictions is not simply a matter of increased access to health information.

I present my empirical setting and data in Section 1.2. Following a discussion of major health events, I provide evidence of their spillover effects and the potential mechanisms driving them in Section 1.3. Then, to evaluate the welfare effects associated with these responses, I present the details of my model in Section 1.4 and its results in Section 1.5. The model output informs several counterfactual analyses assessing the role of consumer responsiveness to information, which I present in Section 1.6. Finally, I discuss the relevance of my findings and directions for future work in Section 1.7.

1.2 Empirical Setting & Data

My primary data on household plan choice, health utilization, and major medical events come from the IBM/Truven Marketscan *Commercial Claims and Encounters* Data. These data contain detailed inpatient, outpatient, and pharmaceutical claims for a sample of households enrolled in ESI through large U.S. firms which contracted

with participating payers. Each observation includes diagnostic, procedural, and payment information, as well as household, firm, and insurance plan identifiers. I obtained data from 2006 to 2018, with the exception of plan identifiers, which are only available until 2013. Throughout, spending data has been normalized to 2020 USD using the Consumer Price Index for All Urban Consumers series.

My final sample includes households with two or more members observed for two or more years and insured with one of eight large firms. I required that each household have full eligibility and continuous enrollment across their window of observation. My final sample consists of 353,403 households and 5,439,482 individual-year observations.

Table 1.1 presents summary statistics for the full sample as well as the subset of the sample with insurance plan identifiers. It is important to ensure that the two samples are relatively balanced given that I use only the plan-identified sample in my structural estimation (Section 1.4). In general, the two groups have similar demographics, spending trends, and health states. A notable exception is that households in the plan-identified sample incur lower out-of-pocket (OOP) costs than the full sample, suggesting that they possess more generous insurance coverage on average. However, this is likely due to time trends arising from the fact that the plan-identified sample runs only through 2013. Medical spending, as expected, is highly skewed, with average annual household spending in the range of $2,500 compared to a median of about $400. Observed switches in plan choices are low, consistent with prior work (Handel, 2013).

1.2.1 Major medical events

I model the ways households respond to information about their health risk communicated through major health events within the family. I identify these events based on observed diagnostic codes in the claims data, using a subset of the Department

Table 1.1
Household Summary Statistics

	Full Sample	Plan-Identified Sample
Panel A: Household demographics		
Family size	3.0 (0.00)	3.0 (0.00)
Employee age	45.0 (0.01)	44.4 (0.01)
Enrollee age	30.9 (0.01)	30.4 (0.01)
% female employees	41.6 (0.00)	42.4 (0.00)
% female enrollees	50.2 (0.00)	50.3 (0.00)
Panel B: Medical spending & plan choices		
Total medical spending	\$2,504.41 [\$679.75] (4.51)	\$2,454.88 [\$624.16] (7.12)
OOP medical spending	\$443.07 [\$109.66] (0.53)	\$337.98 [\$80.33] (0.89)
% individuals w/ zero spending	15.4 (0.00)	16.6 (0.00)
% individuals w/ zero OOP	21.0 (0.00)	22.2 (0.00)
% switching plans	–	5.3 (0.00)
Panel C: Major medical events		
% experiencing chronic diagnosis	6.3 (0.00)	5.2 (0.00)
% experiencing acute event	1.0 (0.00)	0.6 (0.00)
Diagnosis OOP, chronic	\$1,082.05 [\$464.69] (11.59)	\$854.62 [\$329.90] (17.72)
Diagnosis OOP, acute	\$2,494.42 [\$1,419.91] (68.05)	\$2,107.09 [\$964.62] (122.50)
Recurring OOP, chronic	\$983.03 [\$521.39] (17.32)	\$683.60 [\$446.69] (19.20)
Years	2006–2018	2006–2013
N_{families}	353,403	179,044
$N_{\text{individuals}}$	1,087,353	555,733

Notes: Values based on Marketscan claims data, 2006–2018. Enrollees are employees plus their covered dependents. Spending values are reported in 2020 USD. Standard errors are reported in parentheses and sample medians (when reported) are in brackets.

of Health and Human Services' Hierarchical Condition Categories (HCCs). These HCCs, which are typically used in risk adjustment models, identify a basic set of chronic illnesses that may alter overall health utilization and spending. I limit my classification of health events to non-pregnancy HCCs that occur with high frequency as discussed in Appendix A.2.

To ensure that I identify new diagnoses, I require that relevant diagnosis codes appear during or after an individual's second observed year. Additionally, I drop households for which the diagnosed individual is not present for at least a full year after their medical event to exclude individuals who might have passed away during or shortly after their event.

An important feature of my analysis is the separate treatment of health costs for major medical events, including the costs associated with maintaining the health of someone with a chronic condition. To measure these costs, I collaborated with Rebecca Hughes, MD, to identify a set of disease-specific procedures and prescriptions associated with each health condition in my sample.[5] I then identify household spending on these health events based on the claims for these procedures and prescriptions, both in the year of diagnosis and following years. As reported in Table 1.1, the average (median) household in my sample spends $683.60 ($446.69) out-of-pocket on recurring costs needed to care for chronic conditions.

1.2.2 Plan characteristics

Heterogeneity in each household's choice of plans provides a plausibly exogenous source of variation in how major medical events and chronic health costs impact household spending decisions. I exploit the claims data to estimate the characteristics of each plan in my households' choice sets, which will be important inputs in my theoretical model.

[5]Appendix A.3 lists the relevant codes used for each diagnosis.

I define a household's plan choice set at the firm-state-year level, and limit attention to plans covering at least five percent of all covered lives within a firm-year to rule out executive plans.[6] In reality, health plans are defined by a complicated set of cost-sharing measures, including copayment and coinsurance rates that vary widely across provider specializations, networks, and procedures. For tractability, my structural model takes in a simplified version of these measures: a family deductible, a simplified non-specialist coinsurance rate, and a family OOP maximum. I construct measures for each plan's individual and family deductibles based on the empirical distribution of payments in the claims data (Zhang et al., 2018). I then estimate the other two cost-sharing parameters as those that minimize the sum of squared residuals between predicted and observed OOP spending for households within each plan year (Marone and Sabety, 2022). Appendix A.1 describes this methodology in more detail and evaluates the quality of these inferences. I find that these simplified measures capture a wide degree of variation in my data and harmonize well with measures from earlier work. Finally, I estimate each plan-year's family premium as the average cost of all households enrolled in the plan over a year, and assume that employee premium contributions are consistent with the national averages for household coverage (on average about 28% of the household premium; KFF (2020)).

There is substantial variation across firms, regions, and years in the generosity of coverage offered to employees, which I describe in Table 1.2. As I describe in Section 1.4.3, such variation provides an intuitively useful means of attributing household behaviors to changes in risk *preferences* versus risk *beliefs*; households who are more risk averse tend to minimize their overall variation in *ex-ante* expenditures by choosing

[6]My data does not distinguish whether there exist plan "tiers" within firms (for example, a university that offers one set of plans to its faculty and a different set to its graduate students). These unobserved barriers may cause measurement error in the plan choice sets used in the structural model in Section 1.4; however, such error would not affect any of my primary results, which focus on how new health information alters spending choices conditional on the choice of plan.

more generous health plans, while households who are less risk averse but believe they are at higher risk for major health events may choose less-generous plans overall that instead provide more targeted coverage. The average household has between two and four plans to choose from in a given year, with a wide degree of variation in the average family deductible. This variation is comprised of both heterogeneity in the frequency with which firms offer zero-deductible health plans as well as in the size of nonzero deductibles. Similar variations exist in other plan characteristics, including copayment rates and OOP maxima.

Table 1.2
Average Plan Characteristics, 2006–2013

	Firm							
	A	B	C	D	E	F	G	H
# of plans offered	3.5	2.5	3.0	2.0	2.0	2.6	2.8	3.0
Spending/Enrollee ($000s)	12.7	9.8	9.7	10.2	9.3	8.9	9.1	11.5
Family deductible ($000s)	0.4	0.4	2.1	1.0	1.0	0.7	0.9	0.5
% of 0-deductible plans	64.3	46.7	0.0	0.0	0.0	22.2	31.8	38.9
Family OOP max. ($000s)	3.5	4.6	5.1	5.9	4.3	4.1	5.2	3.9
HHI of all plans	0.4	0.6	0.4	0.6	0.9	0.6	0.7	0.4

Notes: Averages are pooled across all plans and years in a given firm.

1.3 Spillover Effects of Household Health Events

This section presents my main reduced-form empirical results. I first show that after experiencing a chronic major health event, households increase their overall medical utilization by about 10% annually, as well as increasing their investment in billed spending on preventive care. I illustrate that the observed responses are consistent with a reevaluation of one's own risk by showing that households are more likely to invest in preventive care that is specific to the illness their family member experienced. I then consider other potential mechanisms, including financial incentives, salience effects, and general learning about the health care system. Finally, I show that household members increase their utilization of "pseudo-preventive" low-value

services—such as extraneous screenings and imaging services—showing that while health events generate strong spending responses, these responses are not necessarily targeted at high-return services.

1.3.1 Induced spending changes

To estimate the causal impact of health shocks on health choices, I first estimate two-way fixed effects (TWFE) "event study" regressions of the following form:

$$\sinh^{-1}(y_{ft}) = \alpha_f + \tau_t + \sum_{k=-T}^{T} \gamma_k \mathbb{1}\left\{t - E_{ft} = k\right\} + \epsilon_{ft}. \tag{1.1}$$

The variable y_{ft} represents a spending outcome for a household f in year t; in my main specification, this outcome is annual OOP payments made by all family members *except* those who experience the major health event. I adjust for highly-skewed distributions of spending variables by using the inverse hyperbolic sine transformation.[7] An added advantage of this transformation is that the resulting regression coefficients can be interpreted as approximate percentage changes in the outcome variable, relative to the year prior to the shock. I include household and year fixed effects, as well as dummy variables indicating when an observation occurred relative to E_{ft}, a household's event year. The coefficients on these indicator variables, $\{\gamma_k\}$, are the objects of interest. I also adjust for potentially correlated responses within a household by clustering standard errors at the household level.

This approach allows me to identify the potentially time-varying effects of health shocks—which might have decaying influence on household choices over time—while

[7] I use the inverse hyperbolic sine transformation to accommodate the approximately 15% of individual-years in my data with 0 spending (Harris and Stöcker, 1998). Bellemare and Wichman (2020) show that for a model with continuous variables x and y and specification $\sinh^{-1}(y) = \beta x + \varepsilon$, the elasticity of y with respect to x is $(\beta x/y)\sqrt{y^2+1} \approx \beta x$ whenever $y \geq 2$. Bellemare and Wichman (2020) also discuss the ways using this measure may refine estimates using the more common $\log(y+1)$ transformation. I show in Appendix B that my results are not substantively altered when using the logarithm transformation.

simultaneously controlling for any unobserved household- or year-specific deviations in behavior. However, recent work has highlighted that TWFE estimators can be difficult to interpret without strong modeling assumptions (Callaway and Sant'Anna, 2018). In particular, coefficients estimated by TWFE models represent the weighted average of many two-by-two comparisons. When treatment effects are heterogeneous across groups—and hence, these comparisons—some comparisons may be assigned negative weights (de Chaisemartin and D'Haultfoeuille, 2019; Goodman-Bacon, 2018). This makes the interpretation of estimated treatment effects—static or dynamic—difficult to interpret. Furthermore, when estimating dynamic treatment effects, researchers must take care that dynamic parameters of interest (including both pre-trends and estimated time-varying treatment effects) are separately identified from time fixed-effects included in the regression (Borusyak and Jaravel, 2021; Sun and Abraham, 2021). Without including a control group of observations which are never treated, separate identification of time fixed effects and dynamic treatment effects is impossible.

I demonstrate that my analysis is robust to both concerns. First, I show that my coefficients of interest do not suffer from problems of negative weighting by considering a number of additional specifications in Appendix B. These include both robust estimators proposed by de Chaisemartin and D'Haultfoeuille (2019) and Sant'Anna and Zhao (2020), as well as simple recentered time series graphs and standard difference-in-differences coefficients.[8] This provides evidence that my results are not idiosyncratic to my estimation method; rather, my results appear even in the raw data.

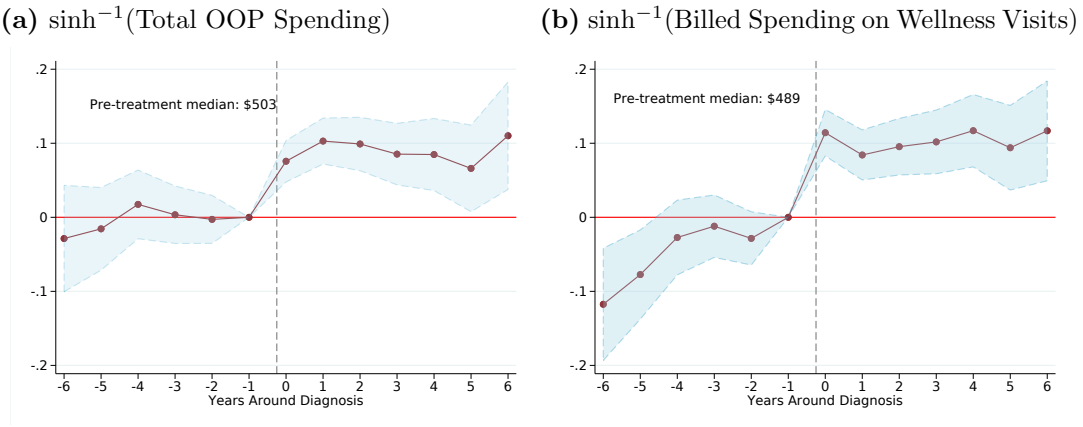
Second, I utilize a large control group in my sample, allowing me to separately identify the time-varying treatment effects from yearly fixed effects. Previous work examining health spillovers within families has restricted the control group to only

[8]Using the Bacon decomposition reveals that the estimands in my primary specification are not constructed using negative weights (Goodman-Bacon et al., 2019). However, I present these additional robustness results for completeness.

those who experience a similar diagnosis in the future in order to utilize a control group that more closely matches the treatment group on unobservable characteristics. I include never-treated households in my sample in order to identify dynamic treatment effects. The central tradeoff in doing so lies in the validity of the parallel trends assumption: namely, that in the absence of major health events, the treated and control groups would continue to have similar spending and utilization trajectories. Given that my setting spans a large range of chronic conditions—many of which are neither directly related to health behaviors or particularly life-threatening—concerns about violations of the parallel trends assumption are less plausible in my setting.

Figure 1·1

Effect of Chronic Diagnoses on Non-Diagnosed Household Members' Spending



Notes: These figures show estimated coefficients and 95% confidence intervals for the effect of a new chronic diagnosis on medical spending. In both panels, the sample includes spending for all household members without major health events. In panel (a), the dependent variable is the inverse hyperbolic sine of total OOP spending; panel (b) estimates the effect on total spending (insurer spending + OOP spending) on wellness visits only. Coefficients are presented relative to the year prior to diagnosis. Spending is measured in 2020 USD. Standard errors are clustered at the household level.

Figure 1·1 presents the time-varying causal effect of a health shock on household OOP spending for all non-diagnosed individuals. The first panel illustrates that non-diagnosed household members increase their annual OOP spending by about 10% relative to the year prior to the event. For the median (average) household, this

corresponds to an increase of about $50 ($115) annually. This effect begins in the year of the health event and persists following the diagnosis. Additional results in Appendix B corroborate this finding with other outcome variables including total billed spending or visit frequencies.

Importantly, this increase in utilization encompasses an increased investment in preventive care. The second panel of Figure 1·1 illustrates this by limiting the scope of the analysis to household spending only on wellness visits. Wellness visits are non-problem-based visits with a family or primary care physician that are generally recommended about once a year; these visits include routine screenings for important chronic conditions including cancers and mental health conditions. These visits constitute an important jumping-off point for the use of other preventive services (Jiang et al., 2018) and are therefore generally considered to be an important form of high-value care (Tong et al., 2021). Here, too, I find that new diagnoses in a household are associated with strong responses. Affected, non-diagnosed household members increase their overall spending on wellness visits by about 10%, matching the increase in overall utilization.[9]

1.3.2 Changes as responses to new health risk information

These results suggest a meaningful, persistent change in how non-diagnosed household members engage with the health care system. I first show that these responses are indicative of household members updating their beliefs about their own health risks following the receipt of health information from a major event. Such observed

[9]Even before the Affordable Care Act (ACA)'s cost-sharing exclusion took effect in 2010 (or 2012 for certain women's health services), OOP costs for preventive care were steadily declining for those with ESI (Hong et al., 2017). Once the ACA took effect, the majority of wellness visits should be free to enrollees in my sample (Shafer et al., 2021), a feature I observe in the data. Although time fixed effects in the regression specification should absorb these trends for both pre- and post-ACA trends, I use billed spending rather than OOP spending as my outcome variable of interest. Note that in my data set, billed spending represents the sum of individual OOP payments and insurer payments to the provider; it does not reflect any price negotiations or other discounts that were provided at the time of service, and therefore does not reflect the listed prices of services.

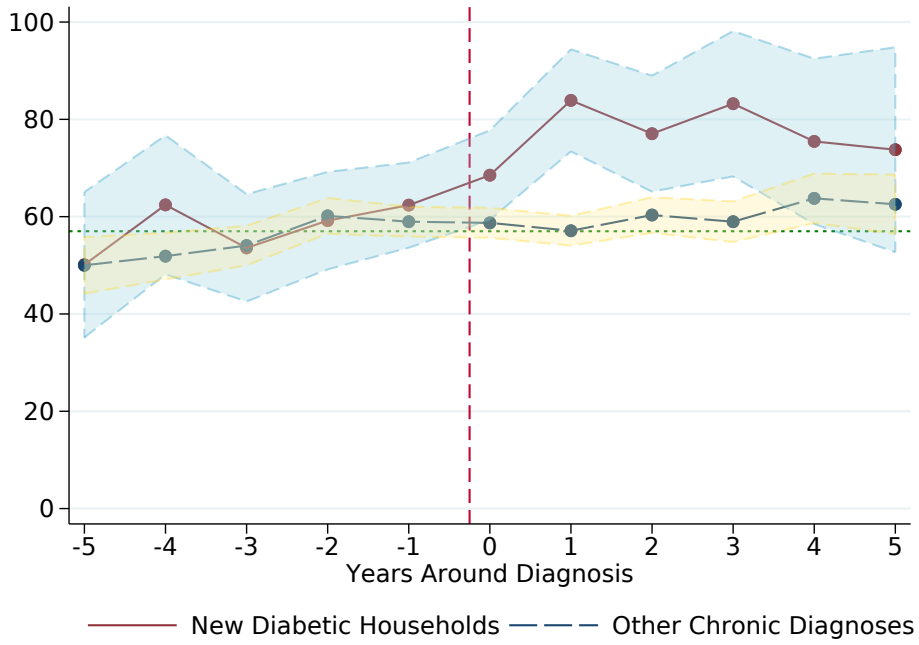
responses could also be driven by factors beyond changes in a household's assessment of their health risks, including changes in the price of care, salience effects, overall exposure to the health care system, or improved physician relationships. I explore these alternative mechanisms in Section 1.3.3.

To more explicitly explore the link between major health events and risk beliefs, I estimate the causal effects of health shocks on preventive services that are specific to an affected household's diagnosis. Here, the intuition I rely on is that household exposure to risk information is more targeted than other forms of health information; hence, the extent to which I observe households selecting into preventive services that are disease-specific rather than general provides evidence of responses specifically to new risk information.

For example, individuals who have learned that they are at increased risk for developing diabetes may have a higher likelihood of seeking out screenings for abnormal blood glucose levels than individuals who have learned that they are at increased risk for another chronic condition. Figure 1·2 plots re-centered time series that depict the associations between household diagnoses and the takeup of diabetes screenings for adults within a household. The figure plots average utilization rates of diabetes screenings for two groups: those who are exposed to a diabetes diagnosis in their home and those who are exposed to a different chronic diagnosis. Individuals whose family members are diagnosed with conditions other than diabetes do not appear to significantly alter their screening behaviors from unaffected households (whose average is depicted in the horizontal, dotted green line). On the other hand, household members of those diagnosed with diabetes increase screenings in the first three years following the diagnosis, being about 36% more likely to be screened for diabetes than unaffected individuals.

Figure 1·2
Rate of Diabetes Screenings Around Time of Diagnoses

(a) Diabetes Screenings (Rate/1,000 Adults)



Notes: Figure shows average utilization rates of diabetes screenings for non-diagnosed household members 18 years of age and older, measured in rates per 1,000 adults. Point estimates and 95% confidence intervals are presented. The top (solid maroon) line indicates average rates for households who experience a diabetes diagnosis, and the bottom (dashed navy) line indicates rates for those affected by other chronic diagnoses. The horizontal, dotted green line indicates the average utilization rate for all other households in the sample who do not experience a diagnosis, about 59 screenings per 1,000 adults.

To assess the causal effect of multiple diagnoses simultaneously on the utilization of disease-specific preventive care, I use a triple-differences approach. This approach disentangles two competing effects: those arising from experiencing any chronic illness (e.g., salience effects) and a disease-specific informational effect. I estimate the effect of a new chronic diagnosis on a household f's decision to screen for a specific diagnosis

d during time t, as summarized in Equation 1.2:

$$\begin{aligned} Pr(\text{Screening})_{fdt} =& \beta_{\text{DD}}(\text{post}_t \times \text{chron}_f) + \beta_{\text{DDD}}(\text{post}_t \times \text{chron}_f \times \mathbb{1}\{\text{chron}_f = d\}) \\ &+ \alpha_f + \tau_t + \varepsilon_{fdt}, \end{aligned} \tag{1.2}$$

where $chron_f$ is a dummy variable indicating whether *any* chronic diagnosis occurred within the household and $post_t$ indicates periods following a diagnosis. The triple interaction variable includes an additional constraint that the chronic diagnosis $chronic_f$ match the specific diagnosis d (e.g., a diabetes diagnosis when the outcome variable is a diabetes screening). Hence, β_{DD} identifies the effect of any chronic diagnosis on screening, while the triple interaction β_{DDD} identifies the effect for the specific diagnosis of interest relative to other diagnoses.[10] For example, using this approach I can estimate the impact of a diabetes diagnosis on diabetes screenings as $\beta_{\text{DD}} + \beta_{\text{DDD}}$, where β_{DD} indicates the impact of experiencing any chronic diagnosis in the household on diabetes screenings and β_{DDD} indicates the specific differential effect of a new diabetes diagnosis occurring in the household.

The triple difference approach is advantageous because it allows me to compare the causal effect of diagnoses on the use of preventive care across multiple control groups. When the outcome variable of interest is a screening for a specific service (e.g., diabetes), this approach estimates the effect of a corresponding diagnosis relative to all other diagnoses, for which the screening reveals no information. In this context, the identifying assumption for the triple differences approach is the same as the identifying assumption for the simpler difference-in-differences regressions: that spending

[10]The sum of the coefficients $\beta_{DD} + \beta_{DDD}$ identifies the diagnosis-specific effect of receiving a diagnosis, relative to all non-diagnosed households in my sample. Notice that, in Equation 1.2, all requisite interaction terms for the triple differences are either subsumed in the fixed-effects or colinear with the included variables given the unique structure of my treatment variables.

differences between diagnosed and undiagnosed households would have evolved similarly over time in the absence of treatment.[11]

I estimate several versions of this regression for various diagnosis-screening pairs. I select diagnoses and screenings which are commonly utilized and for which there are clear diagnostic codes available. I examine the impact of new diabetes and cancer diagnoses on their respective screenings, as well as the effect of diabetes diagnoses on cholesterol screenings. I also assess the impact of any new chronic diagnosis in a household on the rate of new hypertension diagnoses, relative to all major health events.[12]

Finally, to verify my results, I estimate this model for screenings for which health events communicate little useful information, and hence are expected to change behavior little. This might be because a diagnosis doesn't require a doctor's visit to diagnose (e.g., obesity) or doesn't require preventive screening prior to seeking treatment (e.g., mental health conditions, such as major depressive disorder). Hence, observing a lack of response among these types of preventive services serves to underscore the role that health information, specifically, plays in altering individual behavior. I include "placebo" regressions for the effect of new diabetes diagnoses on obesity diagnoses and the effect of new mental health disorder diagnoses on screenings for depression.

Table 1.3 presents the estimation results from these six regressions in two panels. First, I highlight that new chronic diagnoses alter specific preventive behaviors in

[11]When adding the triple interaction, the identifying assumption is modified only to include the assumption that spending differences between households diagnosed with one condition and households diagnosed with another would have evolved similarly in the absence of treatment, a statement which is subsumed in the initial identifying assumption. Appendix B includes standard difference-in-differences regression results that corroborate the findings reported here.

[12]Given that there is no procedure code for hypertension screenings, this approach proxies the effect of the risk information associated with chronic diagnoses on new general wellness screenings, relative to the other forms of health information accompanying acute events. Coding practices reduce my ability to test this finding for each individual diagnosis in my sample; for example, there are no diagnostic or procedure codes used exclusively for asthma screenings.

cases where they transmit important information about health risk. The occurrence of any chronic diagnoses in a household is associated with a 19.4% increase in the rate of hypertension diagnoses among other affected household members. Furthermore, specific diagnoses such as cancer and diabetes increase the likelihood that a non-diagnosed household member will seek out screening by 13.2% and 21.1%, respectively. Finally, diabetes diagnoses are associated with an increase in cholesterol screenings of 7.2%. Similar to previous work, I find evidence that new diagnoses reduce the rate of other, unrelated screenings (Fadlon and Nielsen, 2019); for example, a non-diabetes chronic diagnosis is associated with a 7.4% *decline* in the rate of diabetes screenings among non-diagnosed household members. These effects, however, are typically smaller than the estimated increases in disease-specific screenings, suggesting that this crowding out is not necessarily one-to-one.

The second panel of Table 1.3 reports results for placebo regressions including obesity diagnoses and depression screenings. Here, I find no strong evidence that health events alter screenings. This is consistent with the notion that individuals respond by altering their use of preventive care only when the major health event communicates health risk information that necessitates preventive care utilization. Other dimensions of a health event—such as learning about the role of preventive care in medical maintenance overall—do not appear to drive individual behavior changes, at least in the use of preventive services.

I report additional results in Appendix B. I find that in addition to selecting screenings based on the health risk information they receive, households are selective in which members they choose to screen. I utilize variation in intrafamilial relationships and corresponding risk to show that households screen those who are most affected by the new health information. When households are affected by a chronic illness with a strong genetic component, such as type 1 diabetes, children and siblings

Table 1.3
Effect of Chronic Diagnoses on Take-Up of Disease-Specific Preventive Care

Own Screening (Dependent Variable)	Household Diagnosis	Pre-Diagnosis Average	Effect of Any Diagnosis (β_{DD})	Effect of Specified Diagnosis (β_{DDD})
Panel A: Main Effects				
Hypertension[1]	Any Chronic[2]	2.01	-0.27**	0.39***
		(0.007)	(0.102)	(0.110)
Cancer	Cancer	20.72	-0.01	2.74***
		(0.021)	(0.113)	(0.509)
Diabetes	Diabetes	6.21	-0.46***	1.31***
		(0.012)	(0.086)	(0.279)
Cholesterol	Diabetes	17.01	-0.22	1.23***
		(0.019)	(0.126)	(0.389)
Panel B: Placebo Regressions				
Obesity[1]	Diabetes	1.04	0.02	0.10
		(0.005)	(0.035)	(0.110)
Depression	Depression	0.36	-0.01	-0.08
		(0.003)	(0.037)	(0.077)

Notes: Table presents results from six triple-difference regressions highlighting the role of household investments in disease-specific preventive care following adverse health events. Each regression uses as its outcome variable a binary indicator for the screening listed in the first column, and a binary indicator for the event in the second column as its treatment variable (see Equation 1.2 for the full specification). Regression coefficients for the typical difference-in-difference effect (β_{DD}) indicate the effect of any chronic health event on screenings; the triple differences coefficients (β_{DDD}) indicate the effect of the specific diagnosis on screening choices. Robust standard errors clustered at the household level shown in parentheses. [1] Due to unavailability/low-use of CPT-4 procedure codes for screenings, these outcomes are measured as new ICD-9-CM/ICD-10-CM diagnosis codes. [2] Here, the reference group is all acute major health events. $^{*}p < 0.05$,$^{**}p < 0.01$,$^{***}p < 0.001$

of the affected individual are more likely to be screened than other household members. On the other hand, diagnoses such as type 2 diabetes—which has a stronger lifestyle component than a genetic one—are associated with more frequent screenings for spouses. Taken together, the observed ways in which major health events affect the use of preventive care are all consistent with a model where households interpret new diagnoses as signals of their own health risk, altering their behaviors accordingly.

1.3.3 Alternative explanations for spending changes

Although individuals appear highly responsive to new information about their own risk, additional factors could separately cause or exacerbate observed changes in health spending, including moral hazard effects, salience effects, and learning about the health care system. In this section, I explore each of these potential competing explanations and show that they are each insufficient to explain my observed results.

Moral hazard

A natural response to observing the phenomenon illustrated in Figure 1·1 is to conclude that the spending increase is driven by induced demand responses among the non-diagnosed individuals. A chronic diagnosis—such as diabetes— implies consistent, predictable costs on a household—such as through insulin prescriptions and endocrinologist visits. These additional costs, which are largely fixed for the individual, shift the cost-sharing characteristics of a health plan for the rest of the household, effectively lowering their spot price of future (non-chronic) health care. These induced-demand responses have been studied within families experiencing sudden acute health shocks that unexpectedly meet their household deductible (Eichner, 1998; Kowalski, 2016).

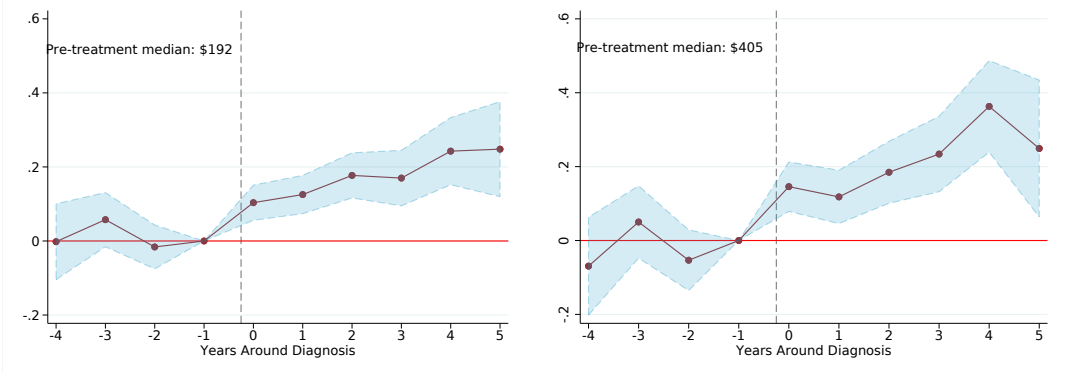
Two features of the results suggest that these induced-demand responses are unlikely to be the principal driver of the results. First, the costs of a chronic diagnosis

are typically larger in the year of diagnosis than in future years, especially when a hospitalization is required to diagnose the illness or there are acute complications that must be dealt with. This would suggest that if other household members were responding to changes in care prices alone, their responses would be much larger closer to the diagnostic event, and more muted in following years. Figure 1·1 does not show this to be true, either for overall utilization or the use of wellness visits specifically. Second, Figure 1·3 illustrates that non-diagnosed individuals respond to health shocks even when those shocks do little to change their spot price of medical care. Were moral hazard responses the principal mechanism of response, households in these plans would have much weaker incentives to adjust their choices.[13]

Figure 1·3

Effect of Chronic Diagnoses on Spending: Households Facing Zero Deductible

(a) $\sinh^{-1}$(Total OOP Spending) **(b)** $\sinh^{-1}$(Billed Spending on Wellness Visits)



Notes: These figures show estimated coefficients and 95% confidence intervals for the effect of a new chronic diagnosis on medical spending. This figure uses a limited sample of only households enrolled in health insurance plans with zero deductible at the time of the event. In both panels, the sample includes spending for all household members without major health events. In panel (a), the dependent variable is the inverse hyperbolic sine of total OOP spending; panel (b) estimates the effect on total spending (insurer spending + OOP spending) on wellness visits only. Coefficients are presented relative to the year prior to diagnosis. Spending is measured in 2020 USD. Standard errors are clustered at the household level.

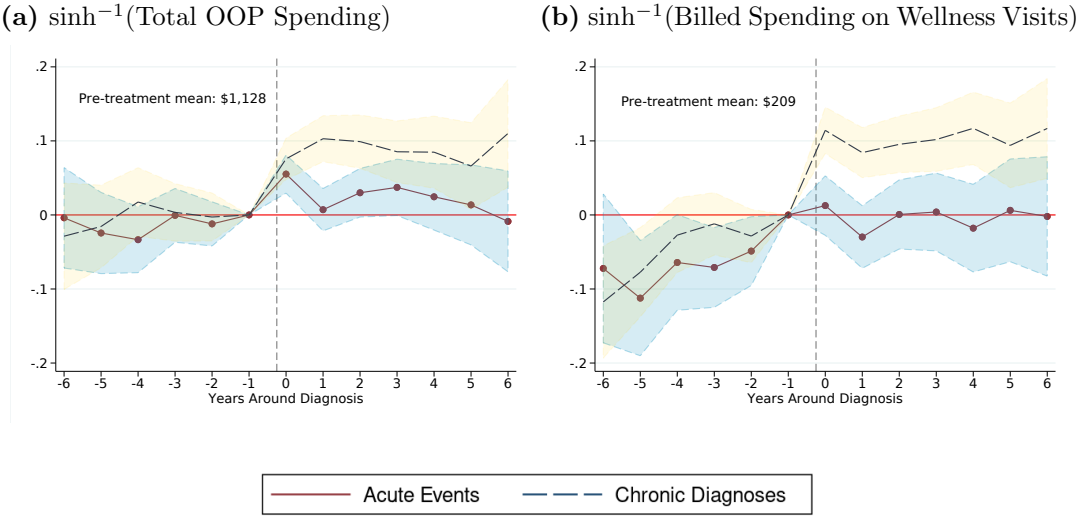
[13] A corresponding result for the subset plans with nonzero deductibles is included in Appendix B. Additional results in this Appendix show that families who are closer to meeting their deductibles prior to a health event are not more likely to increase their spending than those for whom chronic care costs may not meaningfully change family cost-sharing rates.

The effect of salience

It may also be that the intensity of major health events realigns household preferences to prioritize medical care. Individuals who experience the hospitalization of a household member may (over-)respond to the trauma of the event itself, changing their health consumption behaviors in order to avoid future hospitalizations. The critical difference is that when individuals respond to this health trauma, health events alter an household's risk *preferences* by affecting their marginal utility of medical care, rather than affecting risk *beliefs*.

To examine the impacts of these salience effects relative to risk reassessments, I analyze the responses of individuals who experience acute, rather than chronic, health events in their households. These include hospitalizations for family members who experience severe viral infections or other serious conditions unrelated to chronic disease. I use health events that are still assigned HCCs to capture health events of a similar level of seriousness to new chronic diagnoses; however, these events do *not* communicate any information to household members about health risks. Comparing observed household responses to these acute events against responses to chronic diagnoses allows me to assess the extent to which new health risk information alters behavior beyond salience.

Figure 1·4
Effect of Acute Health Events on Non-Diagnosed Household Members' Spending



Notes: These figures show estimated coefficients and 95% confidence intervals for the effect of a new acute hospitalization on medical spending. The solid maroon line indicates estimates from an acute event; the dashed navy line presents estimated results from Figure 1·1 as a reference. In both panels, the sample includes spending for all household members without major health events. In panel (a), the dependent variable is the inverse hyperbolic sine of total OOP spending; panel (b) estimates the effect on total spending (insurer spending + OOP spending) on wellness visits only. Coefficients are presented relative to the year prior to diagnosis. Spending is measured in 2020 USD. Standard errors are clustered at the household level.

Figure 1·4 presents the results. I find that, unlike new chronic diagnoses, acute hospitalizations spur few changes in health behaviors among other household members. Acute hospitalizations are associated with a short-term increase in spending of about five percent (from a baseline of about $1,100) in the year of the diagnosis, but these effects do not persist across time. Acute health events are also not associated with increased investments in preventive care for other household members. In particular, Figure 1·4 compares these regression coefficients to those estimated in response to new chronic diagnoses (Figure 1·1). I find that chronic health events are associated with overall spending responses almost twice as large as for acute hospitalizations, differences which are significant at the 95% confidence level for the first

three years following diagnosis. Furthermore, chronic diagnoses induce significantly more investment in preventive services for the first five years following a diagnosis.

Given that acute hospitalizations make health care at least as salient—if not more so—than chronic diagnoses, these findings suggest that changes in risk preferences arising from a "health scare" are insufficient to entirely explain changes in behavior. Rather, new health risk information, such as about one's inherent genetic risk for a chronic condition, appear to drive observed changes.

Health information

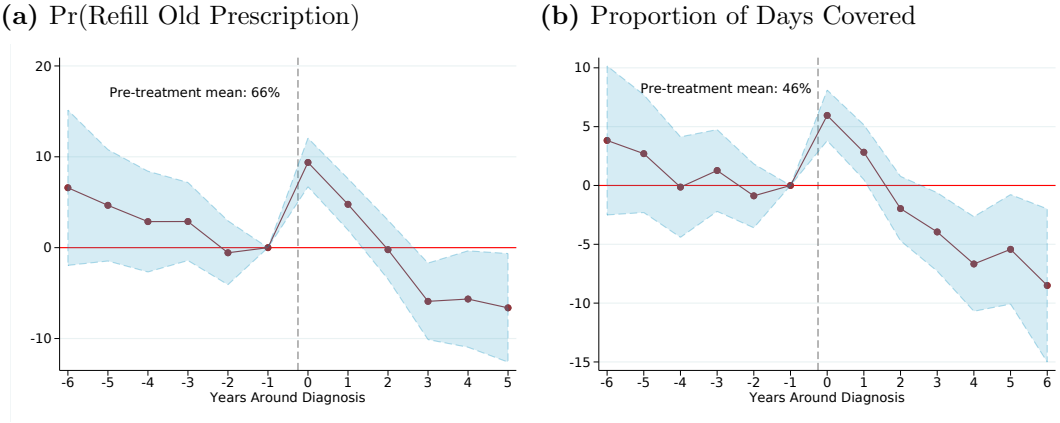
New diagnoses may also alter spending patterns by providing families with more general health information, such as information about the value of medical care, the process of obtaining covered care through an insurer, or how to establish strong provider relationships. Generally, learning about health risks and this more systematic learning imply similar responses among affected individuals, making their effects difficult to disentangle.

I focus on a particular case where new diagnoses provide risk information without more systematic information: non-diagnosed household members who were taking medications to prevent cardiovascular disease *prior* to the diagnosis within their family. Cardiovascular preventive drugs, including statins and other cholesterol-lowering drugs, are an extremely common class of medications and are known to be effective in preventing future health problems when used appropriately (O'Connor, 2006).[14] In this analysis, I limit my sample to those who have filled a prescription for these medications at least once per year during their first two years in the sample. I then measure the effects of chronic diagnoses on utilization and adherence among refills of these prescriptions.

[14]Appendix Table A.5 contains a detailed list of the therapeutic classes used in my sample.

This setting provides a unique environment in which to disentangle the effects of general learning about health systems and learning about one's own health risk. Individuals with existing prescriptions already have sufficient knowledge about the health care system to receive this care from their provider and insurer. Hence, while major health events provide them with information about the potential value of adherence to their medication (along with the potential consequences for not doing so), these events are unlikely to provide new knowledge about how to obtain this medication.

Figure 1·5
Effect of Chronic Diagnoses On Adherence to Existing Preventive Medications



Notes: These figures show estimated coefficients and 95% confidence intervals for the effect of a new diagnosis on adherence to preventive medications whose prescriptions were first written prior to the major health event. The sample is limited to all non-diagnosed individuals who filled preventive cardiovascular medications at least once per year during their first two years in the sample. In the first panel, the dependent variables is a binary indicator for whether the prescription was refilled at all. The second panel uses the proportion of days covered by any preventive cardiovascular medication as the outcome variable (Choudhry et al., 2009). Coefficients are presented relative to the year prior to diagnosis. Standard errors are clustered at the household level.

I show, however, that new diagnoses alter adherence to these prescriptions. I estimate the effect of a chronic diagnosis on both the likelihood of any medication use and overall adherence, measured as the proportion of days covered in a year (Choudhry et al., 2009). This measure is standard in the literature on adherence, and corresponds to the fraction of the year after a patient's first prescription fill for

which the patient has a supply of the medication. One concern in identifying the effect of new diagnoses on adherence is that prescription adherence may decay over time in response to barriers such as financial concerns or apathy (Slejko et al., 2014). Importantly, this may occur at different rates for different individuals both within and across households, meaning that these trends would not be accounted for using only household and year fixed effects. I therefore add a variable controlling for the number of years an individual has been in the sample to Equation 3.23.

Figure 1·5 presents the estimated dynamic treatment effect of a new chronic diagnosis on adherence to existing preventive prescriptions. As expected, in the absence of new health information, individuals become less adherent to prescriptions over time. However, diagnoses in the household spur a resurgence in both the likelihood that individuals will fill their prescriptions at all and the proportion of days covered: affected individuals are around ten percentage points more likely to refill their prescription in the year of a major medical event than in the year before, translating to an additional eight percentage point increase in the average proportion of the year for which they are covered by the prescription. The fact that new diagnoses change individual adherence to prescriptions even among a population which has access to and knowledge of specific preventive care illustrates that individuals are learning about more than just how to obtain care. The estimated causal "re-adherence" to prescriptions is consistent with individuals reevaluating the value of their medication given new information about their health risks.[15]

1.3.4 Quality of induced spending changes

Major health events generate strong spillover effects within a household on both overall utilization patterns and preventive care investments. It is natural, therefore, to

[15]These effects are likewise observed in the sample of households with zero deductible, suggesting that this re-adherence is also not exclusively driven by moral hazard responses.

ask how these responses are distributed within a larger framework of health spending. Do major health events contribute to more informed decisions about the type of care consumers choose to utilize? Or does the salience associated with health trauma lead to further over-utilization of low-return services? I address these questions by examining household use of services typically deemed as "low-value" by medical professionals and health officials (Chua et al., 2016; Colla et al., 2015).[16] Low-value services include both services whose cost typically outweighs any benefits to an average patient (e.g., unnecessary surgeries such as arthroscopies) as well as services which are chronically over utilized in ways that dramatically lower their return (e.g., imaging services such as MRI services for chronic migraines). Avoiding the use of these services can result in an overall higher quality of health care through both cost reductions and the avoidance of unnecessary risks.

I find that new chronic diagnoses are associated with an increase in overall low-value spending of about 5 percent (Appendix B). However, these results mask significant heterogeneity across different types of low-value services. Low-value services may differ in their perceived value to an affected household depending on the ways in which health events induce behavior changes. For example, if a chronic diagnosis communicates new risk information to a household, they may find low-value screening services—such as imaging services and preoperative visits—to be more attractive. On the other hand, households that respond to the price effects induced by a major health event may be more likely to seek out high-cost, low-return services such as elective surgeries. To explore these differences in-depth, I separate my sample of low-value services into five categories: pediatric services, including imaging services and the early use of medications such as antibiotics; adult prescription drugs, such as the

[16]These health services are based on recommendations made with the Choosing Wisely initiative, directed by the American Board of Internal Medicine Foundation and other physician specialty organizations (Bhatia et al., 2015; Wolfson et al., 2014). Appendix A.5 contains more detail about the specific services included in each measure.

use of opiates to treat migraines; unnecessary imaging services for adults, including for lower-back pain; extraneous screening services for adults, including cardiac testing before low-risk surgeries; and adult surgical procedures, such as arthroscopy for knee pain.

Table 1.4
Estimated Effects of Chronic Illness on Low-Value Care Utilization

Population	Pediatric	Adult Services			
Service Category	All Services	Prescriptions	Imaging	Screening	Surgery
$\text{Post}_t \times \text{Diagnosis}_f$	0.051*	-0.004	0.029***	0.103***	-0.096***
	(0.017)	(0.000)	(0.013)	(0.014)	(0.012)
R^2	0.349	0.309	0.293	0.326	0.379

Notes: Table shows estimated difference-in-difference regression coefficients for the effect of a new chronic diagnosis (N=1,538,161). Outcome variables are the inverse hyperbolic sine of billed spending in each category. See Appendix A for service definitions. Spending is measured in 2020 USD. Standard errors clustered at the household level are reported in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table 1.4 presents results estimating the effect of a new chronic diagnosis in each of the five categories using a standard difference-in-differences framework (event study regressions are included in Appendix B). New chronic diagnoses shift households spending and utilization into low-value service categories comprised of screening services, pediatric care, and imaging services. The effect sizes range from an increase as large as ten percent for low-value screenings to three percent for imaging services.[17] I find no effect on the misuse of prescription drugs among adults.

These results suggest that households seek out care that they see as useful in preventing or identifying future illness, even if those services are generally understood by health professionals as being low return. Although I observe households utilizing more of these services—such as preoperative screenings or imaging services—it is

[17]The results also provide preliminary evidence that major health events provide a deterrent from low-value elective surgeries. However, Appendix Table B.3 highlights the strong presence of pre-trends in these models, which obfuscates the true causal effect of the diagnosis.

unclear whether these are decisions made at the household level or by a physician who knows the family history and hence deems these services as appropriate. This provides new suggestive evidence that the utilization of low-value care may be tied more to risk beliefs rather than ignorance about the actual returns of a service. This is in keeping with recent work (Finkelstein et al., 2021).

In addition to the utilization of low-value care, I explore other ways health events alter the quality of consumers' health care decisions, including their plan choices (Appendix B). In general, I do not find that major health events prompt households to switch their health insurance plans. While new diagnoses in a household are associated with marked differences in observed spending behavior, it is still unclear whether these choices are *ex-post* more optimal for affected households. This motivates a more structural approach to quantify the welfare effects of health information.

1.4 Empirical Model of Belief Formation

In this section, I estimate the impact of health risk information on consumer choices as well as its implied welfare effects in a structural model of health utilization. I build on a canonical two-stage model of health spending (Cardon and Hendel, 2001). In the first stage, households choose an insurance plan to maximize their *ex-ante* expected utility, based on their available information about the distributions of future shocks. In the second stage, individuals within the household choose their spending and utilization based on realized health shocks and their chosen health plan's features.

I extend the existing model in two important ways. First, I allow consumers' types to be adaptive in response to health experiences. In my model, individuals learn about their probability of adverse health events; in addition, health events may alter household risk aversion to capture potential salience effects. Second, I explicitly model the differences between acute and chronic health shocks, as chronic health

shocks impose recurring costs on a family, thereby altering conditional OOP prices for non-chronic care and inducing moral hazard effects within a household.

1.4.1 Model primitives

Consider a household f comprised of individuals $i \in \mathcal{I}_f$. Individuals belong to one of two types—those without chronic illnesses and those with at least one chronic condition. I assume state-dependent preferences, so that the utility of receiving medical care differs across these types. Time is discrete and indexed by t; I am thus abstracting away from the timing of health spending within a year. Households and individuals are characterized by three main variables: individual beliefs about health risks (p_{ift}), household risk aversion (ψ_{ft}), and the distributions of their health shocks (discussed below). New health events—including both new chronic diagnoses and acute hospitalizations—cause all household members to update their beliefs about their health risks, as well as potentially altering household risk aversion and OOP prices.

In each period, two types of shocks are realized. Following typical convention, each individual has an acute health realization λ_{ift} drawn from an individual-specific distribution $F_{\lambda_{ift}}(\cdot)$. Acute health realizations model the uncertain aspect of demand for healthcare, with individuals with higher λ_{ift} being sicker and hence demanding greater healthcare consumption.[18] Second, households in each period receive a chronic health shock, m_{ft}^{CH}. This shock represents the disruptions in health *spending* affecting the household that arise from any new chronic diagnoses affecting an individual in the family; for households without a new diagnosis, this amounts to the expected cost of a new diagnosis. For households with pre-existing chronic conditions, these

[18]Rather than simply having families draw their health expenditure m_i following a plan choice (Handel, 2013; Layton, 2017), I explicitly model these health shocks in order to separately identify how spending choices are reflective of beliefs about major health events, as well as to estimate the effects financial distortions caused by health events contribute to moral hazard in spending.

shocks are the health costs associated with maintaining health for those affected by the conditions.[19]

1.4.2 Model stages

Families make two choices during each period. First, families choose their insurance coverage; then, acute and chronic health shocks are realized; finally, individuals choose their yearly health spending. These choices are static, in the sense that both households choose plans and individuals make spending decisions on the basis of the current period's utility and type parameters only (including their beliefs about health risks). The model is static, in the sense that household decisions in period t do not affect outcomes in period $t+1$. I can therefore ignore forward-looking behavior.[20] However, individual and household type parameters—including beliefs and risk aversion—are responsive to exogenous shocks, including major health events. These parameters adjust at the end of each model period, following individual utilization choices. I model the evolutions of these parameters using a Bayesian framework.

In the following sections, I outline the stages of the model in reverse—that is, I first present details of the individual spending choices in Section 1.4.2, followed by a discussion of household plan choices in Section 1.4.2. I then discuss how type parameters respond to exogenous health shocks in Section 1.4.2.

Utilization choice

After choosing a health plan $j \in \mathcal{J}$ and realizing the vector of acute and chronic health shocks $(\vec{\lambda}_{ift}, m_{ft}^{\text{CH}})$, each individual in the household chooses their medical

[19]Many related models incorporate heterogeneity in individual demand elasticities in order to accommodate heterogeneity in moral hazard effects (Einav et al., 2013; Marone and Sabety, 2022). As my model is concerned with disentangling only moral hazard events induced by major health events, I restrict the demand elasticity parameter ω in my model to be homogeneous across individuals and periods.

[20]Households are, however, forward-looking within a period, as they anticipate second-stage outcomes as part of their first-stage choices. See equation 1.9.

spending on non-chronic medical care, m^*_{ift}. In this stage, individuals make decisions independently to maximize their personal welfare; in the first (plan choice) stage, households make a collective decision. Given the flexibility in health states, which vary across individuals, households will ultimately distribute health spending so that the least healthy members receive the most care, as would be expected. Hence, this assumption makes the model more tractable without imposing restrictions on household behavior.

As is typical for these models, individuals trade off health production and wealth. In my extension of the model, individuals face residual uncertainty as to the likelihood of their own major medical events, which they believe occur with probability p_{ift}.[21] Individuals then choose m_{ift} in order to maximize their expected utility over states:

$$m^*_{ift} \equiv \text{argmax}_{m_{ift}} EU(m_{ift}; p_{ift}) = p_{ift} u_{ift,\text{C}} + (1 - p_{ift}) u_{ift,\text{H}}, \tag{1.3}$$

where $u_{it,\text{C}}$ and $u_{it,\text{H}}$ represent individual utilities when diagnosed with a chronic illness and when not diagnosed, respectively. Note that Equation 1.3 nests the case where an individual has already been diagnosed with a chronic illness, in which case $p_{ift} = 1$. I assume that each individual's utility function is separable in health and wealth for both chronic and healthy individuals:

$$u_{ift,\text{H}}(m_{ift}; \lambda_{ift}, m^{\text{CH}}_{ft}) = h_1(m_{ift}; \lambda_{ift}, m^{\text{CH}}_{ft}) + y_{ift}(m_{ift}; m^{\text{CH}}_{ft}) + \varepsilon_1 \tag{1.4}$$

$$u_{ift,\text{C}}(m_{ift}; \lambda_{ift}, m^{\text{CH}}_{ft}) = h_2(m_{ift}; \lambda_{ift}, m^{\text{CH}}_{ft}) + g(m^{\text{CH}}_{ft}; \lambda_{ift}) + y_{ift}(m_{ift}; m^{\text{CH}}_{ft}) + \varepsilon_2. \tag{1.5}$$

[21] Although the value of chronic care costs are assumed to be made known to a household before they choose their non-chronic spending, the model abstracts away from the specific timing of individual costs within a year. Hence, even within a period, individuals have not learned whether they have a chronic illness, and hence maximize an expected utility across both states of the world. It is not until the end of the period that individuals know their true state and update their beliefs p_{ift}.

The returns to medical spending $h_1(\cdot), h_2(\cdot)$, and $g(\cdot)$ are assumed to be concave, so that within-year health fluctuations λ_{ift} alter the optimal level of utilization m^*_{ift}. Remaining annual income is denoted by $y_{ift}(m_{ift}; m^{\text{CH}}_{ft})$. $\varepsilon_1(\cdot)$ and $\varepsilon_2(\cdot)$ are preference shocks to capture unobserved changes in preferences due to major medical events.

I parameterize these utility functions as quadratic loss functions in the difference between medical spending and acute health status, in keeping with past work, but allow for a potentially state-dependent utility function in which health status potentially alters the marginal utility of medical spending.[22] Individuals without chronic conditions face the typical utility function:

$$u_{ift,\text{H}}(m_{ift}; \lambda_{ift}, m^{\text{CH}}_{ft}, j) = (m_{ift} - \lambda_{ift}) - \frac{1}{2\omega}(m_{ift} - \lambda_{ift})^2 - c_j(m_{ift}). \tag{1.6}$$

Here, $c_j(m_{ift})$ represents the OOP costs associated with spending m_{ift}, conditional on the choice of plan j. Hence, individuals choose medical spending to approximately match their acute health realization λ_{ift}, accommodating the associated OOP costs of that spending.

On the other hand, individuals in the state of chronic illness face a utility function that depends on both acute and chronic health shocks, with potentially differing preference parameters. Their utility, which depends on the same model primitives as Equation 1.6, is given by:

$$u_{ift,\text{C}} = (\alpha_1 m_{ift} + \alpha_2 m^{\text{CH}}_{ft} - \lambda_{ift}) - \frac{1}{2\omega}(\alpha_1 m_{ift} + \alpha_2 m^{\text{CH}}_{ft} - \lambda_{ift})^2 - c_j(m_{ift}). \tag{1.7}$$

In this state, utility is derived from both chronic and non-chronic medical spending, each of which is potentially valued at a different rate than non-chronic medical spending for healthy individuals as indicated by the parameters (α_1, α_2).

[22] Previous work discuss and provide evidence for state-dependence in the utility of *non-medical* consumption (Finkelstein et al., 2013, 2009); this model introduces suggestive evidence for the state-dependence of non-chronic medical consumption as well.

Solving the expected-utility maximization problem is straightforward; however, as the marginal OOP cost changes based on where it is evaluated, the solution depends on which "region" of OOP costs an individual finds themselves in conditional on their health shocks (see Appendix C for details). If the realized acute health shock is negative (or sufficiently small relative to the shift parameter), individuals will choose $m^*_{ift} = 0$ as spending is required to be non-negative; otherwise, optimal spending follows the condition:

$$m^*_{ift} = \frac{1}{1 + p_{ift}(\alpha_1 - 1)} \left(\lambda_{ift} + \omega(1 + p_{ift}(\alpha_1 - 1) - c'_j(m_{ift}; m^{\text{CH}}_{ft})) - p_{ift}\alpha_2 m^{\text{CH}}_{ft} \right). \tag{1.8}$$

The interpretation of Equation 1.8 elucidates the key insights associated with this state-dependent utility framework with separate chronic care costs. In this expansion of the model, individuals choose to consume less non-chronic health care as chronic care costs increase in value, either by increases in magnitude, marginal utility, or likelihood. As discussed in Bleichrodt and Eeckhoudt (2006), the extent to which households mismeasure p_{ift} may artificially alter optimal spending decisions based on both the level of actual risks and the extent of the measurement error. Under the assumptions that households begin with p_{i0} close to zero, major health events could be associated with large (relative) increases in p_{ift}, potentially explaining the dramatic and persistent shifts observed in Section 1.3.

Equation 1.8 also highlights the ways that chronic care costs affect spending decisions through prices. The OOP cost function $c_j(m_{ift}; m^{\text{CH}}_{ft})$ is assumed to account for the price of chronic care first in the timing of health spending, before any other non-chronic spending. This anticipation of chronic care costs shifts the boundaries between optimal spending solutions by depressing the rate at which discretionary medical spending translates into OOP costs. This is the method by which moral hazard effects arise from major health events.

Plan choice

In the first stage of the model, households choose an insurance plan to maximize their *ex-ante* expected utilities without knowing their realization of individual health shocks λ_{ift} or major health costs m_{ft}^{CH}. This expected utility depends on the distributions of both health shocks as well as a household risk aversion parameter, which depends flexibly on household demographics and is allowed to evolve over time to capture the salience effects associated with health events, as discussed in Section 1.4.3. The household expected utility function for a given plan j is therefore:

$$U_{fjt} = -\sum_{i \in \mathcal{I}_f} \left[\int \int \frac{1}{\psi_{ft}(x_{ft})} \exp\{-\psi_{ft}(x_{ft}) u_{ift}^*\} \; dF_{\lambda_i} dG_{m^{\text{CH}}} \right] - c_j(m_{ft}^{\text{CH}}) - \pi_{fj} - \eta \mathbb{1}_{fj,t-1}, \tag{1.9}$$

where u_{ift}^* represents the optimal payoff to individual i in period t given the realization of acute and chronic health states.[23] In addition to each individual's realized OOP costs for non-chronic medical spending, households face OOP costs for chronic care represented by $c_j(m_{ft}^{\text{CH}})$. Households also face plan premiums π_j and a perceived monetary cost η for switching plans ($\mathbb{1}_{fj,t-1}$ is an indicator for whether the family chose plan j in year $t-1$).[24]

[23] One concern with a utilitarian index here is that households may have little incentive to diversify their medical spending across household members. However, the choice of the utility function used in the second (spending) stage of the model makes it optimal for families to allocate care according to each individual's realization of λ_{ift}; hence, this modeling choice does not give rise to families allocating all of their care to a single individual. An alternative approach is to use a CES function for utilities; however, this introduces more nuisance parameters into the estimation framework. Finally, I assume that the von Neumann Morgenstern (vNM) utility index for this decision possesses a constant coefficient of absolute risk aversion, a common choice for these models as it implies no wealth effects.

[24] I do not observe premiums or contributions in my data and therefore follow the methodology of Layton (2017). In particular, I assume that premiums are equal to the average cost among the employees with dependents enrolled in the plan during the prior year plus a fixed overhead cost, and then assume that employee contributions are 28% of that value (KFF, 2020). Note that as Layton discusses, identification of the structural parameters in this model do not depend on accurate estimation of premiums, but rather require that the premium differential across firms is correct.

Parameter updating

After households and individuals have made their plan and spending choices, type parameters evolve in response to health events. Of particular interest is the way that individuals update their beliefs about their unknown transition probability (p_{ift}). Additionally, households update their risk aversion parameters (ψ_{ft}) according to an adaptive framework; I discuss this further in Section 1.4.3.

I model individual learning about health risks as a Bayesian updating process in response to health events. In particular, I assume that initial beliefs depend on individual demographics, including age, sex, health risk scores, and the presence of any pre-existing conditions within the household. Prior beliefs are based on a signal x_{if0}, which is assumed to be normally distributed with mean and variance parameters $(\mu_{pi0}, \sigma^2_{pi0})$; this signal is mapped into a probability $p_{if0} \in [0, 1]$ using the standard logistic function. The center of the distribution μ_{pi0} varies with individual demographics and is potentially correlated with other household type parameters.

Major health events provide individuals with signals y_{ift} about the underlying distribution of p_{ift}, I likewise assume that these signals are normally distributed, so that the mean and variance of an individual's posterior distribution has a closed-form solution in each period. Specifically, if $y_{ift} \sim \mathcal{N}(\tilde{\mu}_{ift}, \tilde{\sigma}^2_{ift})$, the evolution of the mean and variance parameters can be written as:

$$\sigma^2_{pi,t+1} = \frac{\tilde{\sigma}^2_{ift}\sigma^2_{pi0}}{\tilde{\sigma}^2_{ift} + s_{ift}\sigma^2_{pi0}} \tag{1.10}$$

$$\mu_{pi,t+1} = \frac{\tilde{\sigma}^2_{ift}\mu_{pit} + \sigma^2_{pit}\tilde{\mu}_{ift}}{\tilde{\sigma}^2_{ift} + \sigma^2_{pit}}, \tag{1.11}$$

where the variable s_{ift} indicates how many health signals an individual has received by the end of period t.

An important potential difficulty when using a Bayesian framework with rare events is the choice of updating frequency. Given the relative rarity with which chronic health shocks occur, updating of probabilities after each period would result in posterior beliefs that are tightly centered around the initial mean, varying little with new information. In such a regime, individuals would have to perceive health shocks as being impossibly likely (e.g., $\tilde{\mu}_{ift}$ much greater than 1) in order for health shocks to meaningfully change health beliefs. This is inconsistent with the analysis I have presented previously, which shows that individuals are highly responsive to chronic health shocks.[25]

I address this inconsistency in my preferred specification by assuming that households update their beliefs *conditional* on a health event occurring. This reduces the number of uninformative signals individuals process, and hence avoids problems of weight degeneracy, and is consistent with individuals who form beliefs about their health risk once, and then only revisit those beliefs once they have been called into question. Once the individual begins evaluating their health risk beliefs (e.g., after a diagnosis has occurred within the household), they do so in a completely standard way, including updating beliefs in all following years without major health events.

Such an approach is an intuitively appealing way to deal with the issue of Bayesian updating when signals are infrequent. However, my results are robust to alternative specifications, including (i) an adaptive learning framework where individual beliefs are specified as an AR(1) with some dependence $\rho < 1$ on the previous period's beliefs, and (ii) a more traditional setup where individuals update their beliefs with some

[25]In addition to the analysis presented here, I also find that older individuals have stronger responses to chronic health events in their household than younger individuals, even after conditioning for risk score (not shown). If individuals behaved as though they updated their health beliefs in each period—regardless of if a signal or health event occurred—then older individuals would have belief distributions more tightly centered around their mean, hence their posterior distributions following a realized health signal would shift *less* than younger individuals with more flexible priors. I do not observe this to be the case.

probability $p > 0$ in the absence of health events.[26] Additional modeling possibilities include the use of quasi-Bayesian modeling where individuals disregard less salient signals, but still update beliefs in each period (Rabin, 2013), or where individuals over-weight "good news" relative to "bad news" (Eil and Rao, 2011).

1.4.3 Estimation

Parametrization

The unit of observation is a family f comprised of a set of individuals $\mathcal{I}_f$ in year t. Each family faces a choice of plans that varies at the firm-year-state level.[27] Households are characterized by their unobserved type variables $\{p_{ift}, \lambda_{ift}, \psi_{ft}\}_{i \in \mathcal{I}_f}$. I allow the initial parameters $(p_{if0}, \lambda_i ft, \psi_{f0})$ to be arbitrarily correlated, and link them to observable data by assuming that they are drawn from a multivariate normal distribution which depends on observed demographics:

$$\begin{bmatrix} p_{if0} \\ \mu_{\lambda if} \\ \log(\psi_{f0}) \end{bmatrix} \sim \mathcal{N}\left(\begin{bmatrix} \beta_p \boldsymbol{X}^p \\ \beta_\lambda \boldsymbol{X}^\lambda \\ \beta_\psi \boldsymbol{X}^\psi \end{bmatrix}, \begin{bmatrix} \sigma_p^2 & & \\ \sigma_{p,\lambda} & \sigma_\mu^2 & \\ \sigma_{p,\psi} & \sigma_{\lambda,\psi} & \sigma_\psi^2 \end{bmatrix} \right). \tag{1.12}$$

Covariates $\boldsymbol{X}$ include age, sex, health risk score, family size, and the presence of pre-existing conditions in a household. In practice, I use individuals' first year of data in $\boldsymbol{X}^p$ and $\boldsymbol{X}^\lambda$ and within-individual averages in $\boldsymbol{X}^\psi$.

Individual beliefs evolve in response to signals about their health risks as discussed in section 1.4.2. I assume that these signals y_{ift} are normally distributed with variance σ_π^2 (to be estimated) and a mean given by the logit regression:

$$y_{ift} = \pi_1 \mathbb{1}\{\text{chronic}\}_{f,-i} + \pi_2 \mathbb{1}\{\text{acute}\}_{f,-i} + \pi_3 \mathbb{1}\{\text{acute}\}_{f,i} + \pi_4 x_{ift}, \tag{1.13}$$

[26] For a more in-depth review of the relative strengths and weaknesses of Bayesian or adaptive learning in structural modeling, see Aguirregabiria and Jeon (2020).

[27] I ignore plans that have less than five percent of the overall firm-year market share in my data to avoid including executive health plans in employee choice sets.

where *chronic* and *acute* indicate the occurrence of chronic or acute health events within a household and x_{ift} is a variable for the number of years that have passed since the earliest major health event in the family. Hence, π_1 is the main parameter of interest, identifying the effect of a household chronic diagnosis on individual beliefs. On the other hand, the variance of the signals, σ^2_π, reveals the magnitude of unobserved information affecting individual health risk probabilities.

To parameterize the distribution of acute health shocks, I assume that $F_\lambda(\cdot)$ is a shifted lognormal distribution. This is a natural parameterization as the distribution of annual health expenditures is highly skewed (Mitchell, 2020). The choice of shifting the distribution accommodates the approximately 15% of individuals in my sample who choose zero medical spending in a given year. I therefore model an individual's (correct) beliefs about their transient health shocks by

$$\ln(\lambda_{ift} - \kappa_{if}) \sim \mathcal{N}(\mu_{\lambda,if}, \sigma^2_{\lambda,if}). \tag{1.14}$$

When κ_{if} is sufficiently large (and negative), small and negative values of λ_{ift} may lead to zero spending being the utility-maximizing solution for an individual.[28]

Acute health shocks at the individual level are therefore summarized by three parameters: $(\mu_{\lambda if}, \sigma^2_{\lambda if}, \kappa_{if})$. The parameter $\sigma^2_{\lambda if}$ reflects the precision in an individual's beliefs about their transient health state. Both $\sigma^2_{\lambda if}$ and κ_{if} are estimated as a linear projection on individual covariates (see Appendix C).

In contrast, I directly use empirical distributions of chronic care costs from my data in household expected utility. I assume that individuals have rational expectations over the distributions of their chronic health care costs, which change when

[28]Previous work has allowed the distributions of these shocks to evolve over time. In my model, which separates acute and chronic health shocks, such variation would amount to shifts in the need for non-chronic health spending, such as variation in an individual's anticipated office-visit spending from year to year. In addition to being of second-order concern to my setting, such variation seems indistinguishable from the random variation in the draws of λ_{ift} already present.

they experience major health events. This is a simplifying assumption employed for tractability, as my model already allows for the identification of rich heterogeneity governing individual expectations about health shocks. However, although there is evidence that consumers do not fully know the price of health care before selecting services (Lieber, 2017), this is less concerning with chronic care costs, which are typically stable over time and hence more easily predicted by household members. The empirical distributions are similarly assumed to be stable across years, but I use a separate distribution in the year of diagnosis to accommodate potentially higher costs in that year (e.g., for unexpected hospitalizations).

Finally, I allow family risk aversion ψ_{ft} to evolve over time as discussed above. In particular, $\psi_{ft}(x_t)$ evolves linearly according to:

$$\psi_{ft} = \gamma_0 \psi_{f,t-1} + \gamma_1 \left\{ \text{Post}_t \times m_{f0}^{\text{CH}} \right\} + \gamma_2 \left\{ \text{Post}_t \times c_j(m_{f0}^{\text{CH}}) \right\} + \gamma_3 \left\{ \text{Post}_t \times \text{Hosp}_{f0} \right\} + \zeta_{ft}, \tag{1.15}$$

where m_{f0}^{CH} represents the billed spending associated with the diagnostic event, $c_j(m_{f0}^{\text{CH}})$ the OOP spending of the diagnostic event, and Hosp_{f0} indicates whether a hospitalization occurred as part of the diagnosis. I assume that $\zeta_{ft} \sim \mathcal{N}(0, \sigma_\psi^2)$.

I denote the parameters of the model by θ. These parameters include the main parameters of interest $\vec{\pi}$ and $\vec{\psi}$, including the variances σ_π^2 and σ_ψ^2. Additional parameters included in the estimation are the utility parameters $\alpha_1, \alpha_2, \omega$, and η; the five vectors of mean shifters $(\beta_p, \beta_\psi, \beta_\lambda, \beta_{\sigma_\lambda}, \beta_\kappa)$; seven variance and covariance parameters $(\sigma_p, \sigma_\mu, \sigma_\psi, \sigma_\kappa, \sigma_{p,\psi}, \sigma_{p,\mu}, \sigma_{\psi,\mu})$; and the variance of the idiosyncratic shock term σ_ε^2, which scales the choice probabilities. I assume that these idiosyncratic shocks follow the typical Type-1 Extreme Value distribution. Based on θ and the data, I am able simulate values for $p_{ift}, \mu_{\lambda if}, \sigma_{\lambda if}, \lambda_{ift}$, and ψ_{ft}.

I estimate the model via maximum likelihood with the appropriate adaptation for modeling a discrete choice followed by a continuous one (Dubin and McFadden, 1984;

Revelt and Train, 1998; Train, 2009). For a given household, likelihood functions are constructed as the density of their observed health spending conditional on their observed plan choices. I provide additional estimation details in Appendix D.

Identification and interpretation

My model utilizes multiple sources of variation to separate multiple effects arising from major medical events. In addition to any changes in individual risk beliefs, health events may alter health behaviors by changing the price of non-chronic care, increasing the salience of health consumption, providing experiential learning about how to obtain high-quality health care, or altering preferences for medical care in other ways. The critical challenge is that changes in risk preferences, salience, or systematic health learning may also increase the willingness to purchase insurance and utilize medical care.

I use a rich set of major health events that vary in their expected costs, both in the year of diagnosis and in following years. This variation in the expected costs needed to maintain health for someone with a chronic condition changes the extent to which a specific chronic condition significantly alters the expected prices for other, non-chronic medical care. This variation, coupled with variation in plan spending characteristics, allows me to separate moral hazard effects from other drivers of behavior.

To separate risk aversion from beliefs, I use variation in insurance plan characteristics and choice sets faced by different households in my data set. These choice sets vary at the firm-state-year level, and typically include plans with a wide range of cost-sharing parameters (Table 1.2). Under the assumption that risk aversion drives plan choice and not medical spending, and that households with high risk aversion seek to reduce the incidence of high OOP expenditures, highly risk-averse households will gravitate towards the plans in their choice sets that most limit high expenses (e.g., low-deductible plans). Finally, I use data on the circumstances of major medi-

cal events—including the resulting costs and whether a hospitalization occurred—to incorporate the role of salience associated with health trauma in changing household risk aversion.

The principal estimated structural parameters of interest in my model are those governing the evolution of the transition probabilities p_{ift}. Changes in these parameters that arise from new chronic diagnoses encompass both a reevaluation of individual health risk beliefs and other informational effects unaccounted for in the model, which may load onto this parameter. These effects include learning about the health care system more generally or forging better relationships with health care providers. Although section 1.3 suggests that these factors are not the principal mechanisms for responses, they may influence how p_{ift} responds to new diagnoses. I therefore interpret changes in p_{ift} as resulting from an aggregate informational effect, rather than from moral hazard or salience effects.[29]

1.5 Structural Results

Table 1.5 presents the estimated parameters resulting from maximum likelihood estimation. Column 3 shows the preferred specification described in Section 1.4, while columns 1 and 2 present simplifications of the model that are useful both in building intuition and validating the estimated parameters. Additional parameters not relevant to the welfare effects of health information—including incidental parameters such as switching costs and individual mean-shifting regression coefficients—can be found in Appendix Table D.1.

I consistently find strong effects on non-diagnosed beliefs associated with household chronic diagnoses. New chronic diagnoses are associated with an average increase in an individual's belief of a major health event of 33 percentage points, an effect

[29] Appendix C discusses an alternative interpretation of p_{ift} as a preference weighting across states rather than explicitly health beliefs.

Table 1.5
Estimated Structural Parameters of Interest

		Model 1		Model 2		Model 3	
		Estimate	Std. Err.	Estimate	Std. Err.	Estimate	Std. Err.
Panel A: Dynamic Parameters							
Belief Evolution							
π_1	Family Chronic Event	0.69	(0.002)	0.17	(0.002)	0.33	(0.002)
π_2	Own Acute Event	0.07	(0.002)	0.02	(0.001)	0.05	(0.002)
π_3	Family Acute Event	0.09	(0.002)	0.03	(0.001)	0.06	(0.002)
π_4	Years since Event	-0.01	(0.000)	0.002	(0.000)	0.01	(0.000)
σ_π	Error Variance	10.29	(0.000)	0.12	(0.005)	1.52	(0.018)
Risk Aversion Evolution							
ψ_0	Persistence, Year $t-1$	–	–	–	–	0.95	(0.025)
ψ_1	Health Event (HE)	–	–	–	–	0.61	(0.015)
ψ_2	HE × Year 0 Cost	–	–	–	–	0.19	(0.020)
ψ_3	HE × Year 0 OOP	–	–	–	–	-0.88	(0.024)
ψ_4	HE × Hospitalization	–	–	–	–	1.51	(0.033)
σ_ψ	Error Variance	–	–	–	–	0.01	(0.016)
Panel B: Heterogeneity in Types							
σ^2_ξ	Idiosyncratic Shock	5.92	(1.006)	6.24	(0.109)	3.56	(0.085)
σ^2_p	Initial Beliefs	16.59	(0.410)	24.43	(0.003)	14.51	(0.001)
σ^2_ψ	Initial Risk Aversion	15.22	(0.289)	5.55	(0.005)	2.57	(0.005)
σ^2_λ	Acute Shocks	–	–	0.58	(0.004)	2.03	(0.001)
$\rho_{p,\psi}$		-0.87	(0.360)	-0.43	(0.002)	-0.54	(0.002)
$\rho_{p,\lambda}$		–	–	-0.91	(0.006)	0.38	(0.002)
$\rho_{\psi,\lambda}$		–	–	0.12	(0.002)	0.09	(0.002)
Beliefs Evolve		Yes		Yes		Yes	
Acute Shock Heterogeneity				Yes		Yes	
Risk Aversion Evolves						Yes	

Notes: This table presents estimates for selected parameters of the structural model of health choice; Appendix Table D.1 presents estimates for the remaining parameters. Belief evolution parameters $\vec{\pi}$ are reported as marginal effects. Standard errors are derived from the analytical Hessian of the likelihood function. Column 3 presents my primary estimates used in later calculations. All models are estimated on an unbalanced panel of 179,044 households over eight years. Preference coefficients are relative to thousands of dollars.

which is far larger than those estimated for acute events for either the individual or their family members, which are estimated to only increase risk beliefs by five and six percentage points, respectively. These increases are persistent, with little evidence that risk beliefs decrease over time (the estimated time trend coefficient is only one percentage point each year). The estimated variance for the unobserved dimension of belief changes is low, indicating that unobserved events are not contributing to large changes in risk assessments.

Table 1.5 also presents parameters illustrating how the effects of new chronic illnesses alter behaviors in other meaningful ways. Major health events—both acute and chronic—are associated with strong salience effects that increase household risk aversion. On average, experiencing a major health event increases the coefficient of household risk aversion by 0.61, a 34.9% increase over the pre-diagnosis average coefficient of 1.75.[30] These effects are stronger when the household event entails either a higher amount of total billed spending or a hospitalization, suggesting that households respond differently to the intensity of an event.

Panel B reports additional information regarding the distribution of household types and the value of incorporating the full richness of the model in rationalizing observed plan choices and spending. In particular, I estimate a high degree of variance in individual health risk beliefs (prior to any health event). These beliefs are weakly positively correlated with acute health status and negatively correlated with household risk aversion. These facts suggest that variation in individuals' estimated beliefs reflects variation in individual health status, as expected. Finally, in the full

[30]To put these numbers into context, I follow the results of Cohen and Einav (2007) and consider the amount $X that would make the average household in my sample indifferent between a sure payoff of $0 and an equal-odds gamble between winning $100 and losing $X. Prior to a diagnosis, the average value of $X is roughly $85.08; after diagnosis, this value changes to $80.85. These results are comparable with previous estimates of household risk aversion for health insurance (Einav et al., 2013; Marone and Sabety, 2022)—however, as mentioned in Einav et al. (2013), the coefficients from models incorporating both health and financial risk do not compare to those of models with pure financial risk (Cohen and Einav, 2007; Handel, 2013).

version of the model, the variance of the idiosyncratic error term is small, suggesting that most of the observed variation in consumer behavior can be explained by heterogeneity in individual types, responses to major health events, or both.

Models 1 and 2 of Table 1.5 illustrate simplifications of the model that help validate the estimated parameters and build intuition. In Model 1, I estimate a version of the model with no heterogeneity in acute health shocks or changes in household risk aversion. That is, $\mu_{\lambda,i}, \sigma_{\lambda,i}$, and $\kappa_{\lambda,i}$ are not allowed to vary based on individual covariates, and ψ_{ft} is fixed over time. A key difference between Model 1 and my preferred specification is that the estimated impact of chronic health shocks on risk belief distributions is much higher when I do not accommodate heterogeneity either in period-level health shocks or salience effects. This result is intuitive, as the absence of this heterogeneity leads to the inaccurate "loading" of belief changes onto specific events.[31] This loading is observed on a comparable scale for coefficients for acute major health events as well; however, note that these effects are associated with higher overall variance in belief evolution, presumably because the simplified model attempts to explain multiple sources of variation through a single channel.

Column 2 adds variation in acute health status to the model while continuing to hold household risk aversion constant over time. Accounting for this heterogeneity explains a substantial portion of the belief evolution pattern suggested by the most simplified model, decreasing the size of the effect of all major health events by about two-thirds and the variance of unobserved belief shocks (σ_π) even more drastically. Similarly, including acute health shocks in each period reduces the estimated variation in initial coefficients of risk aversion and the correlation between risk aversion and beliefs, suggesting that including that accounting for variation across health states is important in estimating both health learning and salience effects. A key difference

[31] This is exacerbated by the fact that acute health states and chronic diagnoses are correlated, as presented in Panel B of Table 1.5.

between column 2 and column 3 is that after incorporating the explicit modeling of salience effects, the estimated effect of major health events on belief changes is almost double. Notice that there is a strong negative correlation between household beliefs and risk aversion; this means that when estimated together, salience effects may have muted the estimated effect of belief changes. Hence, it is to be expected that separating salience effects from belief changes increases the estimated effect of events on beliefs.

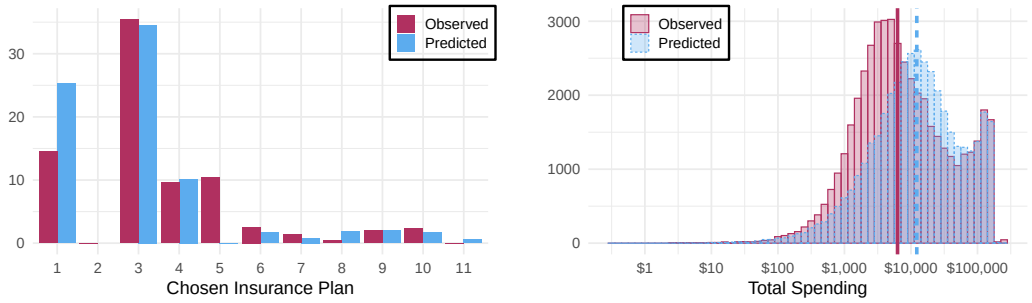
1.5.1 Model fit

I evaluate the fit of my estimated model at both the plan choice and spending stages. To evaluate plan choices, I compare plan choices for households observed in the data with those predicted by the model in Figure 1·6. Predicted choice probabilities are influenced by premiums, inertia, and household expectations of their acute and chronic health shocks, valued based on their level of risk aversion. At the level of household spending, I compare observed household spending distributions to those predicted by the model. As spending decisions are made after the realization of two random variables (acute and chronic health shocks), I base the model predictions off of a single draw of these underlying variables. I pool all individuals within a firm across years.

Figure 1·6 presents the results. The first panel shows the observed and predicted market shares for enrollment in plans offered in the largest firm in my sample. Overall, predicted shares are closely matched. The panel on the right presents observed and estimated spending conditional on a plan choice. Here, the model predicts slightly higher levels of billed spending than are typically observed, with a difference of about $1,000 between the means of the two distributions. The model appropriately predicts the extensive margin of spending, appropriately capturing the fraction of individuals who choose zero medical spending in a given year.

Figure 1·6
Predicted and Observed Insurance Plan and Health Care Spending Choices

(a) Insurance Plan Choices **(b)** Health Spending Choices



Notes: Figures show overall match between estimated model predictions and observed household choices, at both the plan choice (left) and spending (right) stages of the model. In the first panel, market shares for each insurance plan offered to employees of the single largest firm are shown (see Appendix D for other firms). All years are pooled, so each observation is a household-year. The overall match rate is 82.2%. The second panel plots distributions of predicted and observed household health care spending, conditional on predicted/observed spending greater than zero (the observed rate of zero spending is 16.6% and the predicted rate is 13.2%). All years are pooled, so an observation is a household-year. Vertical lines represent the mean of the respective distribution.

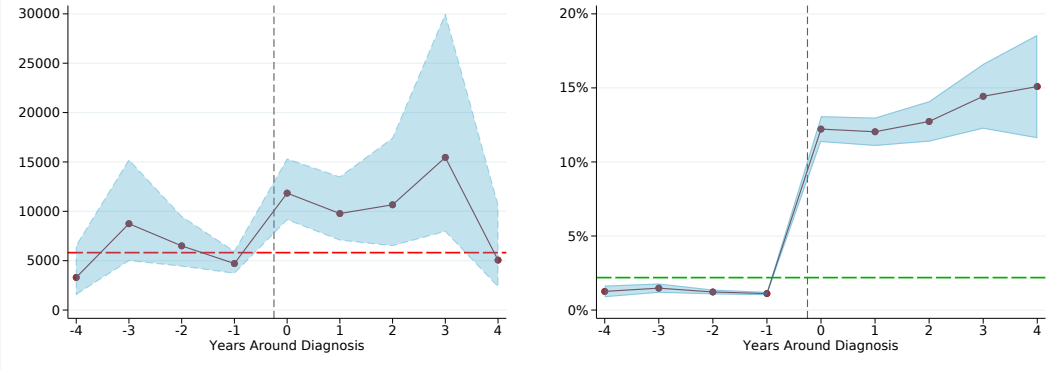
1.5.2 Spending response to major health events

Figure 1·7 illustrates the model's predictions surrounding behavior following new chronic diagnoses in a household as recentered time series graphs. Similar to the results in Section 1.3, I examine how these diagnoses alter the spending patterns of other household members in the panel (a). I also present estimates for how diagnoses affect estimates for individuals' underlying transition probabilities p_{it} in panel (b). In my model, household diagnoses are associated large increases in OOP spending (about 20%, a difference which is statistically indistinguishable from the 10% reported earlier).

Importantly, I predict large accompanying changes in individual health risk beliefs following a new chronic diagnosis in the family. The horizontal green line in the Panel (b) of Figure 1·7 depicts the pooled average risk of diagnosis within my sample, which is roughly 2.5%. Prior to health events, individuals tend to under-

Figure 1·7
Model Predictions: Non-Diagnosed Spending and Beliefs Around a New Diagnosis

(a) Effect on Non-Diagnosed OOP Spending **(b)** Effect on Health Risk Beliefs



Notes: Figures show recentered time series for model predictions of spending and beliefs for non-diagnosed household members who have experienced a diagnosis with a new chronic illness in the household. The first panel illustrates percentage changes in the inverse hyperbolic sine of OOP spending, measured in 2020 USD. The second panel illustrates estimated changes in predicted beliefs, averaged over draws from individual posterior distributions. The green horizontal line in Panel (b) illustrates the average in-sample rate of diagnosis with a new chronic condition, roughly 2.5%.

weight their health risks by about 58%; however, following a diagnosis, individuals move to *over-weighting* their risks by over *six* times the true in-sample rates of diagnosis. Instead, these households make choices as though they perceived their risk of a chronic diagnosis to be greater than one in ten. This provides suggestive evidence that individuals in affected households may over-respond to these events. I explore the welfare implications of these facts in the following section.

1.6 Welfare & Counterfactual Simulations

Based on the estimated model parameters, I am able to construct a measure of each household's willingness to pay for information associated with their own health risks. I use this measure to provide a benchmark for the value associated with this information, with particular focus on whether major health events meaningfully alter individual expected utility and social surplus.

1.6.1 Welfare effects of information

Households who receive health information alter their plan choice and medical spending decisions, thereby altering their *ex-ante* expected payoffs from care. My model allows me to estimate the spillover value of new health information for non-diagnosed household members by comparing these expected payoffs in the observed data—where household members use information to alter choices—and a counterfactual regime where the information is not revealed. In this counterfactual state, non-diagnosed household members experience the observed sequence of acute health shocks without any of the changes to p_{ift}, ψ_{ft}, or $c_j(\cdot)$ that would arise from a chronic event in the household.[32]

A household's willingness to pay for health information is equal to the difference in certainty equivalents across these two regimes. Certainty equivalents are given by

$$CE_{fjt} = -\psi_{ft}^{-1}\log(-U_{fjt}), \tag{1.16}$$

where U_{fjt} is the total *ex-ante* expected utility family f expects when enrolling in plan j at time t, as defined in equation 1.9. I assume that conditional on the estimated parameters, households are fully rational and enroll in the plan that gives the highest expected utility at the time of choice.[33] Throughout, I report differences between CE_{fjt} across the benchmark state of the world and regimes where information is partially or fully revealed; hence, reported values are "marginal" willingness to pay measures.

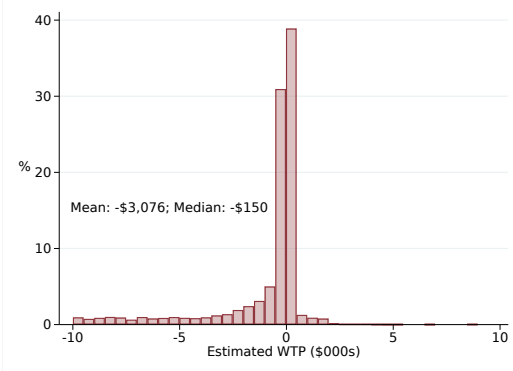
[32] I limit attention to non-diagnosed household members in order to estimate the *spillover* value of new health information, as well as to ignore the mechanical changes in household expected welfare that arise when a household member's health state is permanently altered, as with a new chronic diagnosis.

[33] The model allows for rich heterogeneity in the prediction of health states as well as rationalizations for common choice mistakes, including switching costs. Hence, such an assumption is reasonable. Similarly, I assume that the idiosyncratic shock parameter is not relevant for the context of estimating welfare gains from health information.

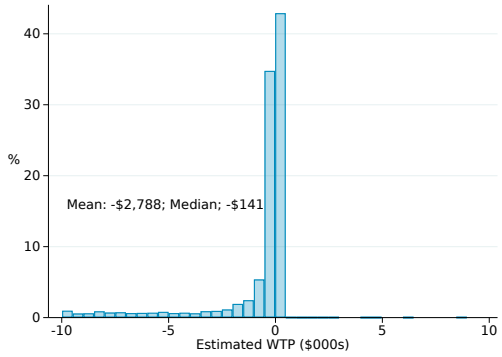
The utility-maximizing decision in my model is one where agents choose an appropriate level of spending relative to an uncertain multi-dimensional health shock; new health risk information changes the relative weight agents place on the dimensions of that shock when making their decisions. Hence, this welfare criterion measures how much households would be willing to pay for the information, based on their resulting changes in utilization choices during that period. My model does not allow me to measure the welfare effects of information in terms of long-term health production, for example from an increased investment in preventive health services. Such welfare effects are interesting particularly in conjunction with feasible health policies that jointly reveal information about health risk *and* the relative quality of health services. However, these returns would take more years to be realized than my sample permits me to analyze.

Figure 1·8
Variation in Welfare Effects Associated with Health Events and Health Information

(a) Full Welfare Effect



(b) Health Information Effect Only



Notes: Figures show estimated changes in household willingness to pay associated with major health events. The panel on the left shows differences in household certainty equivalents in the case of a full response to a new diagnosis, including adjustments to risk aversion and moral hazard effects; the panel on the right shows only differences arising from adjustments to household risk assessments. Welfare effects are calculated in the year of the diagnosis relative to a benchmark in which no information is transmitted.

Figure 1·8 depicts variation in household willingness to pay for health information in the year of the new chronic diagnosis.[34] Household members who are exposed to a new chronic diagnosis experience a welfare penalty that averages $3,076 per household per year. However, there is substantial heterogeneity in these effects, including 28% of treated families who have a higher resulting expected utility following the realization of health information.

The right panel of Figure 1·8 shows the distribution of welfare effects associated solely with receiving new health information. A novel feature of my structural model is the ability to separate changes to household welfare that arise from dimensions of a health event other than the realization of health information. I recalculate welfare changes associated with *only* changes to household beliefs by holding constant changes to both household risk aversion and any moral hazard effects that arise from changes to spot prices. My analysis reveals that these dimensions contribute little to overall changes in household welfare, with 90% of welfare changes being explicitly attributable to changes in household beliefs. The average household experiences a welfare penalty of $2,788 associated with changes to how they evaluate their risk of developing a chronic condition. This corresponds to an average decrease in welfare of about 11.6% (Appendix Figure D.1).

Although at first glance associating new information with a welfare penalty seems counter-intuitive, my results are consistent with a story of household over-responsiveness to information. The observed choice data which informed the estimated model parameters suggests that new chronic diagnosis spur large swings in household members' assessments of their health risks; however, these welfare calculations make clear that in many cases, households would be better off if they had acted as though they had

[34]These welfare effects are stable in the first few years following the diagnosis; hence, for ease of interpretation, I only focus on the year of diagnosis itself.

not received the information. This is precisely because of the magnitude of the shifts in household beliefs, as I illustrate in the following section.

Importantly, the returns to health information vary with key household characteristics, including household risk levels and estimated risk aversion (Appendix Figure D.2). Households who are less averse to negative outcomes prior to the diagnosis experience lower welfare penalties, on average, than those with higher risk aversion. Differences in this parameter are intuitively meaningful: households with greater risk aversion experience greater "translation" of new health information into changes in insurance plan choices and, subsequently, health spending. Hence, households with lower levels of risk aversion tend to respond less to new information, presumably contributing to the lower estimated welfare penalties associated with the event. Similarly, households with high expected health risks prior to a new diagnosis experience lower welfare penalties. This, too, is related to overall muted responses to health information. However, this low level of responsiveness is attributable not to low variation in expected utility but to an already high level of expected spending, meaning new health events change outcomes (in percentage terms) less.

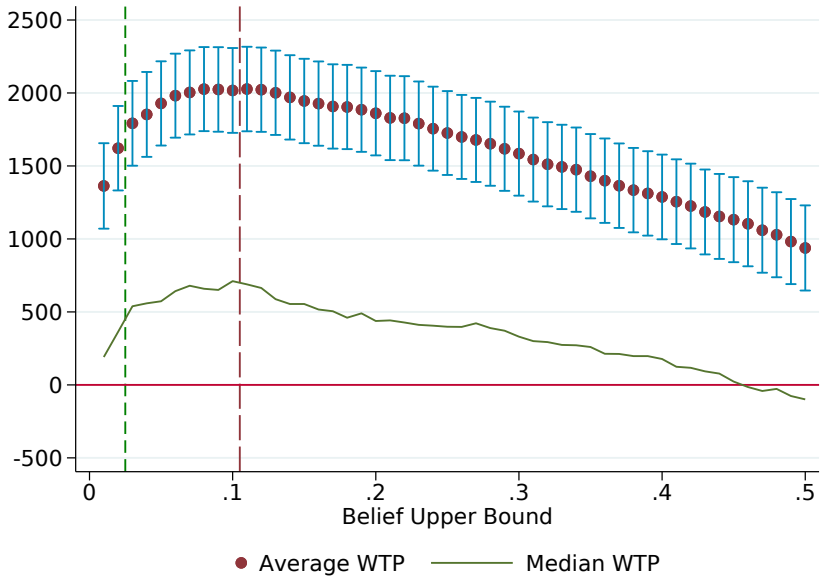
1.6.2 Evaluating household over-responsiveness to information

The results above imply that while households respond meaningfully to new health information, they may not be doing so in ways that are welfare improving. Given these estimated welfare penalties, in this section I assess the extent to which consumers' over-responsiveness to health information dampens potential welfare gains. The model predicts large swings in consumer beliefs when exposed to chronic diagnoses in a household. I therefore first assess the extent to which limiting the magnitude of these changes affects estimated welfare differences. I then turn to practical policy questions surrounding when information revelation is optimal, and whether targeted revelation can improve social outcomes.

Bounding belief updating

I first consider how limiting household responsiveness to adverse health events alters estimated welfare gains or penalties from new health risk information. Here, I present estimated effects from imposing arbitrary upper bounds on an individual's beliefs about their own health risks; that is, imposing that any predicted value p_{ift} in the model be no greater than some threshold $\overline{p}$. This exercise illustrates that if consumers' responses more closely matched their true expected risk (conditional on the household member's diagnosis), health information would be associated with welfare gains rather than losses.

Figure 1·9
WTP ($) for Health Information After Bounding Responsiveness



Notes: Figure depicts estimated household willingness to pay for new health information across multiple counterfactual scenarios in which post-event health beliefs are capped at $\overline{p}$. Each point represents a distinct scenario with $\overline{p}$ indicated along the x-axis. Average and median household WTP for new information are depicted as the maroon scatter plot (with 95% confidence intervals) and the smoothed blue-gray line, respectively. The vertical dashed green line represents the in-sample rate of diagnosis (about 2.5%), while the long-dashed maroon line represents the upper bound at which welfare is maximized (about 10%).

Figure 1·9 presents the results. The figure summarizes household WTP for information across multiple scenarios, each with a varying degree of restrictiveness on $\overline{p}$. Average and median welfare gains are plotted; notice that the distribution of welfare gains is skewed as suggested in Figure 1·8. As opposed to a scenario with no restrictions—where the median household's informational WTP was -$141—the median household would be willing to pay a positive amount for information whenever $\overline{p}$ is less than 45%. Welfare gains continue to improve as this bound becomes more restrictive until $\overline{p}$ is about 11% (shown in the Figure as the marroon long-dashed line). At this point, the average (median) household's welfare is estimated to be $2,027 ($711); in addition, about 86% of households receive welfare benefits from information, compared to 0.2% in the baseline scenario.

As the upper bound moves past this point, average household welfare gains begin to diminish. The belief upper bound which achieves an average WTP maximum is larger than the true in-sample risk of diagnosis (shown in the Figure as the green dashed line); this is because declines in consumer welfare following this point represent heterogeneous returns to new health information. Although the generic household in the model prefers, *ceteris paribus*, to have beliefs matching their true risk of chronic diagnosis onset (due to the state-dependence of preferences for non-chronic care), these risks vary across households. For some, these risks skew much higher than the average rate of illness onset, meaning that arbitrary bounds such as $\overline{p}$ risk harming households for whom information *does*, in fact, reveal large changes to beliefs.

To examine this further, I estimate individual-specific health risks $\hat{p}$ based on demographics including age, sex, and relationship with diagnosed household members. Although these predicted health risk probabilities do not capture the full range of private information, they address individual differences in potential responsiveness to new information. I estimate predicted health risk probabilities on a validation sample

constructed from all Marketscan households not in my main sample who experience at least one chronic event during their observed period. Additional details about this estimation and summary statistics for the resulting probabilities are provided in Appendix D. The predicted probabilities are small and match in-sample diagnostic risks.

When I impose these predicted probabilities as individual-specific upper bounds, I find that the average household would be willing to pay $2,385 for information, an 18% increase in average returns over the welfare-maximizing point in Figure 1·9. This underscores that exploiting individual risk characteristics to further refine household responsiveness can increase welfare. Importantly, accommodating for these heterogeneous returns to information explains the average differences between the welfare-maximizing upper-bound $\overline{p}$ predicted by the model and the in-sample rate of diagnosis demonstrated by the data. I explore methods to harness these heterogeneous returns to maximize social welfare of information-revealing social policies in the following section.

Targeting information to maximize gains

In addition to concerns about individual over-responsiveness to health information, policy guiding the revelation of health information must also balance the potentially heterogeneous returns from such revelation. In the face of such variation, full information revelation may not be socially optimal. This includes cases where a full screening regime is not financially feasible, where the information itself may result in consumers declining actuarily fair insurance (Posey and Thistle, 2021), or where there is a disconnect between privately and socially optimal information revelation (Oster et al., 2013). In these cases, the ability to target policies that reveal health risk information may improve the social returns as well as the fraction of households who benefit from these programs.

I estimated strong heterogeneous returns to health information (Appendix Figure D.2). Based on these results, I consider the effects of targeting information revelation based on observable characteristics, such as individual risk scores.[35] I consider a scenario in which individuals can receive a one-time update to information about their health risks, modeled as changes to their probability of adverse health events p_{ift}. When individuals receive this information, this probability is adjusted to be equivalent to their predicted risk probability $\hat{p}_{ift}$ defined above. I assume that following this information, individual beliefs are constant at their predicted risk level, with no residual uncertainty or updating across periods.[36] As before, I assume away salience and moral hazard effects.

This scenario therefore mirrors a hypothetical transmission of health information that informs consumers of their health risks as perfectly as population-data allows.[37] I present results of the individual and social value of this revelation based on 50,000 households in my sample which do not experience major health events. These individuals may still have erroneous beliefs about their health risks and may benefit from new health information. Furthermore, the estimated welfare effects of this policy validates the results presented earlier, documenting the value of information transmitted in a more quasi-random setting.

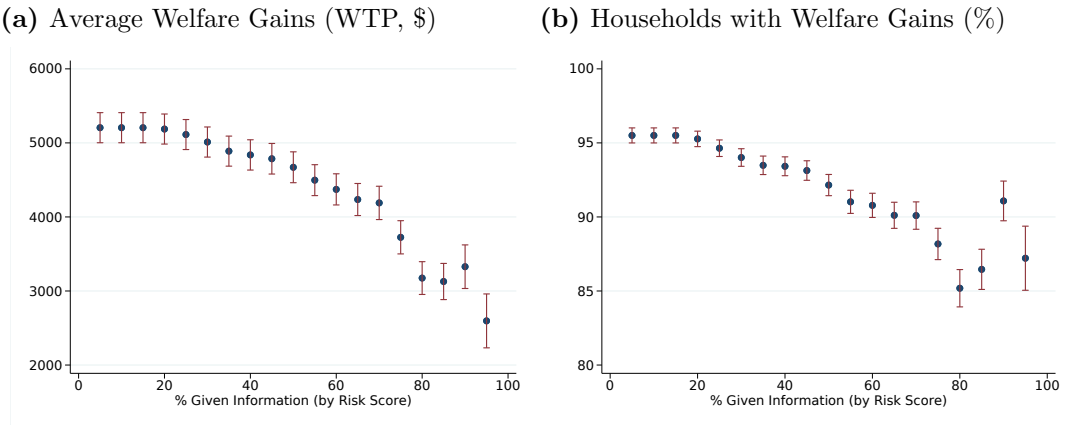
Figure 1·10 presents the results, showing both average welfare gains and the fraction of targeted households benefitting from the information. Each point represents a scenario in which only individuals with risk scores falling in the top $x\%$ of the sample receive health information. The average household in the full sample would be willing

[35] Appendix Figure D.2 illustrates that other, less-easily observable characteristics (e.g., household risk aversion) may also be beneficial.

[36] I ignore residual uncertainty that would arise from individuals treating information revelation as a single signal, rather than true information. For the purposes of this exercise, such fluctuations would serve only to obfuscate the potential benefits of targeting information revelation relative to universal revelation.

[37] Note that $\hat{p}_{ift}$ is not equivalent to one's *true* risk as private information (e.g., underlying health status) is not incorporated.

Figure 1·10
Changes in Welfare Gains From Targeted Revelation of Information



Notes: Figures show estimated welfare gains from revelation of health information. Individuals are organized by their average risk scores, from highest to lowest. Each point in both panels represents a different counterfactual scenario, where individuals with risk scores in the top x% of the sample are given information about their predicted health risks, $\hat{p}$, as described in the text. Returns to health information are presented as (a) average expected welfare changes, measured as willingness to pay in 2020 USD, and (b) the percentage of households with non-negative welfare gains.

to pay approximately $2,500 per year for updated health information (the right-most point in Panel (a)); this information benefits roughly 85% of households (the right-most point in Panel (b)).[38] In contrast, revealing information only to higher-risk individuals improves welfare gains: revealing information only to individuals within the top quartile of the risk score distribution increases average welfare gains to over $5,000 per household per year, benefiting more than 95% of households.

Hence, even policies that are capable of revealing information that closely matches individuals' true risks without inducing salience responses, moral hazard effects, or

[38]Not every household in the sample benefits from information about predicted risk. There are two reasons why even such high-quality information may make a household worse off. First, the household may have private information regarding their true risks, making public information counter-productive. Second, highly risk averse households may benefit from placing smaller weights on the adverse state of the world than are objectively accurate; this is similar to an "optimal expectations" model where individuals do not benefit from information when it lowers utility in an anticipation period (Oster et al., 2013). Overall, this highlights a central tension inherent in the dissemination of health information: even high-quality information can incur individual welfare costs based on how households value health care across states.

over-responsiveness may still benefit from using demographic information to identify households that are most likely to benefit from the policy. For example, policies such as universal genetic screening programs—such as common programs in the U.S. providing risk information to newborns in many developed countries—may incur private welfare costs to specific households, even as they improve societal welfare more generally.

1.7 Conclusion

This paper assesses the extent to which information about one's health risks alters individual and household decision-making in health care. I demonstrate that households who receive new information about health risks from a new diagnosis in the household increase their overall levels of spending, including investments in both preventive and low-value services. These changes in behavior are best explained by individual household members reassessing their risks, rather than responding to financial incentives or salience effects. However, these reassessments do not meaningfully improve the quality of their health care choices. While access to new health information changes behavior in meaningful ways, it does not necessarily do so in welfare-improving ones.

To explore this further, I use a structural approach to quantify a household's willingness to pay for health information, isolating the specific effects of new health information from other mechanisms. The model implies low realized returns to health information, most likely due to individual misinterpretation of their health risks following the health event. Bounding the extent to which individuals increase their beliefs about risks post-diagnosis substantially improves realized welfare. Finally, my analysis illustrates that information revelation is privately most optimal for individuals with high *ex-ante* risk and those with low risk aversion.

The analysis I present could be extended in several meaningful ways. First, future work could relax the assumption that individuals have no control over their chronic care health costs. This would be particularly interesting in non-ESI covered populations, such as those covered by public insurance programs or without any coverage, for whom chronic diagnoses may impose large financial burdens (Hadley, 2007). Another important consideration left out of the model is how liquidity constraints change *ex-post* spending adjustments as health risks change (Gross et al., 2020). Finally, future work might integrate this model with other costs incurred through living with a chronic condition, including earnings penalties and job lock (Biasi et al., 2019; Eriksen et al., 2021; Garthwaite et al., 2014).

Increasing an understanding of how consumers interpret new information is at least as vital as improving their access. Family health experiences are powerful forces in shaping individual behaviors and decisions; however, witnessing these experiences may lead individuals to "over-react" when making future consumption decisions. Individuals and families living with the risk of chronic illness may be better off as they are taught to seek out high-value medical care and temper high expectations of negative outcomes.

Chapter 2

Who Do Innovations Reach? The Influence of Training on Mental Health Treatments

2.1 Introduction

Innovations rest at the heart of many endeavors, and their development, diffusion, and deployment pose critical questions across the spectrum of economic investigation. Generally, innovations studied in economic models are all treated alike, either as random shocks changing a technological process, or a simple event disrupting an equilibrium. In these senses, innovations can be evaluated as though they were policy changes, utilizing many of the simple causal inference tools popular in the field.

However, a more in-depth study of how innovations are discovered and proceed to sway equilibria requires an explicit differentiation of innovation types. Some innovations, for example, are mechanical, such as a software update to a technology that can improve performance for a one-time fixed cost. Others require a more hands-on approach, such as those that require learning-by-doing (Arrow, 1962) or similar methods. Innovations—like many other economic objects—are heterogeneous, and can take on a continuum of values in a potentially high-dimensional characteristic space.

One question that has yet to be asked is how these characteristics affect each innovation's success. It is reasonable that innovations with higher fixed costs, more variation in outcome, or other frictions may diffuse more slowly than innovations with a more straightforward one-time updating cost. Hence, especially as a landscape of innovation tends to the more intangible and artisanal, the spread of new ideas in a field may slow, resulting in gaps between the cutting edge of research and the use of these techniques in practice. Such a gap—commonly referred to as a research-to-practice

gap (RPG)—constitutes an important problem in many areas of research, including healthcare (Glanz et al., 2008; Wandersman et al., 2008; Glasgow and Emmons, 2007) with particular emphasis on mental health (Kazdin, 2011; Kazdin et al., 2017; Kazdin, 2017, 2018; Jensen et al., 1999). Other important fields investigating RPGs include management practices (Bansal et al., 2012; Burke and Rau, 2010; Rynes et al., 2002), education (Coburn and Penuel, 2016; Strohman, 2014), and civil practices such as social work (Rountree and Pomeroy, 2010).

This project studies a RPG in mental health care. Mental health care is a burgeoning field of both research and practice, especially as mental health issues become more prominent in the United States (Olfson et al., 2015). Developments in mental health treatments are typically of two types: pharmacological (e.g., new drugs) or therapeutic (e.g., new models of psychotherapy). My aim is to exploit the differences in these innovations to examine a potentially differentiated rate of innovation take-up among practitioners. I exploit quasi-random attendance of professional trainings (in the form of professional conferences) in both psychotherapy and psychopharmacology among mental health professionals, and assess the impact of each. I implement a panel event study design to assess changes in treatment patterns for therapists who are most likely to attend professional conferences in eating disorder treatments. I explore potentially differentiated responses by provider type and patient demographics, and conclude with an exploration of potential mechanisms for these responses and a validation of my treatment assigning algorithm.

I find muted response among mental health professionals to either kind of professional conference. While this may be the result of an overtaxing estimation process, it provides some suggestive evidence that continuing medical education is not the driver for changes in the treatment behaviors of therapists. Therapists did increase their use of olanzapine (an atypical antipsychotic occasionally prescribed for eating disorder treatments, discussed more in Section 2.2.2). Interestingly, this response occurred only among non-psychiatrists (e.g., psychiatric nurse practitioners) and was used on adolescent patients. However, therapists did not have a similar response to therapeutic education; in fact, when exploring the overall variation in a provider's treatment profile, I find suggestive evidence that a conference *discourages* experimentation.

For clarity, in this paper I make the (somewhat informal) distinction between *tangible* and *intangible* innovations. Tangible innovations are algorithmic in nature: while they may require specific skills and training to be able to implement, their

implementation requires little creativity and varies little across implementations and practitioners. Many of the innovations that come easily to mind—new drugs, medical equipment, etc.—fall into this category. In contrast, intangible innovations depend more heavily on human capital, and therefore can vary widely based on who is implementing it (or even across cases with the same practitioner). The example of intangible innovation used in this project is psychotherapy, which is a rigorous and scientific medical treatment, but also requires a conscious cultivation of relationship between therapist and patient that is impossible to achieve algorithmically. While new therapeutic techniques can be proposed and validated by mental health researchers, the passing on of these guidelines from researcher to practitioner will inevitably leave room for practitioners to adapt the practice to their own treatment style, potentially altering the benefits of the development. Other examples of intangible innovations in health care include testing and prescription guidelines (Obermeyer et al., 2019), as well as any other behaviors subject to clinician interpretation.

Of course, this distinction is a simplifying one, as nearly all innovations contain elements of both "art" and "science". For example, Graham et al. (2019) discuss the implementation of new digital mental health technologies, an ostensibly algorithmic innovation (e.g., a cell phone application) that requires specialist understanding of the mechanisms at play in order to be successfully integrated into a treatment plan. While elements of artisanal and algorithmic innovations exist in almost every development (particularly in a field such as mental health), I have attempted to choose two key innovations that are as close to purely tangible and intangible as possible: psychotropic medication and psychotherapeutic techniques.

The contributions of this study are both methodological and practical. First, this study proposes a way to point identify dynamic treatment effects even in the presence of classification error. This extension of recent work (most notably, Calvi et al. (2019)) increases researchers' flexibility to answer causally motivated questions in the presence of limited data, as well as suggesting ways predictive algorithms (such as machine learning techniques) could be used in causal designs. In addition, this paper discusses how an interpretation of these results might change when the necessary assumptions are implausible or hold only partially, and outlines how validation samples can be used to test the necessary assumptions.

The factors and frictions affecting technology diffusion is a central question in economics—and health economics in particular—and this paper contributes to this

rich literature by assessing diffusion under heterogeneous take-up costs *on the innovation side*. Recently, this literature has been concerned with proposing explanations for heterogeneous rates of innovation take-up; these solutions explore factors such as differences in social network structures (Arieli et al., 2020), the presence of network effects (Ackerberg and Gowrisankaran, 2006), and variations in take-up costs among potential users (Ryan and Tucker, 2012). However, each of these projects considers only one innovation at a time in order to prioritize demand-side heterogeneity (Young, 2009). In contrast, the current project examines how different innovations—with potentially varying take-up costs—compete for takeup among practitioners. The current setting allows for identification of reduced form evidence exploring the ways these costs drive differences in take-up within a specific clinical population (mental health professionals).

From a clinical perspective, this project also contributes to a broad discussion on gaps between research and practice by highlighting one of the most common frictions in the diffusion of ideas: communication. Some papers find strong responses of medical professionals to randomized trials (Depalo et al., 2019), but the dissemination of this information is not always straightforward (Grimshaw et al., 2001; Casper, 2007). Continuing education is the most common method by which medical professionals receive information about new medical research (Church et al., 2010). However, as even medical conferences become more specialized, tailored either to academics[1] or professionals, continuing education has the potential to devolve into a "blind leading the blind" environment, where the trainers are as removed from medical research as the trainees. This, and many other factors, warrants an evaluation of continuing education as a potential source of research-to-practice gaps. This study contributes not only to a discussion on the uses of continuing education, but also a much larger literature on innovation diffusion in intangible settings.

Finally, this paper is also tangentially related to a burgeoning literature on the diffusion of ideas, a discussion on how intangible goods such as international ideals (Gilardi, 2012) and social movements (Rane and Salem, 2012). For example, Ash et al. (2019) examine the spread of economic language among judges following a training program. Their particular type of policy evaluation (with dynamic treatment effects) is similar to the aims of this paper.

[1]For example, the Eating Disorders Research Society holds an annual conference limited only to its members. As a result, only academics attend, not professionals.

2.2 Background & Data

The diffusion of innovation into practice is a central issue for nearly every area of technological advancement. In simple cases, standard economic models predict that technologies that increase marginal benefit or decrease marginal cost will have quicker take-up by practitioners, becoming a new norm until further innovation disrupts the equilibrium again (Christensen et al., 2009, 2006, 2015). However, in the presence of frictions, the diffusion of innovations may depend on much more than their simple benefit/cost contributions, and standard models may be insufficient to predict how a field will develop. This is particularly true when innovations are intangible in nature, as this makes them particularly vulnerable to frictions.

2.2.1 Research to practice gaps

One friction that is particularly salient in the diffusion of intangible medical innovation is a growing divide between academics and professionals (Kazdin et al., 2017; Kazdin and Blase, 2011). With increasing specialization, a burgeoning field such as mental health care becomes split into two camps: one performing and reporting the results of clinical trials and other research, and a second that interprets and incorporates these results as they treat real patients. However, as this specialization progresses, the distance a new idea must travel from the laboratory to the patient increases, raising the chances that it will either not be adopted, or adopted in some stunted capacity.

Communication between these two groups—especially in the medical profession—is incentivized through continuing medical education (CME) programs for practitioners. These programs are motivated by the documented fact that physicians who have been practicing longer tend to stall in updating their practices, putting them at risk for delivering lower-quality care (Choudhry et al., 2005). While the structure of CME programs tends to vary across states and facilities, a typical curriculum generally requires a mix of completing courses taught by state-approved providers, preparing and teaching courses to other professionals, and presenting at professional conferences, with additional options for research, publications, or media involvement. Licensures may be awarded following the completion of certain milestones in a CME program, allowing a mental health professional to advertise as "licensed" in an attempt to increase demand.

In recent years, CME programs have evolved to allow online learning through approved online classes, webinars, and presentations. This has been done largely to reduce the burden continuing education places on rural physicians (Curran et al., 2006) and improve access more generally. In fact, Hugenholtz et al. (2008) have demonstrated that online continuing education is just as effective as traditional, in-person lectures. Despite this, most states still require at least some continuing education to be done in person. Because of this, professional conferences continue to be hubs for continuing education presentations, exams, and courses.

The potential benefits for professional conferences are inherent in the nature of the event, and tend to be highly favored by practitioners (Dysart and Tomlin, 2002). In fact, according to Dysart and Tomlin, professional conferences are attended with about the same frequency as all other continuing education events combined;[2] however, their work also highlights the difficulties associated with receiving education through expensive and travel-intensive methods such as conference attendance. Healthcare facilities are rarely generous in providing time off for conference attendance, and conference and travel fees are typically borne by the provider rather than the employer.

2.2.2 The case of eating disorder treatments

In an attempt to assess the quality of communication and training in inducing innovation takeup, the current project examines continuing education on practices in the treatment of eating disorders. These mental disorders centered around unhealthy relationships with food and eating. They include *anorexia nervosa*, typified by body dysmorphia and severe restriction of food intake; *bulimia nervosa*, characterized by purging excessive food consumption; *binge eating disorder*, a disease marked by superfluous food consumption (but no purging); and other unspecified diseases. This study will focus on patients with diagnoses of either anorexia nervosa or bulimia nervosa exclusively.[3]

These diseases are ideal for the current study for three principal reasons. First, these diseases have the highest mortality rate of any mental illness (Arcelus et al., 2011), making them a pragmatically relevant area of focus. Second, treatment of

[2]Their study examined occupational therapists, rather than mental health professionals.

[3]Note that binge eating disorder did not have its own diagnosis code until the release of the ICD-10-CM Diagnosis Codes, which were used beginning in October 2015 (after my sample started).

eating disorders involves both algorithmic and intangible processes: for example, the refeeding process of severely malnourished anorexic patients is medically more straightforward than the psychotherapeutic aspects of treatment. However, as discussed in more detail below, many of the algorithmic treatment methods—such as pharmacological treatments—have much weaker empirical support than psychotherapies. Hence, in the absence of a research-to-practice gap biasing treatments towards algorithmic interventions, one should observe the use of intangible treatments (e.g., psychotherapy) dwarfing the number of pharmacological interventions. Finally, the study of eating disorders meshses well with available data. While it is a myth that they affect only female adolescents from the middle- and upper- classes (Mitchison et al., 2014), a substantial number of those suffering from this disease will have private insurance. Additionally, there are easily identifiable diagnosis codes for each eating disorder and treatment codes for the two treatments of interest (family-based therapy and olanzapine prescriptions, discussed below). Therefore, I have a clean identification of the populations and outcomes of interest.

While treatment patterns vary for each individual patient, treatment of eating disorders is recommended to follow a team-based model of care (American Psychiatric Association, 2006), with the team generally comprised of a principal psychotherapist, a dietitian (and other general medicine professionals if needed to deal with secondary effects of the disorder), a psychiatrist, and occassionally a social worker. Treatment proceeds in stages, with early stages focused on rectifying any secondary effects of an eating disorder (e.g., a re-feeding or rapid weight gain process), and later stages focusing on mental health treatments. Hospitalizations—if any are required—typically take place in the first stages, with the latter stages largely taking place in an outpatient setting. It is this latter, mental health-oriented stage, with which this project is concerned. This stage typically consists of two major treatment modalities: psychotherapeutic and psychopharmacological.

Family-based therapies (FBT) are considered an optimal therapeutic intervention for the treatment of anorexia nervosa, bulimia nervosa, and eating disorders not otherwise specified (Loeb et al., 2012). In this treatment, family members of a patient are integrated into a team of health professionals, as opposed to other psychological practices, which at best ostracize family members and at worst paint them as responsible for mental illnesses (Le Grange et al., 2010). FBT, developed at the Maudsley hospital in London by Dare and Eisler (2000) and manualized for anorexia nervosa

by Lock and Grange (2015), currently boasts the strongest empirical support of any psycho-therapeutic intervention for treating anorexia nervosa, including hospitalization.[4] The two most recent meta-analyses—Couturier et al. (2013) and Bulik et al. (2007)—each conclude that family-based treatments are more efficacious than many routine treatment methods, particularly for adolescents and youth. Importantly, the advantages of FBT are most notable in the long-term, with positive impacts 6–12 months after treatment that outweigh even the benefits of individual cognitive-based therapy (Couturier et al., 2013). As these authors write:

> "Family therapy focusing on symptom interruption of eating disordered behaviors should be recommended as the first line of treatment for adolescents with eating disorders. Given the growing evidence base for FBT for adolescents with eating disorders, it would be prudent to study implementation strategies and effectiveness of this treatment in the community." (Couturier et al., 2013)

Family-based therapy is recommended by the American Psychiatric Association and the National Institute for Health and Care Excellence in the UK as the main intervention for eating disorders (American Psychiatric Association, 2006). Despite this, however, the overall use of FBT in eating disorder treatments in the outpatient setting remains consistently low. Figure 2·1 shows the percentage of all eating disorder patients in the MarketScan data receiving any form of family-based treatment over time. The graph shows that only around 15% of the 23,000 patients in the sample (and around 26% of the 10,000 youth and adolescent patients) ever receive FBT in their treatment. Furthermore, the graph shows the publication dates of major RCTs and meta-analyses positively evaluating FBT, with little implied physician response shown as a result. This suggests that providers may already have sorted into those who provide FBT to their patients and those who do not, and that the current stream of ongoing research does not affect their decision to provide this treatment.

Of course, FBT will not be ideal for every eating disorder patient. Factors such as family instability, need for longer treatment, and co-morbid psychiatric disorder may influence a patient's lack of response to FBT (Lock et al., 2006). Additionally, FBT has been proven more useful for adolescents than adults (Bulik et al., 2007).

[4] A complete list of randomized control trials (RCTs) evaluating the effectiveness of FBT for eating disorder treatments can be found here.

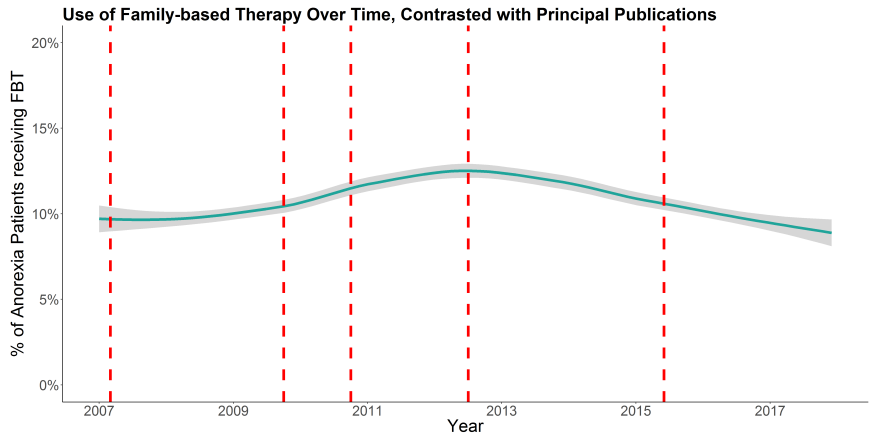


Figure 2·1
Therapist Response to Publications on Family-Based Therapy

Finally, some specialists are able to use family-based techniques across a wide range of diagnoses outside of eating disorders; this may incentivize certain mental health practitioners to specialize in a family-based approach, allowing specific patients to seek out this treatment modality if they feel it may be a good match for their needs.

The second major treatment modality for eating disorders is pharmacological; however, the evidence base for this style of treatment is scant relative to that of therapeutic techniques. There are only two FDA-approved medications for eating disorder treatments: fluoxetine for bulimia nervosa (approved in 1994) and vyvanse for binge eating disorder (2015), both of which suppress purging behaviors. Additional medications—particularly selective serotonin reuptake inhibitors (SSRIs) or other antidepressants—are typically prescribed to assist in mitigating co-morbid depression and/or anxiety symptoms (American Psychiatric Association, 2006). Overall, there are no good pharmacological treatments to handle a patient's relationship with food, making the therapeutic treatment arm essential.

Even without empirical support, an increasing number of prescribers have begun engaging in off-label experimentation in the treatment of eating disorders (Maglione and Hu, 2011). Much of this experimentation uses atypical antipsychotics—which are FDA approved and typically prescribed for schizophrenia and bipolar disorder—to manage weight gain. For example, olanzapine (the most commonly prescribed atypical antipsychotic for eating disorders) is known to induce weight gain as a common side effect, and hence has been viewed as potentially useful in anorexia nervosa treatments (Flament et al., 2012). While there have been some studies examining these

medications (see Maglione and Hu (2011) for a meta-analysis), there is not enough conclusive evidence that these medications are effective in treating eating disorders to warrant a change in their FDA approval status presently; however, continuing education and professional conferences still include discussions of incorporating off-label drugs into psychopharmacological practice in an ED treatment profile.

2.2.3 This project

This project focuses on a single potential friction between academic research and practice: the impact of professional education. Specifically, I focus on the implementation of FBT and prescription of olanzapine in eating disorder treatments, two innovations that embody different styles of innovation and may thus diffuse differently. The prescription of olanzapine, while an off-label practice with relatively little empirical support, has a straightforward implementation, and constitutes a more algorithmic innovation. However, the use of family-based therapies requires specialists to provide a higher level of care, and its implementation therefore varies widely across therapists, in keeping with intangible innovation. This heterogeneous implementation of FBT in eating disorder treatments has been documented in Kosmerly et al. (2015).

To evaluate this takeup, I use a list of about 70 conferences targeted at eating disorder professionals and clinicians. For each conference whose online program is available, I am able to ascertain if the conference ran any sessions or presentations on either FBT or olanzapine use in ED treatments, as well as creating a registry of the conference locations and times. Table 2.1 shows the organizations and conferences examined. Aside from conferences whose programs are not available, I have the universe of such professionally-oriented conferences.[5] I couple this with a sample of 4,476 therapists and professionals from the Truven MarketScan data to examine treatment profiles of specialists before and after conference attendance.

The main complication is that I have no data on who actually chose to attend each conference;[6] instead, I will estimate treatment status based on each specialist's cost (in travel) of attending a conference. By assuming that therapists are more likely to attend conferences that are low-cost to them, I am able to artificially assign

[5]Note that this excludes academic conferences which are limited to members of the academic organization only (for example, the Academy of Eating Disorders) as not all clinicians would have the opportunity to attend.

[6]Note that I have data on conference registration for a few conferences, which I will use in a validation exercise in Section 2.5.

Table 2.1
Professional Conferences on Eating Disorder (ED) Treatments Examined

Organization	Conference Name	Frequency	Total Programs	FBT	Olanzapine
Academy for EDs	International Conference on ED	Annual	9	7	4
American Academy of Child & Adolescent Psychiatry	AACAP Meetings	Annual	10	2	4
Annual Eating Recovery Foundation	ERF Conference	Annual	3	3	0
Center for Change	National ED Conference for Professionals	Annual	5	1	
International Association of ED Professionals	IAEDP Symposium	Annual	9	6	2
Maudsley Parents	One-Day FBT Conferences	Sporadic	3	3	0
Multi-service ED Association	MEDACon	Annual	3	2	1
National ED Association	NEDACon	Annual	4	3	0
Renfrew Center Foundation	Conference for Professionals, Seminar Series	Annual+	15	8	0
Center for ED at Sheppard Pratt	Professional Symposium	Annual	6	3	0
Summit for Clinical Excellence	National ED Conference	Sporadic	4	1	1
UCSD ED Treatment Center	Trainings for Professionals	Sporadic	2	1	0
		Total:	73	40	12

[1.] *Abbreviations:* ED = eating disorder; AACAP = American Academy of Child & Adolescent Psychiatry; ERF = Eating Recovery Foundation; IAEDP = International Association of Eating Disorder Professionals; FBT = Family-based therapy; UCSD = University of California at San Diego.

specialists to treatment and control groups, as discussed in more detail in Section 2.3.1. Finally, I extend recent work on dealing with classification error in treatment effect models (Calvi et al., 2019) to approximate the local average treatment effect of attending these conferences.

Hence, this paper provides two distinct contributions. The first is methodological in nature, and presents a toolkit of econometric techniques to assist researchers in overcoming data limitations. Specifically, this paper introduces an instrumental variables technique for the event study approach, integrates predictive algorithms into a causal framework, and extends results that adjust these frameworks for classification errors. Due to the reasonably complicated procedure by which my results are derived, several sections of this paper are dedicated to the exposition of the algorithm and intuition behind its use.

Secondly, I present information detailing how medical professionals respond to continuing education in the form of professional conferences. I argue that these responses are potentially differentiated on the basis of which techniques or tools are being discussed, and examine heterogeneity by audience (specialist type) and population of interest (patient demographics). Ultimately, the results of this exercise provide little evidence that continuing education changes behavior in the aggregate, either for intangible or algorithmic innovations (psychotherapy and prescriptions, respectively).

This finding warrants future research in light of the severe data limitations and complex econometric procedure, which is taxing for the available data; however, if true, this finding suggests a need to better understand the optimal way to transmit information to practicing professionals.

2.3 Empirical Design

Dynamic treatment effects are at the heart of questions surrounding innovation adoption. I am ultimately interested in how professional conferences impacted the use of FBT and olanzapine *over time* for each specialist who attended. I have concrete data on each specialist's treatment profile for their subset of patients who are covered by an insurer in the MarketScan database; however, I do not have reliable data on conference attendance for these physicians. My empirical approach will (i) estimate treatment status for each medical professional and conference, (ii) estimate a dynamic treatment effect of professional education using an event study framework, and (iii) adjust for potential classification error in the first step.

This project combines various econometric approaches to attempt point identification of my dynamic treatment effect of interest. To fill in data gaps, I employ a predictive algorithm that infers who attends each conference; this suggests a place for more sophisticated machine learning techniques in causal research designs. To deal with the flaws inherent in any such algorithm, I extend an estimator that is robust to measurement error in a treatment variable to an panel event study framework.[7] Using this estimator in tandem with a transformed IV approach allows me to approximate a Dynamic Local Average Treatment Effect (D-LATE) for the specialists in my sample who are induced to take-up treatment (the compliers).

Event study designs have become increasingly popular in recent years (see Sun and Abraham (2021) and Borusyak and Jaravel (2021) for important reviews on the subject). These designs rely on variation in treatment time (with or without the presence of a control group to explore treatment effects in periods both leading up to and following the treatment period, as well as the presence of a control group to correctly control for time fixed effects (Hull, 2018). This design can flexibly be used

[7]The Mismeasurement Robust LATE Estimator of Calvi et al. (2019), discussed in more detail in Section 2.3.3.

to explore heterogeneous responses in an appealing way when the number of groups to compare is relatively small, as in Johannesen and Stolper (2017).

2.3.1 Estimating treatment categories

While my data are ideally suited for the study of a medical professional's treatment profile, they contain no information on continuing education or conference attendance. Hence, I use a predictive algorithm to infer each specialist's decision to attend a CME conference based on their travel costs. The algorithm is based off of the assumption that given that opportunities for continuing education are nearly ubiquitous, decisions to attend conferences for professionals will be driven largely by costs: an ED specialist in Boston is far more likely to attend conferences when they are held in New England than when they are held in California.

Details of this algorithm are relegated to Appendix B for brevity. In general, for each mental health specialist and each conference, I compute a measure of travel cost taking into account both the physical cost of travel and the opportunity cost of time. From this continuous measure, I infer a treatment group as the smallest η-percentile of specialists when ranked by their travel costs. This move from a continuous variable to a discrete one is motivated by the classification error framework laid out in the next subsection; by varying this threshold I change the probabilities of misclassifying a treated/control therapist in a near-monotonic fashion.[8] This is useful for the assumptions of the MR-LATE estimator discussed in Section 2.3.3. However, future research might explore the potential use of this continuous measure in a propensity-weighting framework, as well as how such a framework compares to that of Calvi et al. (2019). Additionally, future research could integrate more sophisticated machine learning techniques to improve prediction accuracy, providing a better approximation of the true LATE (as discussed in Section 2.3.3).

Figures 2·2 and 2·3 show an example of the algorithm's output for a sample conference for professionals that took place in September 2012 in Boston. Figure 2·2 shows the estimated travel cost to attend the conference for each specialist in the sample at that time, while Figure 2·3 shows the estimated distribution of travel costs, including various cities as reference points. Specialists in cities farther away from

[8]That is, as η increases, I tend to increase the probability of classifying a control therapist as being treated, while decreasing the probability of classifying a treated therapist as part of the control group.

Boston incur greater travel costs to attending the conference, but those in distant *rural* areas (such as Mountain Home, Idaho) incur even greater travel costs. By selecting the lowest η-quantile of the distribution, different treatment groups are created, with differing levels of austerity in selecting the treatment (or control) groups[9]. Notice that these treatment groups are not merely centered around the conference location—indeed Atlanta, Georgia, which is a hub for major airlines, has a lower travel cost to a Boston 2012 conference than does New Haven, Connecticut. This illustrates that incorporating travel costs into the predictive algorithm may provide an improvement in prediction quality over a simple geographic distance calculation.

When repeated for all conferences, this procedure creates estimated treatment groups for each of the conferences in the sample (40 conferences for FBT trainings and 12 for olanzapine prescriptions). In order to conduct an event study analysis, it is important that treatment be an absorbing state for each therapist;[10] hence, each therapist is assigned a treatment date as the earliest time period for which it is estimated that they attended a conference on FBT or olanzapine. From this estimated treatment time, relative time dummies typical for an event study are created, completing the necessary data configuration.

2.3.2 Instruments in event study designs

While the use of instruments in an event study is rare compared to their prevalence in other empirical designs, the generalization from the instrumented difference-in-differences design (DDIV) to an instrumented event study design (ESIV) is straightforward. As explained in Hudson et al. (2017), the basic model for the DDIV is Equation 2.1:

$$y_{it} = \alpha_i + \tau_t + \beta D_{it} + \epsilon_{it}, \tag{2.1}$$

where α_i and τ_t represent individual and time fixed effects, and D_{it} is the binary (potentially endogenous) treatment status.[11] To deal with the endogeneity of treatment D_{it}, a binary instrument Z_{it} is used.

[9]The red line in Figure 2·3 illustrates a treatment group based on the lowest 10% of travel costs.

[10]That is, each specialist ought to be treated only once, and remain treated throughout the duration of the sample after that.

[11]Note that additional controls can be added if desired. I ignore them in this section to simplify the exposition.

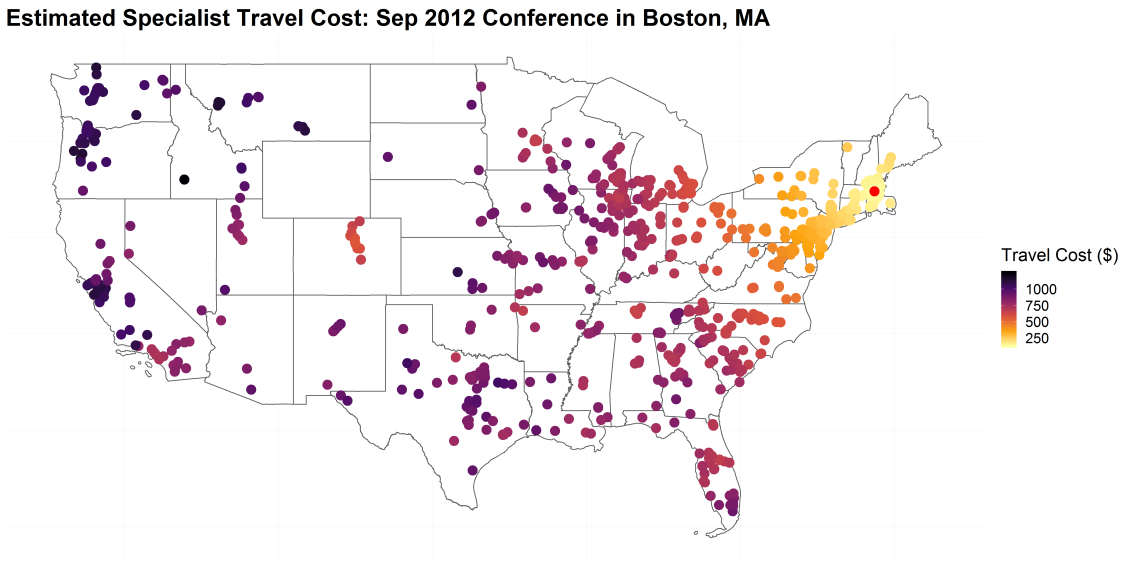


Figure 2·2
Estimated Travel Cost for all Specialists, Boston 2012 Conference

The event study framework generalizes this by mapping between a single treatment indicator D_{it} and a vector of *relative time dummies*, which indicate how much time has elapsed since the treatment event. For each individual i in a panel, the event is denoted as $E_i = \min_t\{D_{it} = 1\}$; given this, each period t can be assigned a value $K_{it} = t - E_i$. This essentially re-orders the time periods in a panel so that each individual appears to have been treated simultaneously. Once this is complete, the estimating equation can be written as Equation 2.2

$$y_{it} = \alpha_i + \tau_t + \sum_{k=-\infty}^{\infty} \gamma_k \mathbb{1}\left\{K_{it} = k\right\} + \epsilon_{it}. \tag{2.2}$$

In this setup, each parameter γ_i indicates the effect of the treatment event on the outcome variable i periods before or after the event itself. See Borusyak and Jaravel (2021); de Chaisemartin and D'Haultfoeuille (2019) for a more detailed discussion of the event study approach. Generally, applied researchers do not estimate the fully dynamic specification (where k ranges over all integers), but limit $k \in [-A, B]$ for two positive integers A and B[12]. This establishes the vector parameters $(\gamma_0, \gamma_1, ..., \gamma_B)$ as the parameters of interest (sometimes referred to as the *dynamic treatment effect* parameters).

[12]For identification, such an approach requires omitting a dummy as a reference group, which is typically chosen to be γ_{-1}.

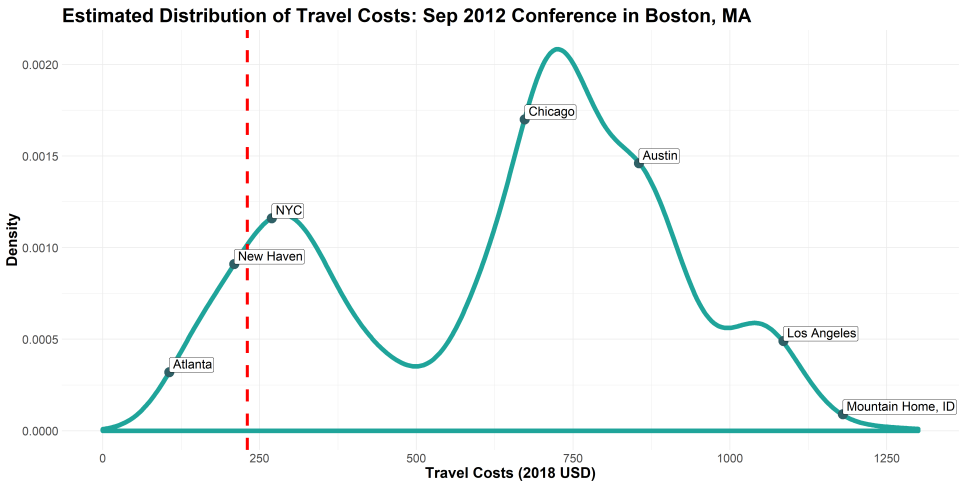


Figure 2·3
Estimated Distribution of Travel Cost for all Specialists, Boston 2012 Conference

Suppose now that there exists a valid instrument Z_{it} for D_{it}. To transform this instrument to be a valid one for the event study approach, one need only follow the same procedure outlined above: for each individual i, define the instrumented event Z_i' as the point that is most likely to induce the treatment event, then define the relative time periods $Z_{it}' = t - Z_i'$ as before. Given that Z_{it} is correlated with D_{it}, the transformed instrument Z_{it}' will be correlated with K_{it}, ensuring that the procedure is valid.

My main instrumental variable is the presence of a "slow spell" for a therapist in the months leading up to a conference. Specifically, if a specialist's average patient volume 4–6 months prior to a conference is lower than their overall average volume, the binary instrument is given a value of 1 (and 0 otherwise). By indicating a potential decline in patients treated during a registration period for a conference, I hypothesize that this instrument will be positively correlated with true conference attendance. Additionally, since such a measure is uncorrelated with both (i) distance between therapist and conference and (ii) therapist treatment profiles, this measure is a valid instrument for the treatment. I therefore define Z_i' as the period with the lowest measured *lagged* patient volume for each specialist i.

2.3.3 Dealing with classification error

Given that I infer treatment status based on an imperfect proxy (travel costs), dealing with classification error is a first order concern in my estimation approach.[13] There is a small, but vibrant, literature on dealing with classification errors in applied microeconometric models. The most notable of these papers, Lewbel (2007a) point identifies the average treatment effect (ATE) in a simple treatment effects model. Other important papers extend this result to include covariates or discretized treatment levels (Lewbel, 2007b; Mahajan, 2006; Hu, 2008). Most recently, these researchers have turned to the problem of estimating Local Average Treatment Effects (LATEs) in the presence of potentially endogenous selection into treatment. This paper extends the recent work of Calvi et al. (2019), who identify a mismeasurement-robust estimator of the LATE (the MR-LATE) used for bias reduction in classification error problems.[14]

The MR-LATE estimator of Calvi et al. (2021)

Calvi et al. (2019) propose an estimator that is "mismeasurement robust" in the sense that it can approximate the LATE under weak assumptions. In their framework, there is a true treatment status $D \in \{0, 1\}$, which is unobserved and cannot be consistently estimated. In addition, there exists a binary instrument Z such that the typical LATE assumptions of Imbens, G. W. and Angrist, J. D. (1994) are satisfied (as replicated in Assumption 1).

Assumption 1: LATE Assumptions. The outcome Y and true treatment status D, together with a binary instrument Z satisfy:

i. $0 < \mathbb{E}[D] < 1$, $0 < \mathbb{E}[Z] < 1$, and $Z \perp (Y_1, Y_0, D_1, D_0)$.

ii. (Y_1, Y_0, D_1, D_0, Z) are independent across individuals and have finite means.

iii. There are no defiers, hence $\mathbb{P}(D_0 = 1 \cap D_1 = 0) = 0$,

[13] Classification error refers to measurement error in a variable denoting treatment status. Ignoring this error—which by construction is nonclassical—can lead to serious problems in estimating a treatment effect, as discussed in detail in Millimet (2010). Kreider (2010) shows that even in a case of infrequent classification error—from 2% or less—can result in estimated effects whose confidence intervals do not overlap the true treatment effect, and may even suggest the opposite sign of the true ATE.

[14] There is another recent paper that tackles this issue (Yanagi, 2018), but this requires additional assumptions and applies to a more restricted class of circumstances.

where subscripts are indicative of potential outcomes following the typical framework. While D cannot be consistently estimated, it is approximated by two imperfect measures $T^a, T^b \in \{0, 1\}$. These measures satisfy an extended set of the LATE assumptions in Assumption 2 (where compliers are denoted by C):

Assumption 2: Mismeasured LATE Assumptions. T^i is such that the following conditions are satisfied for $i \in \{a, b\}$:

i. $Z \perp (Y_1, Y_0, D_1, D_0, T_1^i, T_0^i)$.

ii. $(T_1^i, T_0^i) \perp (Y_1, Y_0) | C$.

iii. $\mathbb{E}[T_1^i - T_0^i | C] \neq 0$.

That is, in addition to the typical unconfoundness assumption, Assumption 2-i assumes the instrument is independent of the potential measurement errors in T^i. The second part of the assumption indicates that the potential outcomes of each mismeasurement are independent of the potential outcomes of the dependent variable Y; combined with the first assumption, this asserts that any measurement errors are uncorrelated with outcome variables. Finally, Assumption 2-iii requires only that T provide some information about D.

Given these two assumptions, Calvi et al. (2019) apply the reasonable well-known fact that a transformed two-stage least squares (2SLS) regression of YT on T (using Z as the instrument) can be written as a mixture of the potential outcomes for compliers:

$$\frac{\text{Cov}(YT^i, Z)}{\text{Cov}(T^i, Z)} = \frac{\mathbb{E}(YT^i | Z = 1) - \mathbb{E}(YT^i | Z = 0)}{\mathbb{E}(T^i | Z = 1) - \mathbb{E}(T^i | Z = 0)} \tag{2.3}$$

$$= \mathbb{E}\left[qY_1 + (1 - q)Y_0 | C\right], \tag{2.4}$$

where q is a weight related to the probability of measurement errors in T given true treatment D. Given this result,[15] Calvi and coauthors define the MR-LATE estimator as the difference in two 2SLS estimators, given the two mismeasured treatments T^a

[15] This result is not unique to Calvi et al. (2019), but has been mentioned in earlier work, including Abadie (2002) and Ura (2018).

and T^b:

$$\text{MR-LATE} \equiv \rho = \frac{\text{Cov}(YT^a, Z)}{\text{Cov}(T^a, Z)} - \frac{\text{Cov}(YT^b, Z)}{\text{Cov}(T^b, Z)}.$$

Using this definition and the result from their first theorem (Equation 2.4), it follows immediately that the MR-LATE is a multiple of the LATE, with the weighting $(q^a - q^b)$; Hence, the MR-LATE is equal to the true LATE when $(q^a - q^b) = 1$. A sufficient condition for this to hold is that of Assumption 3:

Assumption 3: Sufficient Condition for MR-LATE = LATE. T^a and T^b are such that the following two conditions are met:

i. $p_0^a = 0$. That is, among compliers, the mismeasured treatment T^a never mistakes the actually untreated as treated.

ii. $p_1^b = 0$. That is, among compliers, the mismeasured treatment T^b never mistakes the actually treated as untreated.

These restrictions—that one treatment group is strict in its definition of the treatment group, and the other in its definition of the control group—are related to the no-defiers assumption typical in a LATE framework. By eliminating certain combinations of D and Z, the no-defiers assumption allows for a clean interpretation of the local average treatment effect. In a similar vein, Assumption 3 rules out certain types of measurement errors, thereby eliminating extraneous cases wherein the MR-LATE would be different from the true LATE.

As in cases where the no-defiers assumption is violated, an MR-LATE estimator *approximates* the LATE in cases where Assumption 3's conditions are nearly met (meaning that $q^a - q^b$ is close to one). Judging the extent to which the conditions of these assumptions are met is typically impossible given the limitations of the data; however, I have obtained actual conference registration data from recent ED conferences held by the Academy for Eating Disorders, which I use as a validation sample in Section 2.5. With this new data, I am also able to address concerns about a lack of strong identification arising from an imprecise treatment group estimation.

This paper: The dynamic LATE (D-LATE) estimator

Extending Lewbel's work to the event-study setting is relatively straightforward. Theorem 1 below restates the necessary setup and assumptions for the MR-LATE to be identified for each parameter β_i of the dynamic treatment effect.

Theorem 1 *Let $\{Y, D, Z, T^a, T^b\}$ be such that Assumptions 1 and 2 are satisfied. Consider estimating an instrumented event study regression (equation 2.2) on the transformed variable T^iY using T as the treatment measure and Z as the instrument. Then, for any time period t relative to the treatment period, the dynamic treatment coefficient γ_t satisfies*

$$\gamma_t = \mathbb{E}[qY_1 + (1-q)Y_0|C], \tag{2.5}$$

for a q related to the probability of mismeasurement in the substitute treatment measure T^i.

See Appendix A for a proof of this theorem. This extension of the theorem relies on two facts: first, that an event study design is simply a transformation of the DDIV estimator into one with many dummy variables, as discussed in 2.3.2. Hence, estimating a LATE model with one instrument is equivalent to estimating a corresponding ESIV model with many instruments (one for each dummy). Second, as discussed in Angrist and Imbens (1995), coefficients in two-stage least squares models with multiple instruments can be written as a linear combination of each instrument-specific LATE.

Given the results on Theorem 1, the corollary of Calvi et al. (2019) immediately implies that a dynamic version of the MR-LATE (which I call the dynamic MR-LATE or D-LATE for short) is equivalent to the true LATE under the conditions stipulated in Assumption 3. Hence, in order to resolve issues of classification error while still obtaining a dynamic treatment effect, I use two measures of treatment status—one that never misclassifies the treated, and another that never misclassifies the untreated—and the quasi-randomized instrument of patient volume during the conference registration period, as discussed in Section 2.3.2.

For the two mismeasured treatment estimates, I can use the travel cost algorithm described in the preceding subsection with varying thresholds. That is, I create two estimated treatment groups for each conference, one with a very strict threshold for

attendance (e.g., only the lowest ventile of travel costs) and one with a very liberal threshold (e.g., the 95th percentile of travel costs). In this way, I ensure that one of the mismeasured treatments is unlikely to mistake a truly treated professional as a control member, and the other is unlikely to make the opposite mistake, thereby at least approximating the sufficient conditions for the D-LATE estimator to be equivalent to the LATE.

I therefore obtain estimates and standard errors of the D-LATE using the following procedure. First, I estimate two event study regressions (using equation 2.2) using T^iY as the dependent variable, T^i as the treatment status (that determines the dummy variables), and Z as the instrument. The MR-LATE estimator for each coefficient of interest γ_i is given by $\gamma_i^{MR} = \gamma_i^a - \gamma_i^b$. Finally, I obtain bootstrapped panel errors for each coefficient use the panel bootstrap method.[16]

2.4 Estimation Results & Heterogeneity

The D-LATE estimator was implemented to evaluate two sets of professional conferences: one targetting the use of family-based therapies (FBT) and another the prescription of atypical antipsychotics (olanzapine) in eating disorders. In both cases, I am interested in the effect these conferences have on individual therapist experimentation; I therefore measure short-term responses to a conference by the likelihood of utilizing an innovation in the first 6 months following the event.

The main results of the event study on FBT takeup can be seen in Figure 2·4. The point estimates suggest that in the month following conference attendance, FBT techniques were about 8 percentage points more likely to be employed. However, large bootstrapped standard errors and large pre-trend effects suggest that this result is more attributable to sampling variation than a true therapist response. Even if there is a short-term response, it quickly diminishes in the subsequent periods, suggesting a short period of experimentation without true adoption. As I will discuss in Section 2.4.1, this result is robust to multiple specifications.

A similar result holds for olanzapine prescriptions, as seen in Figure 2·5. The estimated coefficients for this treatment effect are much smaller, with at most a 0.4 percentage point increase in prescriptions after conference attendance. Overall, the results suggest little, if any, change in prescribing behavior. The fact that this

[16]See Kapetanios (2008) for an excellent review of this procedure.

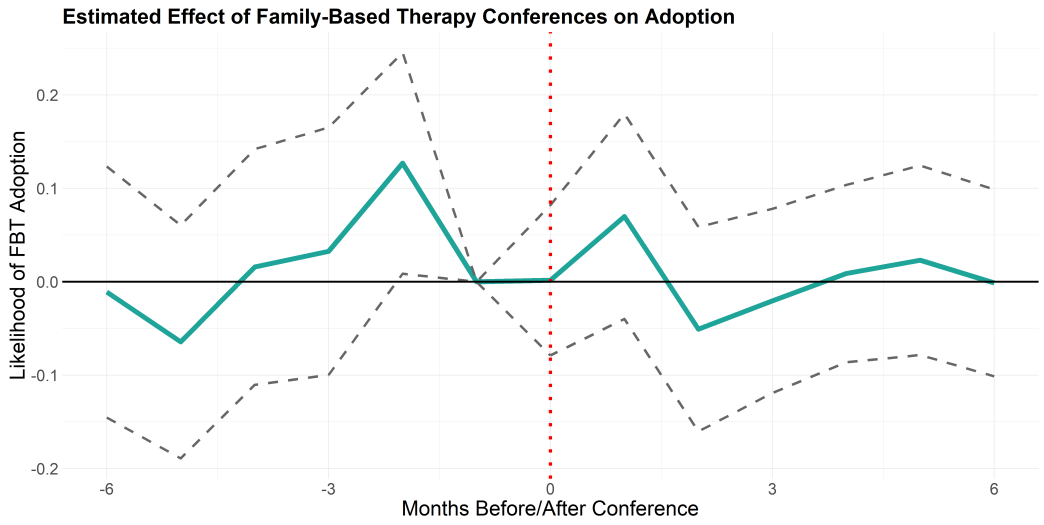


Figure 2·4
Therapist Response to Family-Based Therapy Conferences

response is less dramatic than responses to FBT is somewhat surprising, given my hypothesis about innovation types. I will discuss potential interpretation of these results in Section 2.6.

2.4.1 Robustness

The main results shown above are robust to multiple expressions of the regression specification. In particular, I compared results with binary and continuous dependent variables, the use of all prescriptions (compared to only olanzapine prescriptions), and the decision of whether to normalize the travel costs by specialist salary (as discussed in Appendix B). Figures showing how the estimated coefficients changed based on these varying approaches can be found in Appendix C.

In addition to these typical robustness checks, I also assessed how the results changed relative to my specification for the two mis-measured treatments. My specification uses cutoff thresholds in travel costs to assign treatment status to specialists; however, as discussed in Section 2.3.3, there is a tradeoff between satisfying the conditions of Assumption 3 and maximizing their correlation with the true treatment status (e.g., mitigating concerns of a weak instrument problem). I therefore repeat the estimation procedure using various treatment thresholds, which can also be viewed in Appendix C. The results are quite consistent—if anything, models estimated with more stringent treatment thresholds (smaller η) appear to detect larger estimates, but

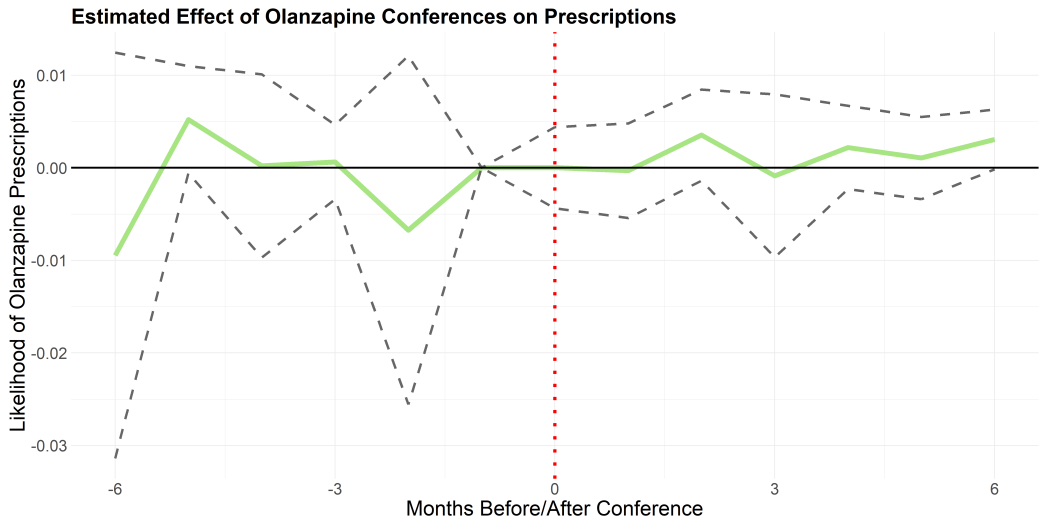


Figure 2·5
Prescriber Response to Olanzapine Conferences

have larger standard errors as well. While future work may elaborate on the optimal decision of treatment threshold to balance the trade off inherent in its selection, this figure provides sufficient evidence that the choice of threshold contributes little to the overall result.

2.4.2 Heterogeneous responses by patient age

While the overall results show little specialist response to professional conferences—whether targetting algorithmic or intangible innovations—a null result may mask interesting heterogeneous responses. To that end, I investigate potentially differentiated responses by patient and specialist type. First, specialists may respond to professional conferences selectively, choosing to implement new techniques on a subset of their patient pool. Particularly, family-based therapies have been shown to be more effective for adolescents and children, for whom family structure is a more integral social context (Couturier et al., 2013). On the other hand, pharmacological interventions may appear more tolerable for adult patients, especially those for whom FBT is not a viable option.

To explore potential heterogeneity along this dimension, I re-estimate the results on the subset of patients who are under 20 years old. Figures 2·6 and 2·7 show the results for the effect of FBT and olanzapine professional conferences on treatment profiles for youth and adolescents. The results for FBT use—a treatment which should

ostensibly be easier to implement among adolescents and youth—are practically identical to those shown in Figure 2·4; however, the results for olanzapine use suggest a small, but more significant, increase in prescriptions for youth following pharmacological conferences. This suggests a certain degree of differentiated response among practitioners based on the type of patients they see, although not in the way one would generally hypothesize.

2.4.3 Heterogeneous responses by specialist type

In addition to potentially heterogeneous response by patient types, specialists themselves may differ in their responses to professional conferences. For example, specialists who hail from a more academic background (e.g., psychologists) may place a higher priority on evidence-based treatments, and may therefore be more likely to integrate FBT or olanzapine into their treatment profiles. To examine this question, I estimate an extended ESIV model using the D-LATE procedure, including interaction terms for specialist types. That is, I examine the specification in Equation 2.6:

$$y_{it} = \alpha_i + \tau_t + \vec{\gamma}\boldsymbol{T}_t + \vec{\delta}\,(\boldsymbol{T}_t \times \boldsymbol{s}_i) + \epsilon_{it}, \tag{2.6}$$

where $\boldsymbol{T}_t$ is the vector of relative time dummies used in the event study and $\boldsymbol{s}_i$ are the relevant specialty types examined in the regression. Then, the coefficients of interest are contained in the vector $\vec{\delta}$. Recent papers such as Johannesen and Stolper (2017) have used this approach as a simple way to explore potential heterogeneous treatment effects.[17]

To examine heterogeneous takeup of FBT, I compare psychologists and therapists to other mental health clinicians (family practice doctors, mental health facilities professionals, etc.); for olanzapine prescriptions, I compare psychiatrists to non mental-health prescribers (e.g., family practice doctors). Figure 2·8 shows the differentiated response for FBT takeup, while Figure 2·9 shows the same for prescribing. In each figure, the first panel illustrates the overall dynamic treatment effect (the vector γ in Equation 2.6), while the other panels are the relevant parts of the δ vector for each specialist type—therefore, these panels are interpreted as the *relative* differences

[17]Notice that it isn't necessary to include level effects for each specialist type $s_i \in \boldsymbol{s}_i$, as these will be picked up by individual fixed effects (for the large majority of the individuals in the sample who don't switch provider types).

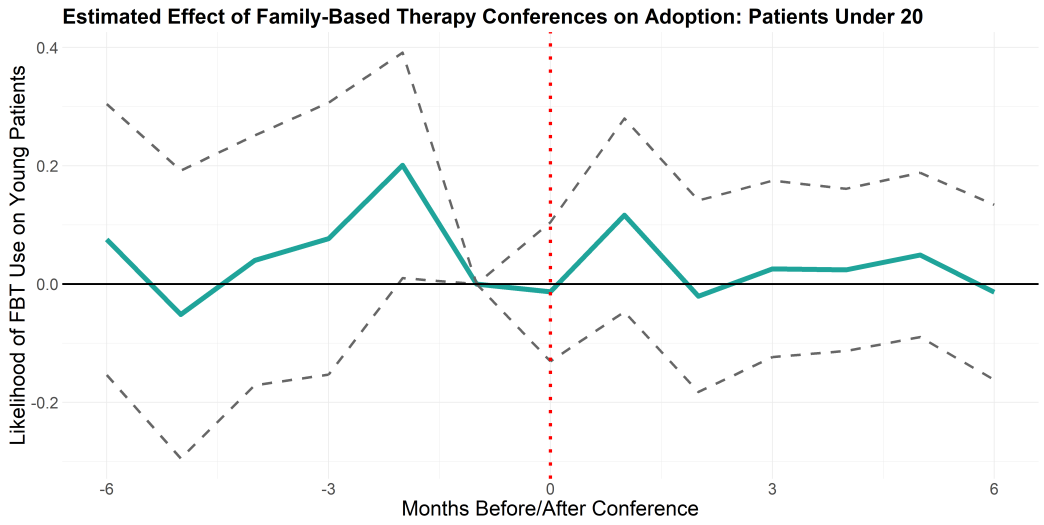


Figure 2·6
Therapist Response to Family-Based Therapy Conferences Among Patients Under 20

in dynamic treatment effects for each group.[18] Psychologists in general appear to have a much higher fluctuation in the use of family-based therapy, but tend to use it overall more than their counterparts. The point estimates suggest a higher positive reaction to the use of FBT for them, but the pre-trends and large standard errors prevent any definitive conclusions. Other mental health professionals (including therapists) appear to have a more subdued response to professional conferences on family therapies.

The results for heterogeneity among prescribers are equally interesting. These estimates have greater power issues than others in this paper due to a smaller group of treated physicians. However, there is still a clear heterogeneous response among prescribers: mental health professionals who are *not* psychiatrists tend to respond positively to these conferences, with a small but significant (and lasting) increase in olanzapine prescriptions following the conference. Other prescribers show a less noticeable change in behavior; general practitioners do not respond at all, and psychiatrists respond for only a few periods following the conference. It may be that psychiatrists are better trained in understanding the risks of a pharmacological approach, or they may have more of an availability to engage in a psychotherapeutic

[18]If one wanted to construct the dynamic treatment effect for psychologists, say, one would add the γ vector to the δ_{psych} vector. The standard errors would stay the same as those around $\delta_{\text{psych},t}$ for all points as this they are bootstrapped standard errors, which do not change under a linear shift.

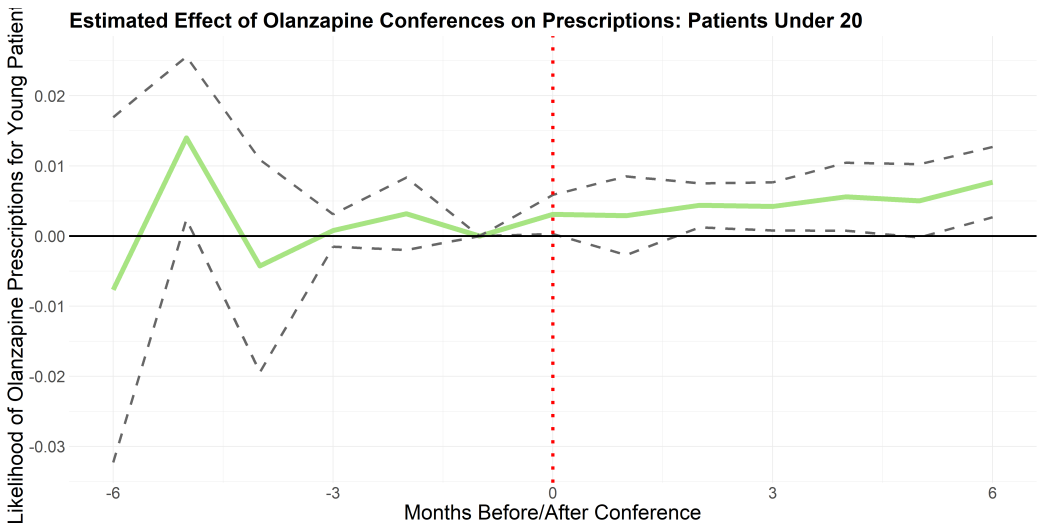


Figure 2·7
Prescriber Response to Olanzapine Conferences Among Patients Under 20

intervention than another mental health prescriber (e.g., a psychiatric nurse practitioner). To the extent that either of these are true, professional conferences may reach those who have less time for therapeutic responses, a higher tolerance for pharmacological risk, or both.

2.4.4 Experimentation as a possible mechanism

Overall, the results suggest a limited and short-lived response to professional conferences. One potential explanation for this fact is that specialists return from conferences and experiment with new techniques, gauging their overall effectiveness and ease of use before integrating them into their regular treatment profile. But therapists who attempt to incorporate FBT, for example, may dislike the increased coordination cost or have a poor first experience with the treatment, which may lead them to revert to their original treatment methods.

To test this hypothesis, I explore the effect of these professional conferences on a specialist's likelihood to expand their treatment set. I measure this likelihood by computing each therapist's Herfindahl-Hirschman index (HHI) of their treatment profile, as measured by variation in their billed CPT-4 codes. The HHI is calculated for each therapist i in period t using the formula:

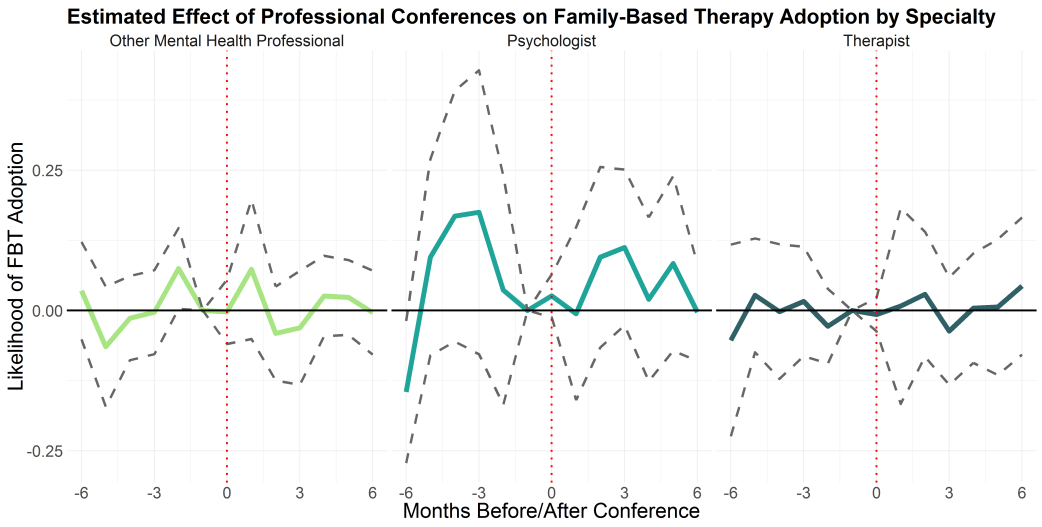


Figure 2·8
Professional Response to Family-Based Therapy Conferences by Specialty

$$\text{HHI}_{it} = \sum_{j=1}^{n} s_{ijt}^2, \tag{2.7}$$

where s_{ijt} represents the fraction of provider i's treatments in time period t that are in the category j. I calculate the HHI using 9 categories, including individual, group, and family therapies amidst pharmacological interventions and other medical and administrative claims.[19]

To the extent that different CPT-4 codes perfectly capture differences in utilized treatment,[20] this provides one measure of how specialized a therapist's treatments are. For example, if a therapist specializes exclusively in family-based therapies, there will be no variation in the treatment profile, leading to an HHI of 1; on the other hand, experimentation with different treatment methods will cause the HHI to *decrease*.

[19]For reference, the 9 categories used are individual therapy, group therapy, family therapy, pharmacological interventions, evaluation and management, intake procedures, general consultations, hospitalization treatments, and other codes used rarely.

[20]There has been a recent discussion on how well physicians agree on the relevant CPT-4 codes for given treatments (Bentley et al., 2002; King et al., 2001) making this calculation an imperfect proxy of true specialization. However, I believe that (given the categories I've selected) disagreement about billing will be minimized in this case, thus making this a useful measure.

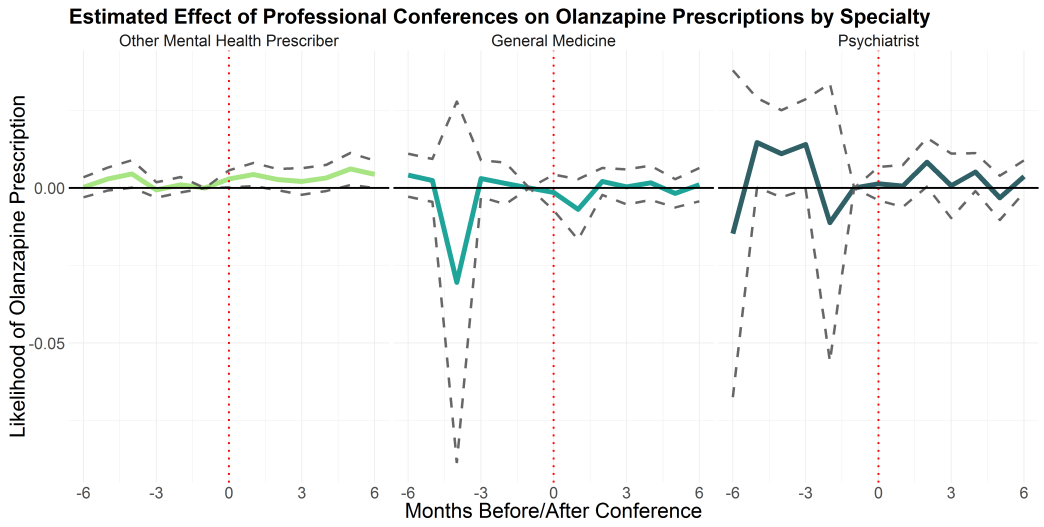


Figure 2·9
Professional Response to Olanzapine Conferences by Specialty

I re-estimate the dynamic treatment effects for these conferences using the calculated HHIs as the new dependent variable. That is, I measure to what extent professional conferences induced specialists to expand (or contract) their treatment methods, inducing experimentation or specialization respectively.

Figure 2·10 shows the effect that FBT conferences have on this measure of specialization. Again, there are no strong results, although there is a slight increase in specialization at the time of the conference (potentially lasting for a few periods). This may result from one of two potential causes: first, the conference itself may impose limitations on a therapist's time for treatment, requiring them to treat only the patients that they are specialized to treat. Second, it may also be the case that continuing education induces therapists to favor their own special skill sets more, as they feel more trained to implement their techniques. Either way, the effects do not suggest an increase in experimentation with new techniques after a conference, which would be indicated by a negative trend. A similar result for pharmaceutical conferences is relegated to Appendix C.

2.5 Travel Cost Validation

Critical to the interpretation of my results is the extent to which the D-LATE estimation technique approximates the true LATE. That is, I would ideally understand

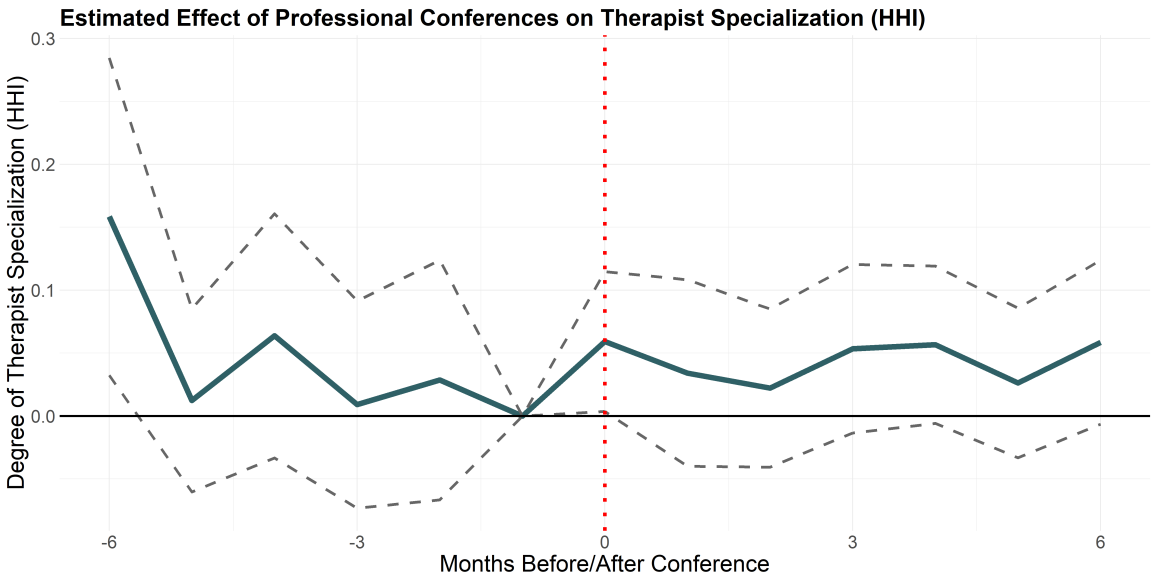


Figure 2·10
Effect of Professional FBT Conferences on Therapist Specailization (HHI)

the probabilities that my treatment/control measures satisfy the conditions in Assumption 3. While I cannot verify this in my sample given the unobservability of true treatment, I have obtained a validation sample of conferences from the Academy of Eating Disorders; this will allow me to obtain a sense of how well these assumptions might be satisfied in my main data.

For now, I have access to registration for the 2019 ICED Conference held in New York City, NY.[21] That is, I have records of each of the 612 unique US-based organizations which sent professionals to the conference, as well as their geocoded locations. Figure 2·11 shows the approximate home location of each attendee, with the conference location shown in red. Notice that, as expected, a large fraction of attendees live in close geographic proximity to the conference location. Interestingly, however, those who travel a greater distance to the conference appear to be based in metropolitan areas, which have greater proximity to an airport and subsequently lower travel costs.

I link this data to my Marketscan data in order to have some idea of a "true" treatment measure relative to a control group. To do so, I identify every therapist in my sample whose main location is within a 10 mile radius of an ICED attendee

[21] I am in the process of widening this validation sample by working with other conference program directors.

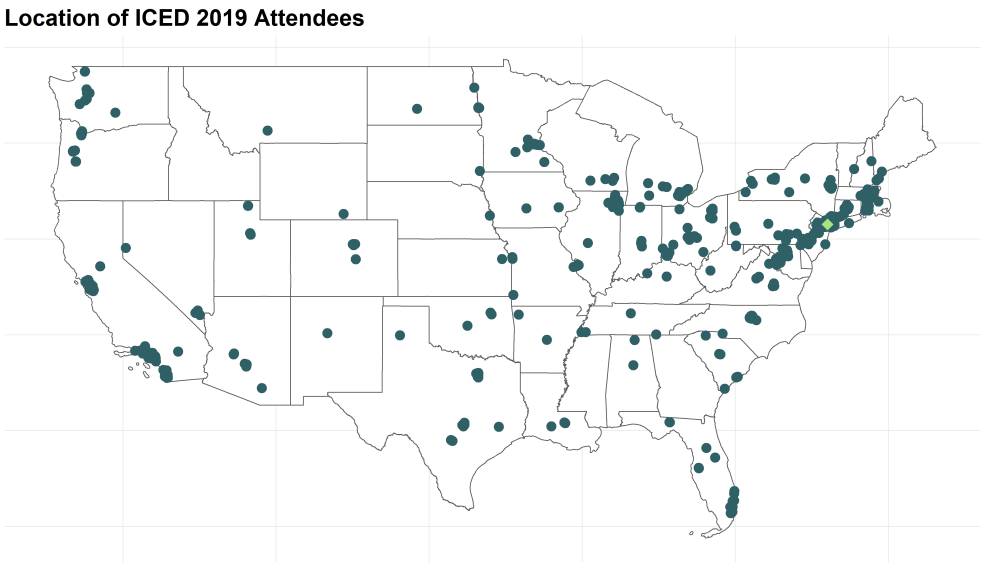


Figure 2·11
Actual Attendees of the 2019 ICED Conference in New York City, NY

location as being "truly" treated. Next, to get an idea of how correlated my predicted treatment measure is with actual attendance, I estimate predicted travel costs for each of the therapists in my sample and the 2019 ICED conference. I then assign treatment groups using the same thresholds used throughout the paper, so that $\eta \in \{1, 2, 5, 10, 15\}$. These are hypothetical, as my sample does not extend to 2019, but will give an idea of how well the prediction algorithm does relative to the truth.

First, I verify the conditions listed in Assumption 3, which are sufficient for the D-LATE estimator to point identify the true LATE. Recall that for the two treatment measures T^a and T^b, one needs to assume that T^a never misclassifies the control group, and that T^b never misclassifies the treatment group. The probabilities of these misclassifications (labelled as p_0^a and p_1^b) are identified for each threshold in Table 2.2 for the unnormed treatment algorithm.

T^a Threshold	p_0^a	T^b Threshold	p_1^b
0.01	0.0000	0.99	0.0074
0.02	0.0000	0.98	0.0086
0.05	0.0172	0.95	0.0103
0.10	0.0645	0.90	0.0360
0.15	0.1155	0.85	0.0406

Table 2.2
Estimated Misclassification Probabilities Using ICED 2019 data (Unnormalized)

In general, these misclassification probabilities decrease with η, suggesting that results using smaller thresholds are closer to the true LATE. In fact, for this particular validation sample, the probability p_0^a decreases to exactly 0 after $\eta < 5$, and the corresponding probability p_1^b decreases to under 1%. This suggests that the conditions in the third assumption are well approximated in my current sample, especially for the smallest two values of η (which include my preferred specification of $\eta = 2$).

Additionally, identification of the D-LATE estimator depends on a nonzero correlation between $T^i, i \in \{a, b\}$ and the true treatment status D. That is, the mismeasured treatments must give some information about true treatment status; without doing so (or with too small of a correlation), a problem similar to that of weak instruments arises. Although this cannot be verified in my sample of interest, I can again utilize the verification sample to assess this correlation. For brevity, the specific correlations are relegated to the Appendix; however, these correlations are strong for most measures and average around 0.15, suggesting little concern of weak identification.

2.5.1 Normed or unnormed?

My travel cost algorithm assigned artificial treatment status based on two types of travel costs: a simple monetary measure (unnormed) and one measured in units of hourly salary (normed by salary). Thus, a simple question is to ask which of these measures best satisfies the assumptions needed for the D-LATE estimator to be meaningfully interpreted. In contrast to Table 2.2, Table 2.3 shows the misclassification measures the predicted travel cost for each therapist as a multiple of their expected hourly salary.

T^a Threshold	p_0^a	T^b Threshold	p_1^b
0.01	0.0004	0.99	0.0149
0.02	0.0062	0.98	0.0257
0.05	0.0315	0.95	0.0389
0.10	0.0645	0.90	0.0691
0.15	0.1019	0.85	0.0915

Table 2.3
Estimated Misclassification Probabilities Using ICED 2019 Data (Normalized)

Overall, the algorithm performs significantly more poorly when normalizing by salary than when using a simple monetary measure. This is additionally advantageous because—as discussed in Appendix B—the normed travel cost measure appears to

over-assign treatment status to those in the sample who make higher salaries (e.g, psychiatrists and family practice doctors) over those who stand to benefit the most from the professional conferences (e.g., therapists and mental health facility workers). Given both of these results, results using the non-normalized treatment algorithm should be taken as closer to the true LATE of interest.

2.6 Discussion & Conclusion

The methodology outlined above and the results arising from its application each have novel implications. In general, my project identifies ways that researchers can augment limited data with powerful statistical learning techniques to answer a broader range of questions than currently accessible, as illustrated by my analysis of professional conferences and mental health treatment behaviors.

2.6.1 Potential uses of methodology

A strong causal inference project typically requires rich data to be compelling. However, the set of questions researchers ask far eclipses the amount of adequate data available to them. The MR-LATE estimator of Calvi et al. (2019), as well as the D-LATE estimator proposed, discussed, and utilized here, offer ways researchers can incorporate imperfect data into an analysis without crippling it.

The estimator used in my project allows for the point identification of a dynamic local average treatment effect (D-LATE) under relatively mild assumptions. Current statistical learning techniques are more than capable of generating the mismeasured treatment assignments necessary for the estimator, and can use validation samples or other techniques to ensure that the assumptions are met at least approximately. Even in cases where the misclassification does not satisfy the conditions of Assumption 3, the D-LATE estimator can be looked at as a method of *bias reduction*, moving the estimated treatment effects closer to the truth by taking into account the possibility of misclassification.

This paper utilized a rather simple predictive algorithm to estimate treatment status of mental health professionals. Future research could integrate more advanced machine learning models in its place, thereby extending the set of questions the D-LATE estimator can answer. Additionally, future econometric research may explore the instrumental variables approach for event studies introduced here, identify more

properties of the D-LATE estimator, and discuss how the work of Calvi et al. (2019) extends to other commonly used causal identification strategies.

2.6.2 Diffusion of mental health treatments

This paper utilizes the methodology of the D-LATE estimator to allow imperfect data to shed light on an important problem in the healthcare industry: the diffusion of ideas. By examining how professionals respond to continuing medical education covering various types of innovations, I am able to assess to what extent research-to-practice gaps are developing in mental health treatments.

The results are suggestive that therapists respond more to tangible innovations than to intangible ones. While there is no discernible response to conferences covering family-based therapies, there are situations in which providers are seen increasing their prescriptions of olanzapine following professional conferences on the subject. The identified heterogeneity discussed in Section 2.4 corroborates this finding; therapists from strong medical training and academic backgrounds (e.g., psychiatry and psychology) respond more positively to family-based therapy, which has a stronger evidence backing, while eschewing the somewhat weaker development of atypical antipsychotics. Interestingly, this increase in prescriptions occurs more among the younger patient population, despite the fact that adolescents and youth stand to gain the most from a family-based treatment approach rather than a pharmacological one.

Of course, future research is critical to confirming these findings. A crucial step will be extending this research beyond the Marketscan data, moving instead towards a holistic assessment of provider behavior amidst patients of various degrees of insurance coverage. Additionally, it will be important to gauge therapist response among other demographics, including experience, academic training, and clinic type. Replicating this project on a richer data set (such as all-payer claims data) can both confirm the validity of this estimator and its findings as well as identify with greater precision the subset of therapists who respond to professional conferences.

There are important questions outside of this domain that must be answered surrounding the impact of continuing medical education and research-to-practice gaps. This project overlooked the role that referring physicians and other members of the treatment team (e.g., dietitians) play in the decision to incorporate new treatments, either pharmacological or therapeutic. However, it may well be the case that these sidelined parties can induce innovation just as well—or better—than a CME program.

Additionally, it may be useful to examine how provider payment mechanisms, network effects, and insurance coverage all dictate the decision to update or experiment with new treatments.

Finally, additional research may move beyond the communication problem of continuing medical education and into other frictions that exacerbate research-to-practice gaps. One might examine how the evolution of academic medical research may have siloed researchers into their own niche, and how this affects researcher involvement with practitioners at all. Additionally, projects might assess how researchers respond to other forms of media surrounding new treatments, including research articles and magazines. Finally, it will be useful to understand how implementing these new techniques affects the ultimate outcome for patients, especially those being treated by an intangible innovation.

Only by obtaining a more holistic picture of the different frictions and mechanisms can we hope to catch a glimpse at a solution to effectively incentivizing the diffusion of better mental health practices. Similarly, recognizing the manifold characteristics of individual innovations will allow a richer study of the economics of innovation. By doing so, future work can provide real solutions to gaps between academic research and real-world practice, as well as foster more efficient channels of communication in a broad spectrum of policy-oriented fields.

Chapter 3

Innovations and Inequities in Access to Medical Services

3.1 Introduction

Improving the quality of medical treatments has immense economic and social value, through both the economic returns from improved health states and the insurance value associated with reduced population risk.[1] Funding, developing, and disseminating novel medical technologies is one of the most promising ways to improve the return on the high levels of health spending in developed countries (Cutler et al., 2007). On the other hand, novel technologies may exacerbate health inequities, which have persisted for over two centuries across socioeconomic status, race, ethnicity, and other group identifiers (Adler and Rehkopf, 2008). Novel, typically high-cost medical interventions typically exacerbate these inequities, especially during the early years following their adoption (Arcaya and Figueroa, 2017).

Achieving the twin ideals of health innovation and health equity requires understanding the tradeoffs involved in pursuing these aims. Physicians may appropriately respond to improvements in one type of medical treatment by increasing their investments in that form of treatment; this results in well-documented inequities in which patients receive access to novel treatments. What is less apparent is how innovation adoption affects other patients who, rather than seeking out the innovative treatment, continue to vie for other, less-intensive interventions that physicians in their local market are now providing more infrequently. If physician specialization affects the returns to a procedure, medical innovations may lead physicians to reduce their volume of older techniques by more than the relative increase in volume of the new treatment, resulting in a second type of inequity: patients who are crowded-out of

[1]For a discussion of the value of medical innovations, see Murphy and Topel (2006) and Lakdawalla et al. (2017).

specialist interventions altogether. These inequities may be further exacerbated by incorrect perceptions of patient risk, either on the part of the physician, the patient, or the health system more generally.[2]

In this paper I present a model of physician decision-making that characterizes the tradeoff inherent in expanding access to medical innovations at the potential cost of these two dimensions of inequities. In the model, physicians select medical interventions for patients of differing risk levels from one of three treatments: two surgical interventions (a high-intensity and a low-intensity procedure), and standard maintenance care. The model incorporates technological spillovers in the style of Chandra and Staiger (2007), so that the returns to a treatment increase as the physician invests more in that technique (e.g., from learning-by-doing). I then consider the impact of an innovation that increases the average return of the high-intensity procedure. The model highlights that physicians may respond to such an innovation along two distinct margins. First, improvements in the high-intensity technique directly lead to an expansion in its use among intermediate-risk patients who previously selected less intensive interventions. Second, and more surprising, the novel technology generates a movement of high-risk patients out of low-intensity interventions and into maintenance care. This is due to a reduced return of the low-intensity intervention due to lower productivity spillovers, resulting in a set of patients who lose access to surgical interventions entirely as a result of the innovation.

The central insight from the model is that the crowding-out of treatment for high-risk patients may be inequitably borne by patients from certain groups within the population. The composition of patients crowded-out from surgical interventions in the model may differ systematically from the overall patient distribution, especially to the extent that patients of different groups are assigned different levels of surgical appropriateness or risk. Moreover, inequities may be exacerbated when risk is imperfectly observed, and certain groups are incorrectly assigned higher or lower levels of appropriateness for care. I quantify the extent to which measurement error in perceptions of patient risk may increase inequities in access not only to medical interventions, but to specialty care overall.

I then present an empirical test of the predictions of my model. My setting is the development and dissemination of transcather aortic valve replacement (TAVR) surgeries used to treat aortic stenosis in elderly patients in the United States. These

[2]For a broader discussion of the inequitable perceptions of patient risk, see Arkfeld (2021).

minimally invasive procedures changed the scope of aortic stenosis treatments in two key ways: first, the procedure allowed surgeries to be performed on higher-risk patients who were previously deemed too risky for surgery; and second, the procedure could be performed by interventional cardiologists instead of cardiothoracic surgeons alone. The rise of this procedure therefore represented a novel disruption in the practice of interventional cardiologists by bringing in a new procedure that could be used on new patients, and therefore meaningfully changed their practice style.

I use this setting to test my model by estimating how TAVR adoption in local markets led interventional cardiologists to change their provision of other surgical interventions, including percutaneous coronary interventions (PCIs) such as angioplasties. I show that interventional cardiologists who began performing TAVR quickly specialized in the procedure, dedicating up to 20% of their time to the procedures in as little as three years. This specialization included an increased rate of screening patients for the appropriateness of valve replacements, and resulted in higher-risk patients receiving TAVR surgeries. However, this adoption caused a reduction in the volume of PCIs performed locally. Importantly, this shift was due to both the increased share of patients receiving TAVR *and* a shift of higher-risk patients out of interventional care altogether. Finally, I highlight that this exclusion of high-risk patients disproportionately affected patients living in low-income areas, with patients in the bottom 40% of the income distribution being 10 percentage points more likely to lose access to surgical cardiac care than those in the top 60%. These findings are consistent with a systematic misperception of a patient's surgical risk across the income distribution.

The model presented in this paper is the first to provide a framework for considering the equity impacts of health innovations. Hence, this project contributes to both the literature on health innovations and health disparities. Recent work has suggested that changes in the allocation of high-value medical services may reduce racial disparities in care, particularly when those reallocations reduce geographic variation in the provision of services (Chandra et al., 2020). The theoretical framework presented in this paper highlights that while innovations may reduce disparities in the populations directly affected by the innovation, other disruptions in the supply of services also need to be taken into account.

Health disparities have been increasing in recent years, with some groups even experiencing disproportionate decreases in life expectancy as a result (Case and Deaton,

2015; Olshansky et al., 2012). This paper highlights that novel technologies may still exacerbate inequities in access even when the playing field of income is leveled, and particularly that these inequities may spillover into access for other specialty care (Arcaya and Figueroa, 2017). Finally, my results highlight that changes in the provision of one medical service may have unforseen consequences that affect the provision of others. In that regard, my work is related to the spillover effects of health services (Fadlon and Nielsen, 2019; Hoagland, 2022).

The paper proceeds as follows. Section 3.2 describes the adoption of TAVR in more detail, as well as providing an overview of the data used in this project. In Section 3.3, I lay out a model of physician decision-making in the presence of technological spillovers, and analyze how such a model implies a tradeoff between the adoption of novel medical technologies and inequities in who is crowded out from accessing specialty care. The model suggests several empirically testable implications, which I outline in Section 3.4; the results of these analyses are presented in Section 3.5. I then conclude with a discussion of the relevance of these results in Section 3.6.

3.2 Setting and Data

3.2.1 The adoption of TAVR

Transcatheter aortic valve replacement surgery is a minimally-invasive alternative to surgical aortic valve replacement (SAVR); TAVR procedures involve the transfemoral placement of either a balloon-expandable valve or a self-expanding valve instead of an open surgical approach used in SAVR procedures. Numerous randomized trials of TAVR (for both valve types) have indicated that the procedure is either superior or noninferior among patients at intermediate or high risk for mortality from typical surgery (Smith et al., 2011; Adams et al., 2014; Leon et al., 2016) and even among low-risk patients (Mack et al., 2019; Popma et al., 2019). These results led to the first TAVR device (from Edwards-SAPIEN) receiving approval from the United States' Food and Drug Administration's (FDA's) Center for Devices and Radiological Health for patients with severe surgical risk in November 2011 (Dvir et al., 2012). Over time, the procedure's use has been expanded to a wider pool of patients as it has continued to be shown to be noninferior to open surgical methods for patients with lower levels of surgical risk (Nishimura et al., 2014; Falk et al., 2017). As of 2017, more surgical

interventions are performed percutaneously than using the traditional open methods (D'Agostino et al., 2018).

The adoption of TAVR is an ideal setting to study the tradeoffs between innovations and inequities for two reasons. First, the adoption of this novel technology was ultimately market-expanding: the median number of surgical interventions used to treat advanced aortic stenosis in the U.S. increased by roughly 1/3 following the adoption of TAVR, with the number of providers supplying these interventions nearly doubling (see Appendix). This increase in the total addressable market provided strong incentives for physicians to change the style of their practice in order to accommodate the opportunity to reach these patients, similar to the rapid expansion of percutaneous coronary intervention (PCI) as an alternative to coronary artery bypass graft (CABG) surgery (Cutler and Huckman, 2003). Second, TAVR—similar to the adoption of PCI as a substitute for CABG surgeries—disrupted the supply of these procedures. Whereas SAVR procedures are performed only by cardiothoracic surgeons, TAVR procedures are performed by a team of surgeons and interventional cardiologists (Adams et al., 2014).

Importantly, these two types of cardiac specialists receive differentiated training. Specifically, as noted by Huckman and Stern (2022), after completing a medical residency, interventional cardiologists complete three additional years of cardiology fellowship and an additional year of an interventional cardiologist-specific fellowship. On the other hand, cardiac surgeons typically complete a general surgery residency followed by multiple cardiothoracic surgery fellowships, a training program that lasts six to seven years. These unique training paths prepare each type of surgeon to hyper-specialize in different surgical approaches, typically open surgical approaches for cardiothoracic surgeons and percutaneous interventions for interventional cardiologists.

3.2.2 Data

I assess the impact of TAVR adoption on treatment decisions for traditional Medicare patients seeking cardiology care using Medicare fee-for-service (FFS) claims data.[3] These data contain 100% of cardiology inpatient procedures performed by both cardiothoracic surgeons and interventional cardiologists on Medicare patients, and include important information about patient risk and demographics as well as demographic

[3]Note that this data excludes individuals enrolled in Medicare Advantage plans.

information for surgeons. I use data from 2010 to 2017, encompassing the years of TAVR's adoption and rapid diffusion. By 2017, surgeons were performing TAVR at higher volumes than SAVR; in addition, IV cardiologists were involved in over 1/5 of these procedures.[4] The adoption of TAVR, therefore, both expanded the pool of patients eligible for medical intervention and fundamentally changed the composition of the surgical team used to treat these patients.

My main sample includes all Medicare patients with aortic stenosis, including both patients who ultimately sought surgical intervention and those who did not. My final data set includes 9,858,536 unique traditional Medicare patients spanning 2010 to 2017.

	Mean	SD	Min	Max
Patient Demographics				
Age	72.58	11.44	0	115
Female	0.52	0.50	0	1
White	0.86	0.35	0	1
Black	0.10	0.30	0	1
Hispanic	0.02	0.14	0	1
Other Race	0.04	0.20	0	1
Median County Income (all)	$55,621.54	$14,677.75	$13,037	$125,003
Median County Income (age 65 plus)	$39,931.21	$8,814.85	$12,709	$91,242
Clinical Characteristics				
# of Chronic Conditions	4.08	2.96	0	20
CC: Congestive Heart Failure	0.21	0.41	0	1
CC: Diabetes	0.32	0.47	0	1
CC: Hypertension	0.62	0.49	0	1
CC: Stroke	0.05	0.22	0	1
CC: Acute Myocardial Infarction	0.02	0.13	0	1
CC: Lung Disease	0.15	0.35	0	1
Surgical History & Risk				
Any Previous Cardiac Surgery	0.00	0.02	0	1
Any Previous Bypass Surgery	0.00	0.02	0	1
Any Previous Valve Surgery	0.00	0.02	0	1
Any Previous Revascularization	0.00	0.01	0	1
Predicted STS-PROM	0.03	0.02	0	0.47

Table 3.1
Summary Statistics of Aortic Stenosis Patients, 2010–2017

Table Notes: Table shows summary statistics for patients seeking interventional cardiologist care to treat aortic stenosis. $N = 43,414,162$ unique patient years spanning 2010-2017. Income is averaged at the zip code level and reported in 2021 USD. Chronic conditions are identified using the 100% Master Beneficiary Summary File (MBSF) Chronic Conditions segment. Surgical history is identified using the 100% Inpatient FFS Claims file. Predicted patient risk (STS-PROM) is predicted as described in Section 3.4.1.

[4]See Appendix Table C.1 and Figure C·1 for details on TAVR's expansion.

Table 3.1 includes relevant summary information for the patients in my sample. I observe demographic information, including a proxy for income at the zip code level (both for the full zip code and specific to residents 65 and older, in order to better approximate Medicare incomes). I also construct relevant clinical information, including the number of chronic conditions and surgical history, as well as specific diagnostic items using the framework of Ellis et al. (2022). The final row summarizes the predicted surgical risk for patients; this is empirically estimated and discussed further in Section 3.4.1.

3.3 Model

This section presents a model of responses to medical innovations, adapted from Chandra and Staiger (2007). The model highlights both how innovations may have unintended consequences on other margins of treatment and how imperfect perception of patient risk may lead these consequences to exacerbate inequities in access to health services.

3.3.1 A model of treatment choice

Suppose there is a continuum of patients suffering from a single disease. Patients and physicians can select from three possible treatments, indexed by $t \in \{0, 1, 2\}$: preventive maintenance ($t = 0$), low-intensity surgical interventions (e.g., PCIs, $t = 1$), and high-intensity surgical interventions (e.g., surgical valve replacement, $t = 2$).

The patient-specific appropriateness of each procedure depends on a patient risk index θ_{it} for patient i. When observed perfectly, θ_{it} captures both the diagnostic severity of each individual as well as their relative risk associated with an intervention—hence, individuals with lower θ_{it} will be more likely to receive intensive surgical treatments. In practice, θ_{it} is not observable; instead, physicians and patients proxy this risk based on a set of observable characteristics Z_{it}; I discuss this more in Section 3.4.3.

The expected utility of each procedure for a specific patient with characteristics $\{Z_{it}\}_t$ is given by

$$U_{it} = \beta_{it} Z_{it} + \alpha_t P_t + \varepsilon_{it}, t \in \{0, 1, 2\}, \tag{3.1}$$

where P_t represents the fraction of the population receiving treatment t. This expected utility incorporates the potential for productivity spillovers in the style of Chandra and Staiger (2007), captured in the second term of Equation 3.1; this allows for specialization to improve the expected utility of that treatment for the marginal patient (if $\alpha_t > 0$).

Since utilities are assumed to be linear, patients' treatment decisions can be characterized as two-way comparisons at any value of θ_{it}. To simplify these comparisons further, I make the natural assumption that treatment intensity levels are perfectly distributed across θ_{it}; mathematically, this is equivalent to the statement that the (absolute value) of the marginal utility of treatment with respect to patient risk is increasing in treatment intensity.[5] Practically, this means that patients make choices only along one of two margins: a choice between valve replacement and valve support techniques, or a choice between valve support techniques and preventive maintenance. Given this assumption, the surgical risk of patient i can be harmonized into a single univariate measure θ_i.

A patient with characteristics Z_i thus chooses the most intensive treatment ($t = 2$) only if $U_{i2} > U_{i1}$. Over the distribution of characteristics Z_i, the probability that a patient receives valve replacement is given by:

$$\begin{aligned}\Pr\{t = 2\} &= \Pr\{U_{i2} - U_{i1} > 0\} \\ &= \Pr\{(\beta_{i2} - \beta_{i1})Z_i + \alpha_2 P_2 - \alpha_1 P_1 > \varepsilon_{i1} - \varepsilon_{i2}\} \\ &= \Pr\{\beta_{21} Z_i + \alpha_2 P_2 - \alpha_1 P_1 > \varepsilon_{12}\}, \end{aligned} \tag{3.2}$$

and the probability that a patient will choose the intermediate treatment ($t = 1$) is:

$$\begin{aligned}\Pr\{t = 1\} &= \Pr\{U_{i1} - U_{i0} > 0\} \\ &= \Pr\{(\beta_{i1} - \beta_{i0})Z_i + \alpha_1 P_1 - \alpha_0 P_0 > \varepsilon_{i0} - \varepsilon_{i1}\} \\ &= \Pr\{\beta_{10} Z_i + \alpha_{10} P_1 + \alpha_0 P_2 - \alpha_0 > \varepsilon_{10}\}. \end{aligned} \tag{3.3}$$

[5] Or $|\partial U_{i2}/\partial\theta_2| > |\partial U_{i1}/\partial\theta_1| > |\partial U_{i0}/\partial\theta_0|$. Note that in the case where θ_{it} perfectly captures patient appropriateness for treatment, this assumption is not a special case. In practice, when θ_{it} is unobserved, these delineations will be less clear.

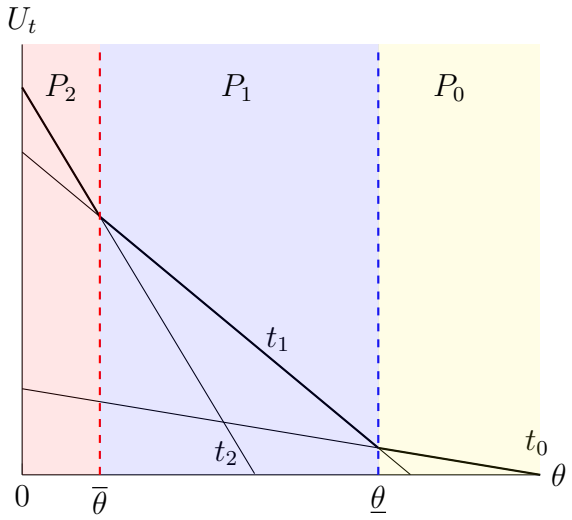
The equilibrium is therefore defined as a fixed point that solves the system of equations:

$$P_1 = \int_Z \Pr\{\beta_{10} Z + \alpha_{10} P_1 + \alpha_0 P_2 - \alpha_0 > \varepsilon_{10}\} f(Z) dZ \tag{3.4}$$

$$P_2 = \int_Z \Pr\{\beta_{21} Z + \alpha_2 P_2 - \alpha_1 P_1 > \varepsilon_{12}\} f(Z) dZ. \tag{3.5}$$

The equilibrium can be conceptualized in a simple single-crossing framework. If an initial allocation is such that all patients are sorted into high-intensity treatments, this will generate utility benefits such that some patients prefer either the lower-intensity intervention or maintenance care. As more patients select out of the highest-intensity intervention, decreases in the survival return to productivity spillovers move more patients out of surgery, until only the most appropriate patients receive high-intensity interventions. A similar market mechanism determines the allocation of low-intensity interventions to patients for whom high-intensity treatments are not justified, but who still receive benefit from medical intervention.

Figure 3·1
Treatment Decisions Based on Patient Risk



Notes: Graphical illustration of the selection of patients into treatment based on risk. The production possibilities frontier for all levels of θ defines the maximum utility for each patient and identifies distinct regions of treatment. Different colored regions indicate the fraction of patients receiving high-intensity treatments (red, defined as P_2); low-intensity treatments (blue, defined as P_1); and maintenance care (yellow, defined as P_0).

Figure 3·1 illustrates the allocation of patients to treatments, based on a perfectly observed patient risk (this assumption will be relaxed in Section 3.3.3. The figure plots patient utility $U_t(\theta)$ for each of the three possible choices of treatment as a function of patient risk θ. As patient risk increases, the utility of each treatment declines; however, by assumption, these decreases occur at faster rates for more intensive treatments. This creates three well-defined regions of treatment, where patients with the lowest risk select the high-intensity intervention t_2, patients with moderate risk select the low-intensity intervention t_1, and the highest risk patients choose to simply receive maintenance care t_0. These regions are defined by the threshold risk levels $\overline{\theta}$ and $\underline{\theta}$; these, combined with the underlying distribution of θ, define the market share of each treatment.

3.3.2 The effect of innovations

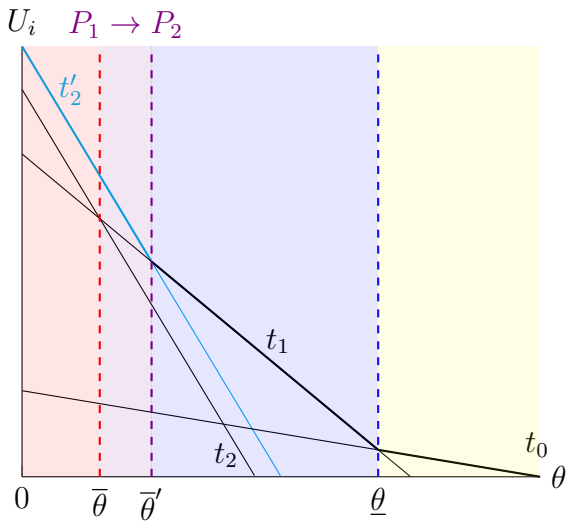
The model allows a direct comparison of how innovations affecting one style of treatment may modify the provision of other treatments. To that end, consider the innovation in valve replacement technology ($t = 2$) reflected in the transition from SAVR to TAVR. This innovation can be characterized as one that reduces the cost of treatment for the patient across all risk levels θ, without affecting the survival utility (based on the noninferiority of TAVR as discussed in Section 1.2). Hence, I model this innovation as a linear increase in the expected utility U_1 by some amount τ.[6]

Figure 3·2 illustrates two separate effects of this shift in U_1. First, panel (a) highlights direct effects of the intervention—as the high-intensity intervention increases patient utility for all levels of risk, a greater share of patients will select the high-intensity intervention over the low-intensity one. This is reflected in a change in the risk threshold between the two interventions, captured by the change in $\overline{\theta}$ to $\overline{\theta}'$ (with the patients switching treatments identified in purple). This increase in the market share of high-intensity interventions results in further increased utility from the intervention and a corresponding decrease in the utility of the low-intensity intervention due to the presence of productivity spillovers in the model. Therefore, second, panel (b) introduces the indirect effects of the innovation's disruption in productivity spillovers. These include both shifts into t_2 as the utility from high-intensity interventions increases further and shifts out of t_1 into t_0 as the return to the low-intensity

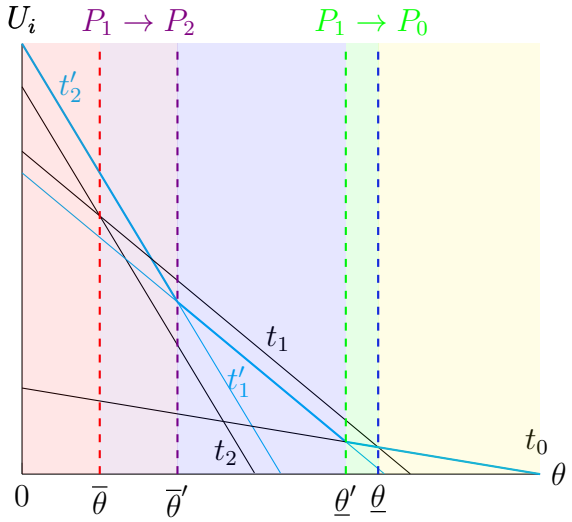
[6]Note that this increase need not be constant across θ, but the resulting implications presented here do not depend on this simplifying assumption.

Figure 3·2
Effects of a Medical Innovation on Treatment Decisions

(a) Direct Effects



(b) Indirect Effects



Notes: Graphical illustration of the effect of an innovation changing the return to high-intensity interventions $U_2(\theta)$. Panel (a) highlights the direct effects of the intervention, which changes the tradeoff between high- and low-intensity interventions and results in a greater share of patients selecting high-intensity interventions (captured in the purple area and the change in $\overline{\theta}$ to $\overline{\theta}'$). Panel (b) highlights the indirect effects of the innovation, which generate changes in the productivity spillovers to both t_2 and t_1, resulting in both a larger share of patients selecting into t_2 and movement from t_1 to t_0 (captured in the green area and the change in $\underline{\theta}$ to $\underline{\theta}$').

intervention decreases. This treatment crowd-out is shown in the change in $\underline{\theta}$ to $\underline{\theta}$' (highlighted in green in the figure). Changes along both margins continue in response to changes in the productivity spillovers affecting the utility from all three treatments until a new equilibrium is reached.

Of particular interest is the magnitude of the shift in $\underline{\theta}$, which defines a share of patients who are crowded-out of (or, depending on the direction of the shift, crowded-*in to*) treatment. To quantify this shift, note that the risk thresholds $\overline{\theta}$ and $\underline{\theta}$ are defined, in expectation over ε, by the equations

$$\beta_2\overline{\theta} + \alpha_2 F(\overline{\theta}) + \tau = \beta_1\overline{\theta} + \alpha_1\left(F(\underline{\theta}) - F(\overline{\theta})\right) \tag{3.6}$$

$$\beta_1\underline{\theta} + \alpha_1\left(F(\underline{\theta}) - F(\overline{\theta})\right) = \beta_0\underline{\theta} + \alpha_0\left(1 - F(\underline{\theta})\right). \tag{3.7}$$

Given the assumptions outlined in Section 3.3.1, the shares P_1 and P_2 can be directly calculated given a distribution for θ, yielding the above result.

Based on this system, I compute the comparative statics of interest measuring how risk thresholds change in response to changes in τ as

$$\frac{\partial\overline{\theta}}{\partial\tau} = \frac{\beta_{10} + (\alpha_0 + \alpha_1)f(\underline{\theta})}{\alpha_1^2 f(\overline{\theta})f(\underline{\theta}) - [\beta_{21} + f(\overline{\theta})(\alpha_1 + \alpha_2)][\beta_{10} + f(\underline{\theta})(\alpha_0 + \alpha_1)]} \tag{3.8}$$

$$\frac{\partial\underline{\theta}}{\partial\tau} = \frac{\alpha_1 f(\overline{\theta})}{\alpha_1^2 f(\overline{\theta})f(\underline{\theta}) - [\beta_{21} + f(\overline{\theta})(\alpha_1 + \alpha_2)][\beta_{10} + f(\underline{\theta})(\alpha_0 + \alpha_1)]}, \tag{3.9}$$

where $\beta_{ij} = \beta_i - \beta_j$ for $i, j \in \{0, 1, 2\}$.

When the innovation is market-expanding, the shift in the extensive margin threshold (Equation 3.9) is nonpositive—meaning that patients are crowded-out from treatment—if and only if

$$\frac{\alpha_1 f(\overline{\theta})}{\beta_{10} + (\alpha_0 + \alpha_1)f(\underline{\theta})} \leq 0 \tag{3.10}$$

$$\Leftrightarrow \underbrace{-\alpha_0 f(\underline{\theta})}_{\partial P_0/\partial\theta} - \underbrace{\alpha_1[f(\underline{\theta}) - f(\overline{\theta})]}_{\partial P_1/\partial\theta} \geq \beta_1 - \beta_0. \tag{3.11}$$

The terms on the left side of the inequality in Equation 3.11 represent the reduction in productivity spillovers for both t_0 and t_1 associated with changing the risk thresholds θ, while the right side of the inequality captures the differences in the marginal utility of each treatment. Note that both sides are necessarily negative numbers. Hence, a

market-expanding innovation in the high-intensity treatment will result in a crowding-out of patients receiving any surgical intervention when the relative change in the productivity spillovers between t_1 and t_0 is less than the difference in the marginal utilities between treatments (in absolute value).

Given that changes in the effectiveness of a treatment due solely to provider specialization are estimated to be much smaller than the marginal returns of an effective treatment itself, this condition is likely to be met in many cases (Chandra and Staiger, 2007). A similar condition allows us to conclude that the innovation itself is market-expanding:[7]

$$\frac{\partial \overline{\theta}}{\partial \tau} \geq 0 \tag{3.12}$$

$$\Leftrightarrow -(\alpha_1 + \alpha_2) f(\overline{\theta}) \geq \beta_2 - \beta_1 \tag{3.13}$$

That is, innovations can easily be seen to be market expanding when the spillovers arising from specialization do not swamp differences in the marginal returns across treatments.[8]

3.3.3 Exacerbating inequities

The crowding-out of lower-value medical interventions due to provider specialization may directly contribute to inequities in who has access to care. In this section, I relax the assumption that patient risk is perfectly observed, and instead assume that risk is proxied based on observable demographic and clinical information. This highlights two features of the crowd-out induced by innovations: first, when risk is *correctly* proxied based on observable demographic information (e.g., income, socioeconomic status, race-ethnicity, sex and gender identity, or sexuality), certain groups may be more likely to be crowded out of care. Second, and pivotally, this inequity may be further exacerbated by *incorrect* rules for assigning patient risk, leaving some groups

[7]Note that Equation 3.13 is simplified by assuming that the extensive margin change is also negative; the full condition—which has the same intuition, albeit less clearly visible—is presented in the Appendix.

[8]Note that it is possible for an innovation to be market-contracting in the model; however, this requires that the productivity spillovers from the low-intensity treatment be so high that any perturbation in $\overline{\theta}$ leads to patients sorting back into t_1 from both t_0 and t_2. This is an unrealistic scenario in practice.

without access to even low-intensity medical interventions despite true underlying medical appropriateness.

Throughout what follows, assume that the condition for crowd-out is satisfied (Equation 3.11), so that there is a region C of patients who received low-intensity interventions prior to the innovation and no medical intervention after its adoption. C is therefore defined by the region of the (true) patient risk distribution on the interval $[\underline{\theta}, \underline{\theta}']$. I suppose that medical care professionals do not observe θ directly but are presented with a proxy for risk $\hat{\theta}$.[9] I assume that $\hat{\theta}$ is a linear combination of observable characteristics Z_{it}, and that it correctly predicts θ except for an idiosyncratic, mean-zero error ε:

$$\theta_{it} = \underbrace{Z_{it}\beta}_{\hat{\theta}} + \varepsilon_{it}. \tag{3.14}$$

Suppose that among the variables contained in Z_{it}, there is a binary variable d_{ig} which is equal to 1 if patient i is a member of a group g, and 0 otherwise. Note that this general form encompasses many different scenarios, including both demographic groups (e.g., patient race or socioeconomic status) and clinical indicators (e.g., patients with diabetes, high BMI, or smokers).[10] The coefficient β_d used in translating d_{ig} to risk captures a discrete shift in predicted risk based on group membership. For ease of exposition, I assume throughout this section that d_{ig} is independent to all other, non-group covariates $Z_{-g} = Z_{it} \setminus d_{ig}$.[11]

It is immediately apparent that if group membership is informative in predicting patient risk (meaning that β_d is nonzero), patients will have different likelihoods of having lower-intensity treatment crowded-out based solely on their group membership. Given information about the underlying distributions of θ and its proxy $Z_{it}\beta$,

[9] Note that this proxy may be the result of physician assessment, patient beliefs, clinical risk information, or some combination of all of these.

[10] Indeed, such indicators, such as patient race, sex/gender, and BMI routinely inform patient risk calculations (van Ryn and Burke, 2000).

[11] Note that this assumption is not critical to the results presented here, but merely simplifies their presentation.

we can identify the fraction of patients in C who belong to g using Bayes' rule:

$$s_{C,g} = Pr(i \in g | i \in C) = Pr(i \in C | i \in g) \frac{Pr(i \in g)}{Pr(i \in C)} \tag{3.15}$$

$$= \frac{s_g}{s_C} \left[Pr(Z_{it,-g}\beta_{-g} + \beta_g \in [\underline{\theta}, \underline{\theta}']] \right] \tag{3.16}$$

$$= \frac{s_g}{s_C} \left[\int_{\underline{\theta} - \beta_d}^{\underline{\theta}' - \beta_d} f(Z_{it,-g}\beta_{i,-g}) d(Z_{it,-g}\beta_{i,-g}) \right] \tag{3.17}$$

$$= s_g \frac{\int_{\underline{\theta} - \beta_d}^{\underline{\theta}' - \beta_d} f(Z_{it,-g}\beta_{i,-g}) d(Z_{it,-g}\beta_{i,-g})}{\int_{\underline{\theta}}^{\underline{\theta}'} f(\theta) d\theta}. \tag{3.18}$$

Here, s_g indicates the relative size of group g in the population, and $s_C = F(\underline{\theta}) - F(\underline{\theta}')$ is the relative size of the crowd-out region. In general, Equation 3.18 does not equal either 0.5 or s_g, meaning that the crowd-out region may be non-representative of membership to g in the overall population. Although this difference arises from true (average) differences in underlying patient risk, such systematic differences in who receives access to care may still have important long-term effects that differ across groups.

Further inequities arise, however, when β_d is not correctly measured. Such imperfect risk proxying may be the direct result of providers who incorrectly gauge the size of risk differences across groups, but may also be the result of other factors, such as patient beliefs or health system measurements such as risk scores, which have been shown to suffer from bias (Obermeyer et al., 2019). However it arises, this measurement error will distort the likelihood that members of group g are represented in the crowd-out region C. To quantify the relationship between measurement error and this inequity, suppose that instead of using β_g in risk calculations, $\hat{\theta}$ relies on the use of a "noisy signal" $\hat{\beta}_g$, defined as

$$\hat{\beta}_g = \beta_g + \nu, \tag{3.19}$$

where ν is an idiosyncratic error in group risk measurement.[12] I define the inequity resulting from the presence of ν as the change in the representation of members of

[12]Note that unlike the classical measurement error readers may immediately associate ν with, this parameter is not random noise (in particular, it is not necessarily centered around 0). In the simplest version of the model, ν is common across provider-patient assessments; however, the model

group g in C, relative to the initial representation $s_{C,g}$. Hence, I define the multiplier increase in members of g represented in C as

$$I(\nu) = \frac{s'_{C,g}(\nu)}{s_{C,g}} \tag{3.20}$$

$$= \frac{1}{s_{C,g}} \int_{\underline{\theta}-\beta_d-\nu}^{\underline{\theta}'-\beta_d-\nu} f(X_{i,-g}\beta_{i,-g}) d(X_{i,-g}\beta_{i,-g}), \tag{3.21}$$

where $s_{C,g}$ is defined as in Equation 3.18. The result in Equation 3.21 follows directly from the calculation in Equations 3.15 to 3.18.

Given information about the parameters governing the initial risk thresholds $\underline{\theta}$ and $\underline{\theta}'$, as well as the distribution of other covariates $X_{i,-g}\beta_{-g}$, the multiplier $I(\nu)$ can be easily calculated. In particular, notice that

$$\frac{\partial I}{\partial \nu} = \frac{1}{s_{C,g}} \left[f_{X_{-g}\beta_{-g}}(\underline{\theta} - \beta_d - \nu) - f_{X_{-g}\beta_{-g}}(\underline{\theta}' - \beta_d - \nu) \right]. \tag{3.22}$$

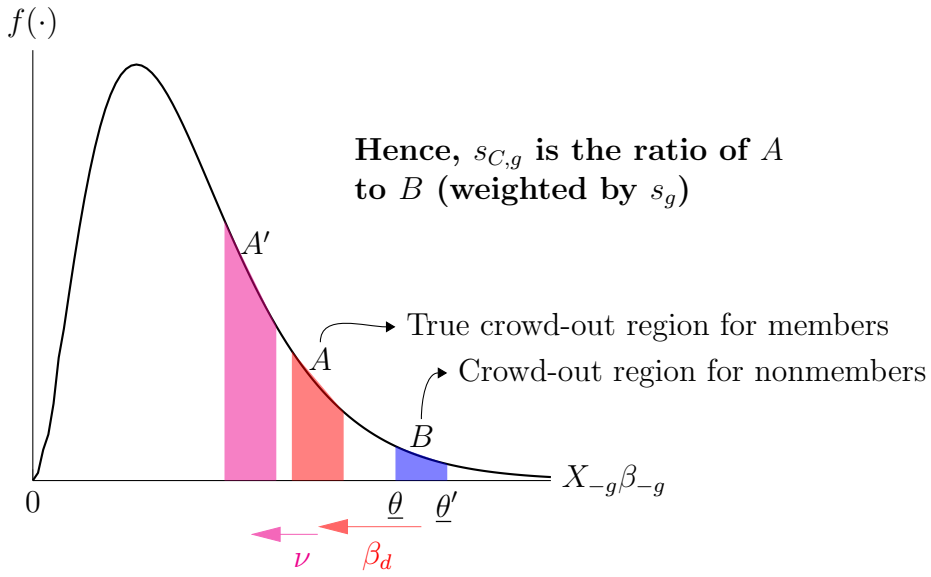
That is, the magnitude of the measurement error ν in $\hat{\beta}_d$ affects the relative crowd-out of members of group g in proportion to (i) the initial composition of g in C and (ii) the relative comparison points used in assessing the risk of nonmembers.

Figure 3·3 captures the intuition behind these inequities. The figure shows, for a given distribution of observable *non-group* characteristics $X_{-g}\beta_{-g}$ and risk cutoffs $\underline{\theta}$ and $\underline{\theta}'$, the regions for which different types of patients will be crowded out of low-intensity interventions by the medical innovation. When the patient is a member of group g, the discrete risk shift β_d results in them being crowded out of treatment when their proxied non-group risk lies in the red region A. Similarly, for patients that are not members of g, the crowd-out region is defined simply by having a proxied risk level $\hat{\theta}_{-g} \in [\underline{\theta}, \underline{\theta}']$ (the blue region B). Hence, the fraction of crowded-out patients in g is given by the ratio of A to B (weighted by s_g).

Intuitively, measurement error in the binary coefficient β_d results in differences in the location of the crowd-out region for the group members. The figure plots the case when $\nu > 0$, or when patient risk for members of the group is overestimated; this shifts the crowd-out region to a pool of lower-risk patients (represented in the figure as the magenta region A'). To the extent that this shift captures a larger

could easily be generalized to allow ν to either vary across providers or patients, as appropriate for the context.

Figure 3·3
Inequities in Crowdout Associated with Imperfect Risk Assessment



Notes: Figure illustrates the relative "crowd-out regions" for members and nonmembers of a group g when used in a proxy for patient risk, as well as the effect of measurement error in β_d on the relative crowd-out rates of members and nonmembers. The figure plots an inverse gamma distribution with parameters $(3, 1)$ for observable non-group covariates used in predicting patient risk, $f(X_{-g}\beta_{-g})$. The figure assumes that the membership variable d_{ig} is independent of all other covariates X_{-g}. The region A (in red) represents the crowd-out region for members of a group g given β_d, and region B (in blue) the corresponding region for nonmembers. Hence, the relative sizes of A and B (weighted by the overall size of the group g in the population) indicate the representation of members of g in the crowd-out region. Changes in ν affecting $\hat{\beta}_d$ shift the region A', ultimately affecting the relative representation of members of group g in the crowd-out region.

share of patients in the population, this will lead to an *over-representation* of group members in the pool of patients who lose access to specialty interventions following the innovation's adoption.

3.3.4 Empirical implications

The central mechanism by which innovation adoptions are linked to health inequities, therefore, consists of two steps. First, when technological spillovers in medical interventions affects physician treatment decisions, innovation adoptions may create "crowd-out regions" that shift patients out of specialty care altogether. Second, the patients in crowd-out regions may be systematically different from the overall population, an inequity that is exacerbated when patient risk is imperfectly observed and incorrectly proxied.

Three empirical implications of interest arise from this model. First, I can directly test for the presence of technological spillovers by assessing the extent to which a medical innovation crowds out the use of other interventions. In particular, given sufficient data on patient risk, I can test whether the development of TAVR affected both the extensive margin risk threshold (e.g., between low-intensity interventions and no intervention) and the intensive margin risk threshold (e.g., between high- and low-intensity interventions). Quantifying the extent to which TAVR's adoption led to decreases in utilization of lower-intensity procedures among the highest-risk patients identifies the existence and magnitude of these crowd-out regions.

Second, in addition to examining heterogeneity in crowd-out across patient risk, I can empirically test the prediction that patients in the crowd-out region will be (potentially inequitably) distributed across members of different groups. The extent to which an innovation leads to differential access to specialized medical interventions may be informed both by true and perceived differences in risk across groups. Nevertheless, given that inequities in access to care may have long-term and/or spillover affects in future health outcomes, even identifying aggregate differences sheds important light on the potential equity problems arising from an innovation's adoption.

The model suggests that I can go one step further to empirically quantify how much of an observed inequity in crowd-out is attributable to true group differences in risk instead of errors in risk proxying. Finally, therefore, I can use the identified model parameters—including both treatment risk thresholds and the distribution of observed patient demographics—to present suggestive evidence on imperfect risk

proxying in selecting cardiac interventions for patients of different groups. These calculations are suggestive as they require strong assumptions about true patient risk, which is unobserved to the econometrician as well as the provider; a further discussion of this is presented in Section 3.4.3.

3.4 Methods

To test the empirical implications of the model, I examine the effects of TAVR's adoption on crowd-out and inequities in access to specialized services at the local level. Although the model is highly stylized and abstracts away from many features complicating physician decision-making, I can test the basic insights of the model by examining how disruptions to the value of surgical intervention (e.g., the adoption of a minimally-invasive technique) altered physician use of closely-related procedures among patients seeking care from interventional cardiologists.

In my empirical exercise, I assess the role of TAVR's adoption in utilization of percutaenous coronary interventions (PCIs) used to treat coronary artery disease (CAD), such as angioplasties and valvuloplasties. Due to the relatively high rate of comorbidity of CAD with aortic stenosis, revascularization surgeries such as PCI were frequently performed on patients whose risk levels made them unfit to receive open surgery for a valve replacement through SAVR. Hence, the adoption of TAVR will have a direct impact on the margin of treatment between a full valve replacement surgery and percutaneous revascularization, particularly when the decision of care is made by an interventional cardiologist.[13]

3.4.1 Estimating patient risk

A patient's risk for cardiac surgery is typically based off of several risk models constructed and maintained by The Society of Thoracic Surgeons (STS). These models account for preoperative factors that may influence a patient's surgical outcomes, and predict patient risks for adverse outcomes such as surgical mortality, permanent stroke, infection, and length of stay, among others (O'Brien et al., 2009).

[13]Note that there is new evidence that PCI can be performed in addition to TAVR in order to treat both CAD and AS (Bajaj et al., 2017; Søndergaard et al., 2019). This evidence comes after the timeframe of my sample, but should be considered in future assessments of this tradeoff.

In my empirical application, I model patient risk θ using the STS Predicted Risk of Mortality model (STS-PROM). This model predicts the likelihood of 30-day surgical mortality following a cardiac surgery using a logistic regression including patient demographics, health conditions, and time trends. Patient demographics include important social determinants of health, including race/ethnicity, gender, and income level (Ash et al., 2017). Health conditions include general counts of chronic conditions as well as finer indicators for specific conditions and symptoms, utilizing the Diagnostic Items framework of Ellis et al. (2022). The full set of covariates used can be found in Appendix Table C.2.

The STS-PROM model is generally used to classify patients into one of three risk categories: low surgical risk (with a risk score $\leq 3\%$), moderate surgical risk (with a risk score between 3% and 8%), and high surgical risk (with a risk score $\geq 8\%$). Patients deemed low risk are those most likely to receive open surgical interventions (e.g., SAVR), while PCI interventions can be performed on intermediate-risk patients as well. There is recent evidence calling into question the effectiveness of using the STS-PROM model as the basis for physician decision-making (Catalano et al., 2020; Khan et al., 2019); however, I continue to use this model as it remains the model most commonly used by practitioners to approximate θ.

3.4.2 Effect of innovation on crowdout

I estimate the causal impact of TAVR adoption on individual interventional cardiologist treatment decisions using two-way fixed effects (TWFE) "event study" regressions of the following form:

$$\Pr(\text{Treatment}_{is}) = \alpha_s + \tau_t + \sum_{k=-T}^{T} \gamma_k \mathbb{1}\left\{t - E_{st} = k\right\} + \epsilon_{st}. \tag{3.23}$$

Here, the outcome variables of interest are treatment decisions for a patient i being seen by interventional cardiologist s.[14] The regression specification controls for both surgeon and time fixed-effects, using quarters as the time unit of interest. Using this specification allows me to estimate a dynamic treatment effect which captures how physician practices evolve in the quarters relative to E_{st}, the surgeon's time of

[14]E.g., s for surgeon.

TAVR adoption. I also adjust for potentially correlated responses within a market by clustering standard errors at the local health market level.[15]

Recent work has highlighted that TWFE estimators can be difficult to interpret without strong modeling assumptions (Callaway and Sant'Anna, 2018). In particular, coefficients estimated by TWFE models represent the weighted average of many two-by-two comparisons. When treatment effects are heterogeneous across groups—and hence, these comparisons—some comparisons may be assigned negative weights (de Chaisemartin and D'Haultfoeuille, 2019; Goodman-Bacon, 2018). This makes the interpretation of estimated treatment effects—static or dynamic—difficult to interpret. In the Appendix, I include robustness checks showing that the results I obtain by estimating Equation 3.23 are robust when using alternative estimators such as those proposed by de Chaisemartin and D'Haultfoeuille (2019) and Sant'Anna and Zhao (2020).

A principal implication of the model is that the adoption of TAVR should have heterogeneous impacts across different values of patient risk; in particular, the model suggests that TAVR's adoption should meaningfully change treatment decisions for patients at both margins of receiving low-intensity care. I therefore estimate potentially hetergeneous treatment effects across the empirically observed distribution of patient risk. I implement two estimators using the methodology of Xie et al. (2012): a parametric estimator that assesses the impact of TAVR adoption across different bins of patient risk; and a nonparametric estimator which this heterogeneity more flexibly.

3.4.3 Inequities in post-innovation access

Finally, I identify how crowd-out inequitably affects groups of differing populations. Throughout, I focus on income-based inequities in access to non-TAVR cardiology services, in keeping with the predictions of the model.[16] These inequities are identified in two stages.

In the first step, I identify the differential probability with which minority individuals are likely to be crowded out of non-TAVR medical interventions. This is done by estimating a version of Equation 3.23 which interacts the dynamic treatment effect coefficients of interest with a dummy variable identifying if a patient resides in

[15]The local market is defined as the commuting zone using relevant U.S. Census data.

[16]Race-based inequities are presented as well in the Appendix.

a county with average income for Medicare patients in the bottom two quintiles of the distribution. I limit attention to patients with estimated risk greater than 5% and use the rate of PCI interventions as my outcome variable. This specification therefore identifies income-based heterogeneity in the likelihood that a patient will be in the crowd-out region between receiving low-intensity medical interventions and maintenance care, as in Equation 3.18.

Under perfect risk perception, the share of low-income patients in this crowd-out region will be proportional to the estimated effect of being low-income on predicted patient risk. Hence, in the second step, I compare the empirically estimated share of patients in the crowd-out region to the share that would be predicted based on the estimated differential risk of being low-income. That is, I reconstruct the estimated share $s'_{C,g}$ using the empirically estimated STS-PROM regression coefficients for income (as shown in Appendix table C.2). I then compare the magnitude of these two shares in order to estimate the extent to which patient risk may be imperfectly proxied by income (the parameter ν in Equation 3.21).

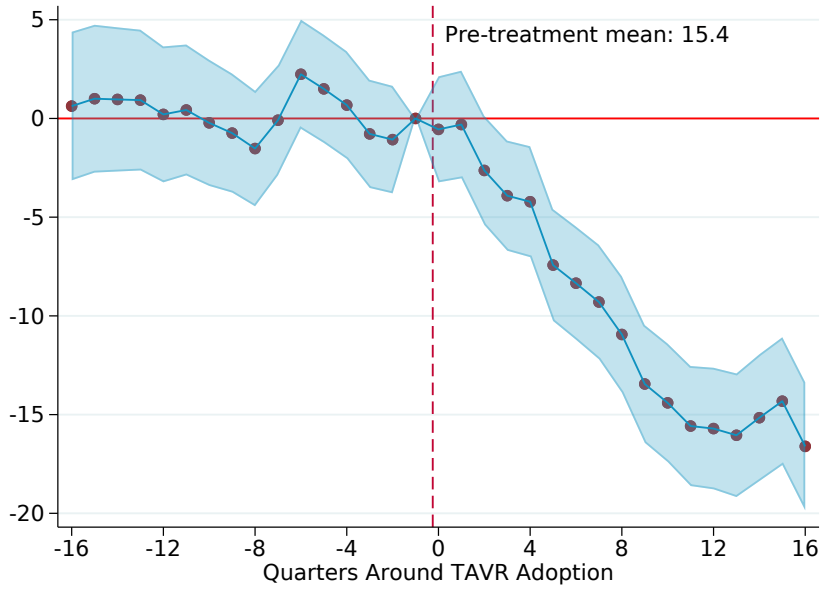
3.5 Empirical Results

The empirical distribution of predicted patient risk closely resembles the population risk distribution predicted by the STS-PROM model (Appendix Figure C·2). The average (median) predicted risk is 3.6% (4.8%). 40% of patients are identified as low-risk, 44% as intermediate-risk, and 15% as high-risk (predicted risk $\geq 8\%$). Patient income is a significant predictor of surgical risk—patients living in the lowest-income counties (measured as the bottom two quintiles of the income distribution) have an expected increase in their risk of surgical mortality of approximately 0.4% ($p < 0.001$).

The adoption of TAVR, as expected, increased both the unconditional likelihood that an individual patient would receive a surgical intervention and the conditional average risk of surgical patients (Appendix Figure C·3). The adoption of this minimally-invasive technique meaningfully expanded the pool of patients eligible for surgery—the adoption was associated with an increase in both the average likelihood an individual would receive surgery (from 1.2% to 1.8%) and the conditional risk of the

average surgical patient (from about 4% to 6%).[17] Overall, TAVR allowed surgeries to be performed on older, higher-risk patients.

Figure 3·4
Effect of TAVR Adoption on Total IVC Surgical Volumes, Commuting Zone Level



Notes: Figure shows estimated impact of TAVR adoption on the total volume of surgical interventions performed by IVCs, including all SAVR, TAVR, and PCI procedures. Interventional cardiologists who perform fewer than 10 inpatient surgeries per year are dropped from estimation, and standard errors are clustered at the commuting zone level. *Abbreviations:* IVC = Interventional Cardiologist

Figure 3·4 shows the estimated impact of this shift on surgical volume. The figure shows event study regression coefficients estimating the impact of TAVR adoption on the total volume of procedures performed by interventional cardiologists—including both high-intensity valve replacement surgeries and percutaneous procedures— at the local market (commuting zone) level. Here, the effect of market-level adoption of TAVR (measured as the first TAVR procedure performed in the commuting zone) is estimated to dramatically reduce the overall surgical volume of IVCs. The average commuting zone performs about 15 such procedures every 3 months prior to adoption; however, within the first 2-3 years following adoption, the total volume is reduced by about 20% for the average CZ. This is driven by a larger reduction in the provision

[17]The Appendix Table further highlights changes in the types of patients receiving surgery following TAVR's adoption in a local market.

of PCI procedures than the corresponding takeup of TAVR, leaving fewer patients receiving surgical treatments in total (see Appendix Figure C·4).[18]

These findings indicate that although TAVR ultimately reached higher-risk patients seeking valve replacement surgeries, fewer patients were treated overall. To examine how adoption creates a crowd-out region of patients who receive neither intervention, I assess how the treatment effect of TAVR adoption on PCI use varies across observable patient risk. Two findings are striking: while the average surgical risk of patients receiving TAVR goes up following adoption (in line with TAVR being a lower-risk procedure than SAVR), there is not a corresponding increase in the average conditional risk of a patient receiving PCI (Appendix Figure C·6). This suggests that there is considerable change in the composition of which patients receive PCI not only at the lower risk threshold, but also at the threshold between PCI and no surgical intervention. Second, the overall likelihood that patients receive any surgical intervention (including TAVR, SAVR, or PCI interventions) declines after TAVR is adopted in the local market (Appendix Figure C·7).

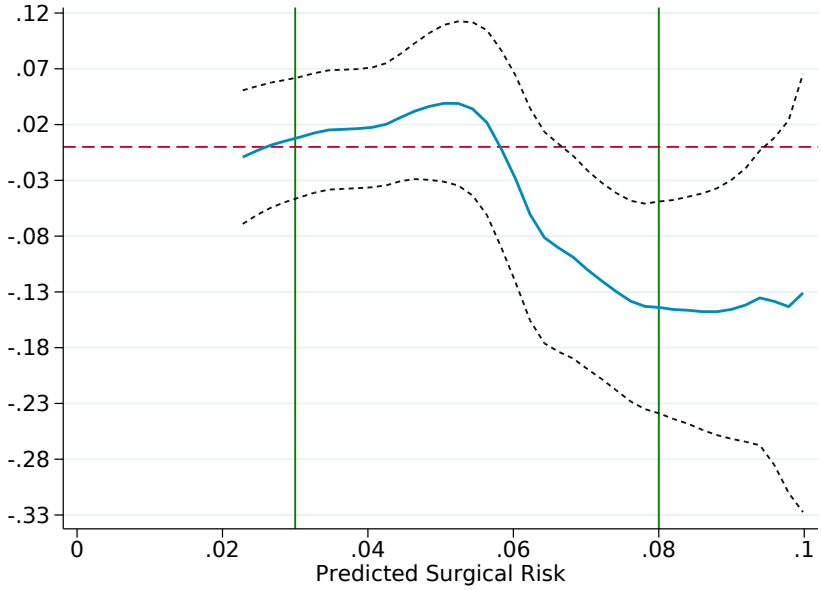
I therefore perform a decomposition analysis in order to assess how TAVR adoption may affect patients in ways that systematically vary with patient risk. Figure 3·5 shows the results, using the method of Xie et al. (2012) in order to estimate a nonparametric relationship between a patient's risk and their likelihood of receiving any surgical intervention. For each level of patient risk, the figure shows an estimated impact of TAVR's adoption on the likelihood that a patient with that risk level will receive any surgical intervention (including both valve replacements and PCIs). Notice that the crowd-out occurs for patients at the boundary between medium- and high-risk, as predicted in the model. In this setting, the region of patient risk where patients lose access to surgical interventions is estimated to be between about 6.5% and 10.0%.

3.5.1 Inequities in access to surgical care

My results indicate that TAVR's adoption led to fewer total surgical interventions for patients in the "crowd-out" region, with patient risks on the boundary between

[18]In the Appendix, I include evidence that shows surgeons also raise their rate of testing for patient appropriateness for valve replacement surgeries. This is in keeping with the recent findings of Mullainathan and Obermeyer (2021). See Figure C·5 for additional results at the individual physician level.

Figure 3·5
Likelihood of Surgical Crowd-out Across Predicted Patient Risk

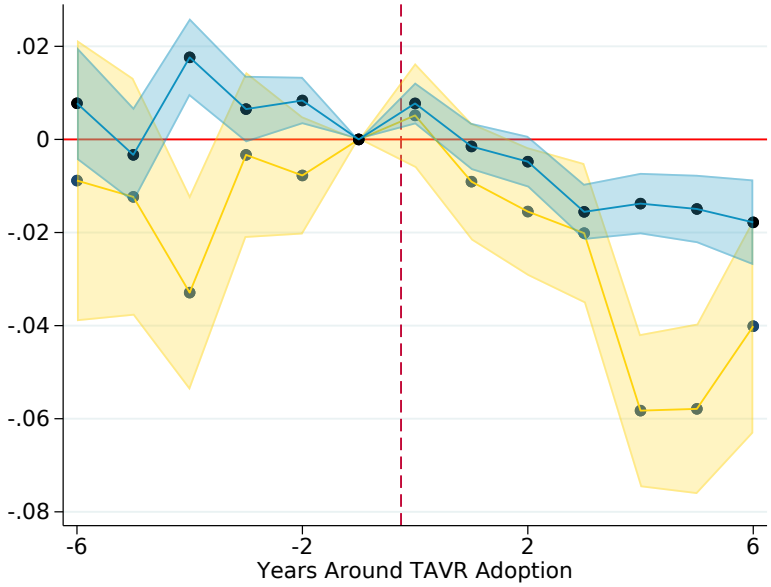


Notes: Figure shows the heterogeneous relationship between the impact of TAVR adoption on the likelihood of receiving cardiac surgery and patient risk. Nonparametric estimator is constructed using the methodology of Xie et al. (2012), and estimated on all patients who have not been diagnosed with aortic stenosis.

intermediate- and high-risk. In this section, I assess the extent to which this crowd-out may exacerbate income-based inequities in accessing cardiac surgeries.

Figure 3·6
Heterogeneous Effect of TAVR Adoption on Cardiac Surgery Use By Patient Income

(a) Total Procedure Volume (SAVR/TAVR + PCI)



Notes: Figure shows estimated impact of TAVR adoption on the total volume of surgical interventions, including SAVR, TAVR, and PCIs. Estimated regressions are shown for two groups—the group of individuals qualifying for either low-income premium subsidies or dually enrolled in Medicaid is shown in gold, while those on traditional Medicare without any such subsidies are shown in gold. In each group, I limit attention to patients in the "crowd-out region," or patients with estimated STS-PROM risk scores in the interval (6.5%, 10.0%). Standard errors are clustered at the commuting zone level.

Figure 3·6 assesses the extent to which TAVR's adoption led to a increased rate of crowd-out for traditional Medicare patients with different levels of income. The figure shows two event study figures assessing the rate at which patients in the crowd-out region (with estimated STS-PROM risk scores between 6.5% and 10.0%) lose access to cardiac surgeries following TAVR adoption for two groups: all Medicare patients receiving premium or copayment subsidies or dually enrolled in Medicaid (shown in gold), and all non-subsidized traditional Medicare patients (shown in blue).[19] Patients with lower incomes are three times as likely to lose access to cardiac surgeries following

[19]Note that about 23% of Medicare patients in my sample are in this "low-income" group.

TAVR's adoption (a difference in total likelihood of receiving surgery of 6 percentage points instead of 2).

This income gap in access to services is an aggregate of two effects: the inequities associated with changes to surgeon risk thresholds for different interventions, and the compounding inequities that arise from incorrect perceptions of patient surgical risk. In the absence of any imperfect risk proxying, I can use the empirically observed distribution of patient risk and the estimated changes in risk threshold to identify the share of low-income patients in the crowd-out region.[20] Based on the empirical distribution of patient risk, the change in the threshold for revascularization from 8% to about 6.5% should have affected low-income individuals only twice as much as high-income individuals. This suggests that approximately 2/3 of the new inequities in access to services arise not from adjustments to physician practice styles alone, but specifically from misperceptions of patient risk.

3.6 Conclusion

In this paper, I present a theoretical framework to consider the potential equity implications of expanding access to novel medical technologies. The model highlights a tension between increased access to a novel technology and overall access to specialized health services when there are returns to physician specialization. Increased provider investment in a high-intensity intervention may result in a "crowd-out region" of patients who lose access to surgical interventions altogether due to the diminished returns of low-intensity procedures. Importantly, the composition of the patients in this crowd-out region may differ systematically from the overall distribution of patients. These inequities in who loses access to medical services are further exacerbated when patient risk is not directly observed and imperfectly proxied.

The predictions of this model can be seen empirically in the diffusion of TAVR among interventional cardiologists. This technology quickly led to an expansion of valve replacements for medium-risk patients, but also reduced the extent to which high-risk patients received less intensive procedures such as PCIs. This loss of access to services among high-risk patients disproportionately fell on patients living in low-income areas, potentially exacerbating geographic and socioeconomic disparities in

[20]Note that this back-of-the-envelope calculation abstracts away from any errors in the STS-PROM model.

access to health care. Back-of-the-envelope calculations suggest that these inequities are magnified due to incorrect measures for patient risk across income groups.

This empirical application highlights the value of using the theoretical framework in considering the general equilibrium effects of innovation diffusion on equitable access to health services. In addition, the findings suggest that corrections to risk prediction models may be far more effective at reducing the inequities associated with TAVR's adoption than adjustments to provider reimbursement for TAVR and other novel procedures. Taken together, the theoretical framework and the empirical exercise suggest that there is room for considering equity implications and potential downstream effects at the time of an innovation's deployment, particularly by large regulators such as CMS.

Future work can build on the central tension highlighted in this paper in several directions. New research may generalize the model to include multiple dimensions of patient risk, or consider further, more long-term consequences of losing access to specialty care. These generalizations may lend themselves well to novel empirical applications using machine learning techniques in novel assessments of patient characteristics and identifying systematic disparities in those assessments across groups (Mullainathan and Obermeyer, 2021). Additionally, future work may identify the extent to which physician selection into innovation adoption affects long-run market outcomes, including equitable access to health services (Huckman and Stern, 2022). Finally, while this project highlighted socioeconomic and geographic disparities exacerbated by medical innovations, this framework can be extended to many other inequities and structural forces that worsen health outcomes for minority patients of many groups. These include a more direct examination of biases and discrimination at the point of care or systematic gaps in seeking out specialty services, either due to coverage or mistrust.

Appendix A

Appendix to Chapter 1

A.1 Data Preparation

This appendix provides detail on sample construction, including the assignment of plan characteristics, health events, and chronic illness costs.

A.1.1 Identifying plan characteristics

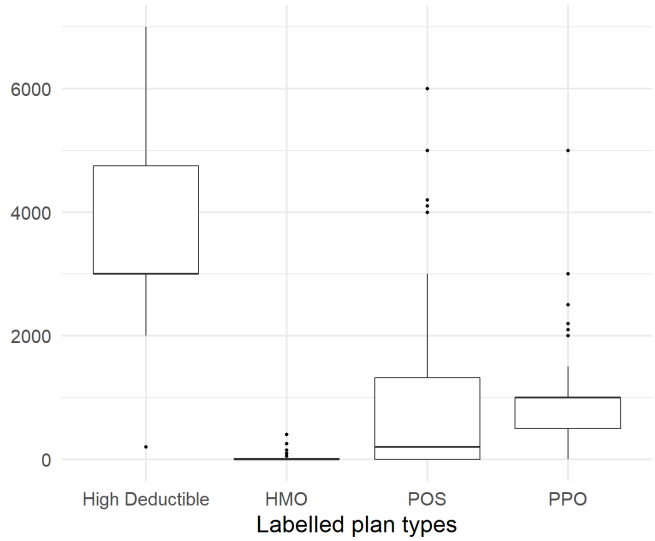
I follow the methodology of Zhang et al. (2018) in inferring individual and household deductibles from the empirical distribution of claims. Given the claims for an individual plan-year, I:

1. Remove claims that are out-of-network, as well as claims with negative values in any of the total paid, plan paid, deductible, and OOP fields.
2. Limit attention to families that had at least 4 consecutive zero-deductible claims after the last positive deductible claim (to ensure that the deductible has really been reached).
3. Calculate each family's total deductible contribution over the year.
4. Estimate the mode and 95th percentile of the deductible within each plan-year.

Figure A·1 illustrates the match quality of these assignments by comparing the distribution of imputed plan family deductibles across listed plan types (Rabideau et al., 2021).

Once deductibles are estimated, average coinsurance rates and out-of-pocket maxima are estimated using the methodology of Marone and Sabety (2021). These cost-sharing parameters are those which minimizes the sum of squared residuals between

Figure A·1
Imputed Family Deductibles by Listed Plan Type



Notes: Box and whisker plot summarizing imputed household deductibles for each listed plan type. Each observation is a plan-year. The box in each boxplot extends from the first quartile to the third quartile of all family deductibles, with a line in the middle for the median. Whiskers extend to 1.5 times the interquartile range (the length of the box) if applicable. All plan years with deductibles outside of the whiskers are shown as outlier points.

predicted and observed out-of-pocket costs, where predicted out-of-pocket costs utilizes the estimated deductible and assumed coinsurance, OOP maximum, and observed spending comes directly from the claims data. This estimation is done separately for each plan-year.

Figure A·2 illustrates the estimated function used in calculating out-of-pocket costs for a given amount of medical spending in a particular insurance plan and year, compared with the realized distribution of total and out-of-pocket spending for all households enrolled in that plan during the year. Each gray dot represents a household, and gold dots are a binscatter plot of all households, using 50 bins. The estimated features of this plan are a family deductible of $2,000, a coinsurance rate of 13.4%, and a family out-of-pocket maximum of $6,750.41.

A.1.2 Identifying major health events

I assign major health events using a set of chronic and acute HCCs, as discussed in Section 2 of the main text. Prior to October 2015, Marketscan claims data relied on ICD-9-CM diagnosis codes, transitioning to ICD-10-CM diagnosis codes thereafter. A

Figure A·2
Inferred Characteristics for a Sample Insurance Plan

Notes: Data shown for a single plan year. Each gray dot corresponds to a single household's observed total and out-of-pocket spending. Gold dots show averages within 50 bins. Blue line illustrates the estimated piece-wise linear function translating observed billed spending into out-of-pocket spending, determined by a plan deductible, coinsurance rate, and out-of-pocket maximum.

table with each major health event as well as its corresponding status (acute/chronic) and accompanying diagnosis codes is available upon request.

When assigning HCCs, I exclude diagnoses associated with the following place of service and procedure codes, due to their high potential for false positive diagnoses, as is done in the HHS-HCC risk adjustment model:

Table A.2 identifies additional demographic information, as well as illustrating the balance in my sample across households with and without a chronic condition in the family. The table also shows the frequency of the various chronic conditions utilized in my sample. Households with chronic conditions are not markedly different in terms of age or sex composition or family size, but do incur significantly higher medical costs in a year. They are not, however, more likely to switch insurance plans from year to year. There is wide variation in the onset of chronic illnesses; the three most common illnesses are asthma, major depressive disorder, and diabetes.

Place of Service Codes		
	12	Private residence home
	31	Skilled nursing facility
	32	Nursing home
	33	Custodial care
	34	Hospice
	41	Ambulance – land
	42	Ambulance – other
	65	Renal dialysis
	81	Independent lab
	99	Unknown
Procedure Codes		
	36415-36416	Drawing blood
	70000-76999	X-ray and ultrasound
	78000-78999	Imaging
	80000-87999	Lab tests
	88000-88099	Autopsy
	88104-88299	Cytopathology
	88300-88399	Surgical Pathology
	88720-88741	In Vivo
	92551-92569	Hearing tests
	93000-93350	ECG and ultrasound
	99000-99001	Specimen handling
	A0021-A0999	Ambulance
	A4206-A999	Medical and surgical supplies
	B4304-B999	Enteral Supplies
	G0001	Drawing blood
	E0100-E9999	Durable medical equipment
	K0001-K9999	Wheelchairs and accessories
	L0100-L4599	Orthotics
	L5000-L9900	Prosthetics
	P2028-P9999	Pathology and Lab
	R0070-R0076	Radiology

Table A.1
Excluded Places and Procedures for Major Health Events

A.1.3 Identifying chronic care costs

An important component of my model is that chronic illnesses correspond to annual diagnostic and maintenance costs that are not strictly choice variables, in the sense that certain health utilization is more or less required. I identify the costs associated with these illnesses from the claims data as procedures which have the major diagnosis listed on that line item. Additionally, in conjunction with Rebecca Hughes, MD, I identify specific therapeutic classes for prescription medications that are associated with treating each chronic condition. Empirical distributions of these estimated diagnostic and maintenance costs for each major health event are available upon request.

	Full Sample	Households with chronic conditions
Demographics & Utilization		
Enrollee age	30.87 (0.008)	29.61 (0.046)
% female enrollees	50.17 (0.000)	50.46 (0.001)
Mean [median] total spending	$2,504.41 [$679.75] (4.510)	$3,378.17 [$957.52] (23.752)
Mean [median] OOP spending	$443.07 [$109.66] (0.525)	$531.93 [$151.18] (3.153)
% switching plans ever	–	
Incidence of chronic illness (per 1,000 individuals)		
Adrenal & pituitary disorders	0.22	7.35
Asthma	2.93	96.08
Breast/prostate cancer	0.35	11.58
Chronic hepatitis	0.10	3.23
Chronic skin condition	0.23	7.46
Congestive heart failure	0.14	4.52
Diabetes with complications	0.39	12.72
Diabetes without complications	1.18	38.57
Fibrosis of lung	0.46	15.10
Heart arrhythmias	0.00	0.00
Inflammatory bowel disease	0.14	4.65
Lupus	0.16	5.20
Major depressive/biploar disorder	1.62	52.76
Multiple sclerosis	1.10	36.17
Personality disorder	0.09	2.81
Rheumatoid arthritis	0.17	5.70
Seizures	0.30	9.82
Thyroid cancer	0.14	4.69
N_{families}	353,403	52,747
$N_{\text{individuals}}$	1,087,353	165,694

Table A.2
Relative Incidence of Chronic Conditions

A.1.4 Identifying cardiovascular preventive medications

Cardiovascular preventive medications are identified using the following set of therapeutic classes. See the review article Albarqouni et al. (2017) for an additional discussion of these classes.

Therapeutic Class	Example Medications
Angiotensin-converting-enzyme (ACE) Inhibitors	benazepril (Lotensin), zofenopril, perindopril
Anticoagulants	warfarin (Coumadin), heparin
Antihyperlipidemic Agents	atorvastatin (Lipitor), fluvastatin, lovastatin
Beta Blockers	propranolol (Inderal), pronethalol
Hypotensive Agents	midodrine (Amatine), norepenephrine

Table A.3
Therapeutic Classes Used in Identifying Cardiovascular Preventive Medications

A.1.5 Identifying low-value services

Low value services are identified at the procedure level using CPT codes for medical procedures and therapeutic classes for prescription medications. I aggregate these services into five broad categories, available upon request.

A.2 Additional Reduced Form Results

A.2.1 Robustness of results to transformations

Table A.5 demonstrates that results are robust to two standard transformations for skewed spending variables: the inverse hyperbolic sine transform, as reported in the main text, and the $\log(y+1)$ transformation.

A.2.2 Robustness of results to event study specification

Table A.4 shows the standard difference-in-differences coefficients for each of the main event study regressions performed in the main text.

I also explore robustness to the problem of negative weights and dynamic treatment effects common in two-way fixed-effects regressions. Implementing the Bacon decomposition of difference-in-differences estimation with variation in treatment timing (Goodman-Bacon et al., 2019) suggests that individuals who experience a chronic diagnosis in the home increase their out-of-pocket spending by 24.6%, more than double the estimates presented in the main text. Additionally, all weighted comparison

Outcome Variable	Treated$_f \times$ Post$_t$	Adusted R^2	N
OOP, chronic, full sample	0.09***	0.51	1,538,162
	(0.012)		
OOP, chronic, zero-deductible plans	0.13***	0.55	390,335
	(0.020)		
OOP, acute, full sample	0.42***	0.50	1,374,481
	(0.031)		
OOP, acute, zero-deductible plans	0.39***	0.54	358,860
	(0.063)		
Billed spending, wellness visits, full sample	0.13***	0.43	1,538,162
	(0.013)		
Billed spending, wellness, zero-deductible plans	0.18***	0.40	390,335
	(0.027)		
Cardiovascular Prescriptions, Prob(fill scrip)	2.56	0.42	439,542
	(1.501)		
Cardiovascular Prescriptions, PDC	1.46	0.48	439,542
	(1.142)		
Billed Spending, Low Value Services	0.06***	0.20	1,538,162
	(0.011)		
Utilization, Low Value Services	0.03***	0.20	1,538,162
	(0.008)		

Notes: This table presents estimates for the standard difference-in-difference coefficients of the event study regressions reported in the paper. Standard errors are clustered at the household level. $^{*}p < 0.05,^{**}p < 0.01,^{***}p < 0.001$

Table A.4

Difference in Differences Coefficients, Main Regressions

	OOP, chronic diagnosis		OOP, acute diagnosis		Wellness spending		Low-value spending	
	$sinh^{-1}(y)$	$log(y+1)$	$sinh^{-1}(y)$	$log(y+1)$	$sinh^{-1}(y)$	$log(y+1)$	$sinh^{-1}(y)$	$log(y+1)$
$t-5$	-0.02	-0.02	-0.11	-0.10	-0.09**	-0.08**	-0.06*	-0.05*
	(0.028)	(0.026)	(0.070)	(0.064)	(0.031)	(0.028)	(0.033)	(0.03)
$t-4$	0.02	0.01	-0.11	-0.10	-0.03	-0.03	-0.04	-0.03
	(0.024)	(0.022)	(0.059)	(0.055)	(0.026)	(0.024)	(0.028)	(0.024)
$t-3$	0.00	0.00	-0.02	-0.02	-0.02	-0.02	-0.03	-0.02
	(0.020)	(0.018)	(0.052)	(0.048)	(0.022)	(0.020)	(0.023)	(0.021)
$t-2$	-0.00	-0.00	-0.07	-0.06	-0.03	-0.03	-0.01	-0.01
	(0.017)	(0.015)	(0.045)	(0.042)	(0.019)	(0.017)	(0.020)	(0.018)
$t-1$	–	–	–	–	–	–	–	–
t	0.08***	0.07***	-0.01	-0.01	0.12***	0.11***	0.05*	0.04*
	(0.014)	(0.013)	(0.041)	(0.037)	(0.016)	(0.015)	(0.018)	(0.016)
$t+1$	0.10***	0.10***	0.10*	0.09*	0.09***	0.08***	0.05**	0.04**
	(0.016)	(0.014)	(0.047)	(0.043)	(0.017)	(0.016)	(0.019)	(0.017)
$t+2$	0.10***	0.09***	0.06	0.07	0.10***	0.10***	0.05*	0.04*
	(0.018)	(0.017)	(0.055)	(0.050)	(0.020)	(0.018)	(0.021)	(0.019)
$t+3$	0.09***	0.08***	0.10	0.09	0.11***	0.10***	0.04	0.04
	(0.018)	(0.019)	(0.062)	(0.057)	(0.022)	(0.020)	(0.024)	(0.021)
$t+4$	0.08***	0.08***	0.14	0.13	0.13***	0.12***	0.09**	0.08**
	(0.025)	(0.022)	(0.074)	(0.068)	(0.025)	(0.023)	(0.028)	(0.024)
$t+5$	0.07***	0.06*	0.12	0.12	0.10***	0.09***	0.12***	0.11***
	(0.030)	(0.028)	(0.088)	(0.081)	(0.030)	(0.027)	(0.033)	(0.029)
R^2	0.51	0.52	0.50	0.51	0.43	0.44	0.20	0.20
N	1,538,161	1,538,161	1,374,359	1,374,359	1,538,161	1,538,161	1,538,161	1,538,161

Notes: This table presents estimates for the main event study regression results reported in the paper. The first column of each pair of results are the results shown graphically in the text, while the second column uses the log transformation. Standard errors are clustered at the household level. $^{*}p < 0.05$,$^{**}p < 0.01$,$^{***}p < 0.001$

Table A.5

Robustness: Inverse Hyperbolic Sine & Log Transformations

groups are estimated to be positive in the primary specification. Furthermore, Table A.6 implements the robust alternative event study estimator described by de Chaisemartin and D'Haultfoeuille (2019) and Sant'Anna and Zhao (2020). Estimations are performed using the appropriate Stata packages (Rios-Avila and Naqvi, 2021; Chaisemartin et al., 2021). The overall ATTs estimated by the doubly-robust method for overall spending responses and prevention spending are 8% and 4%, respectively (Sant'Anna and Zhao, 2020). Figure A·3 illustrates the doubly-robust event study version of Figure 1 in the main text.

	OOP spending, chronic			Billed spending, wellness		
	No Adjustment	CD	SZ	No Adjustment	CD	SZ
t	0.08***	0.06***	0.08***	0.12***	0.11***	0.08***
	(0.014)	(0.013)	(0.014)	(0.016)	(0.022)	(0.22)
$t+1$	0.10***	0.08***	0.10***	0.09***	0.07***	0.05**
	(0.016)	(0.016)	(0.016)	(0.017)	(0.018)	(0.22)
$t+2$	0.10***	0.06***	0.09***	0.10***	0.07***	0.03
	(0.018)	(0.018)	(0.019)	(0.021)	(0.021)	(0.026)
$t+3$	0.09***	0.04**	0.07**	0.11***	0.06***	0.02
	(0.018)	(0.021)	(0.023)	(0.022)	(0.021)	(0.026)
$t+4$	0.08***	0.02	0.05*	0.13***	0.07**	0.07*
	(0.025)	(0.025)	(0.028)	(0.025)	(0.021)	(0.36)
$t+5$	0.07***	-0.02	0.01	0.10***	0.02	0.06
	(0.030)	(0.031)	(0.034)	(0.030)	(0.034)	(0.44)
N	1,538,161	1,538,161	1,538,161	1,538,161	1,538,161	1,538,161

Notes: This table compares regression results from the typical two-way fixed effects event study regression and the robust alternative estimators proposed by de Chaisemartin and D'Haultfoeuille (2019) and Sant'Anna and Zhao (2020). Note that pre-trends are not estimated using the command proposed by Chaisemartin et al. (2021), and are hence not reported). Standard errors clustered at the household level are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.6
Model Comparison: Robust Estimation of Event Studies

As mentioned in the text, the Bacon decomposition suggest that none of the weights used in the typical TWFE regressions are negative. This is illustrated in the following figure.

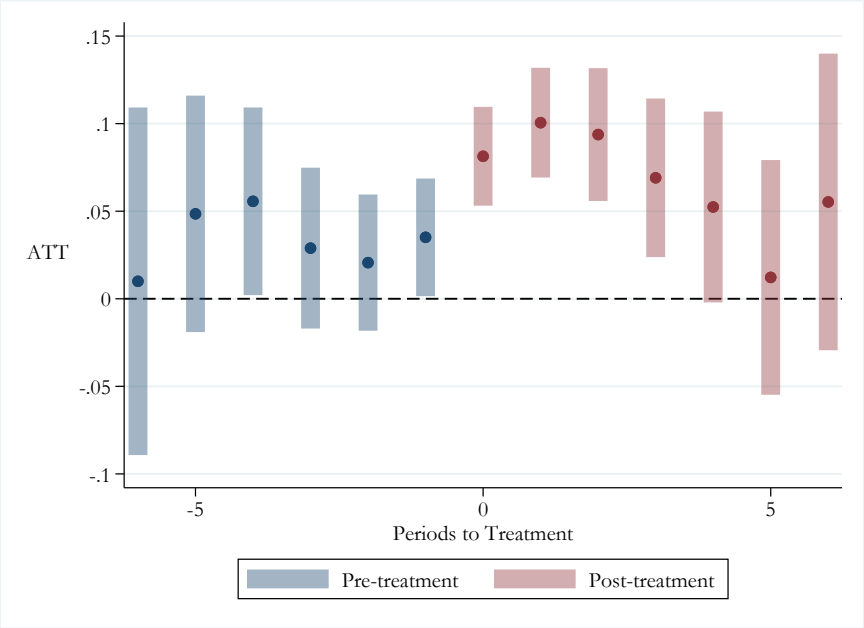


Figure A·3
Effect of Chronic Diagnosis on OOP Spending: Doubly-Robust Estimation of Sant'Anna and Zhao, 2020

Notes: This figure re-presents regression coefficients for the event study regression of Figure 1 in the main text, using the approach of Sant'Anna and Zhao, 2020. Rectangles show estimated average treatment effects and 95% confidence intervals for the effect of a new diagnosis on household OOP spending. Standard errors are clustered at the household level.

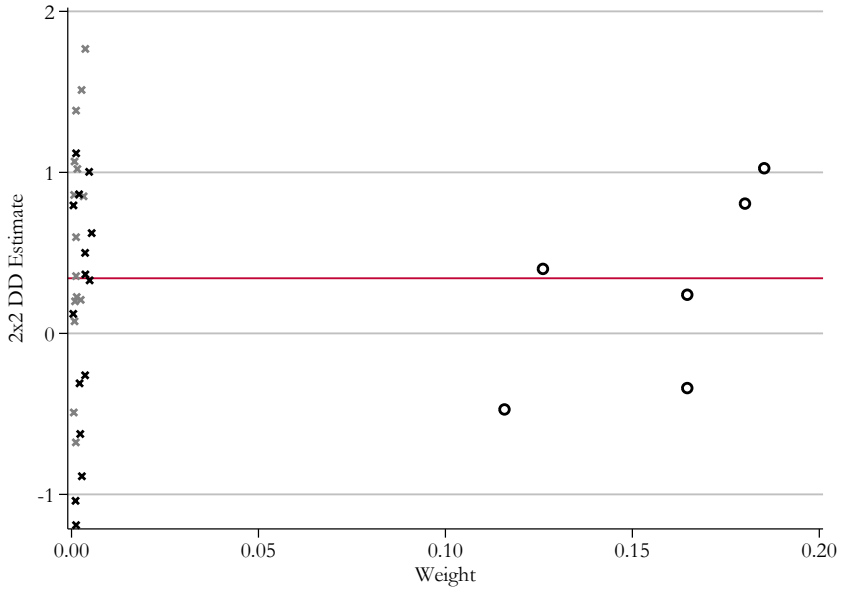


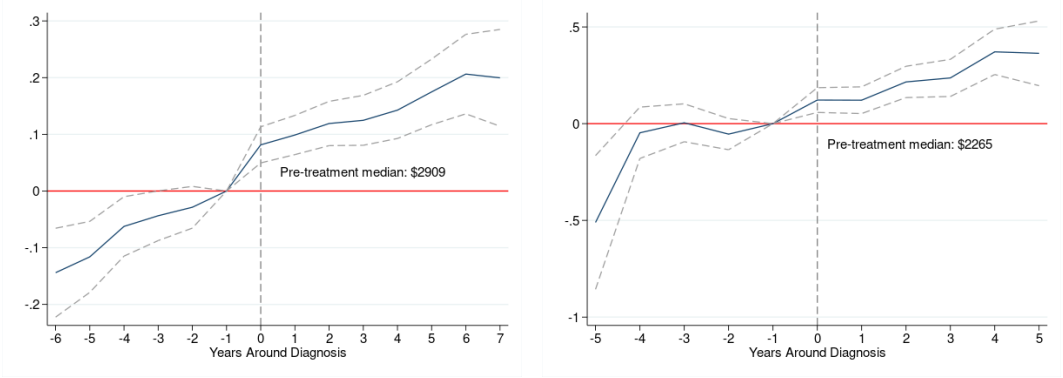
Figure A·4
Bacon Decomposition: Total OOP Following Chronic Diagnosis

Notes: This figure illustrates the estimated decomposition for how individual household-year cells contribute to the overall event study regression of Figure 1 in the main text, using the Bacon Decomposition. Each point represents a single 2x2 regression across a household-period, with its assigned weight shown on the x-axis and the estimated coefficient on the y-axis. All weights are nonnegative, and centered around the overall difference-in-differences coefficient, reported as the horizontal red line. Standard errors are clustered at the household level.

A.2.3 Household response to major medical events

Figure A·5
Estimated Effect of a Chronic Diagnosis on Billed Non-Diagnosed Spending

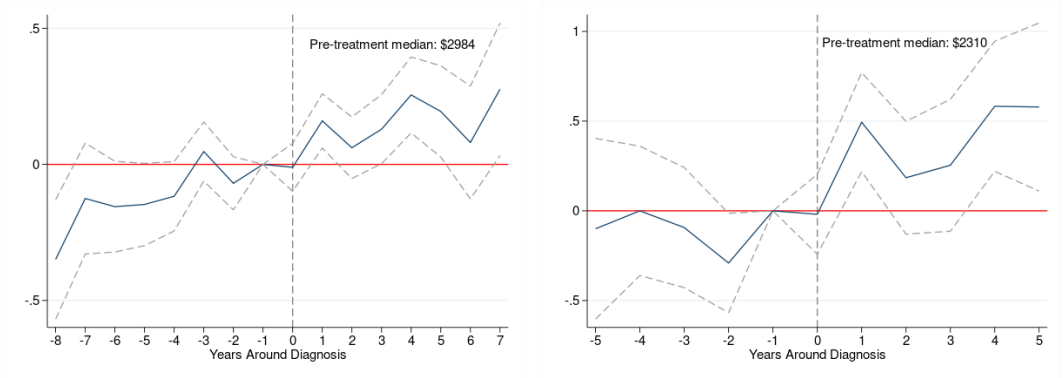
(a) Full sample **(b)** Households facing zero deductible



Note: Dependent variable is the inverse hyperbolic sine of total billed spending for all non-diagnosed individuals in a household. Coefficients are presented relative to the year prior to diagnosis. Spending is measured in 2020 USD. Standard errors are clustered at the household level.

Figure A·6
Estimated Effect of an Acute Health Event on Billed Non-Diagnosed Spending

(a) Full sample **(b)** Households facing zero deductible



Note: Dependent variable is the inverse hyperbolic sine of total billed spending for all non-diagnosed individuals in a household. Coefficients are presented relative to the year prior to diagnosis. Spending is measured in 2020 USD. Standard errors are clustered at the household level.

In this section, I include additional results from a suite of two-way fixed effects models estimating the causal effect of major medical events on health behaviors. Figures A·5

and A·6 illustrate the estimated effect on billed spending for both chronic and acute medical events.

I also explore the effect of acute health events on household out-of-pocket spending, similar to Figure 1 in the text. In general, acute events do not generate the same household response that chronic diagnoses do.

To explore the role that these conditional price changes have on the observed spending responses, I first examine the potentially heterogeneous effects of major medical events by families' typical pre-diagnosis deductible contributions. Figure A·7 illustrates various difference-in-difference estimates for the effect of a major medical event on billed spending, estimated on the sample of families who contributed up to a certain fraction of their deductible on average prior to diagnosis. For this approach, I examine billed spending instead of OOP spending because OOP spending will mechanically rise more for those who tend to have a larger portion of their deductible to pay off, as the deductible is typically the largest contributor to OOP expenses.

The figure shows much larger utilization effects among families that typically spent less than a quarter of their deductible OOP. In fact, families that spent 10% or less of their deductible on average prior to diagnosis are estimated to increase their utilization by about 50%. These large effects decay as more of the sample is included, and I find that even families spending 50% of their deductible may not increase their health utilization following major medical events. Taken together, these results suggest that the families who experience the largest price reductions in care are not the families increasing their utilization the most, suggesting that demand responses are not the major driver of health behavior changes.

Finally, I find a strong extensive margin response among household members who experience major medical events in their families. The following table shows that individuals are more likely to spend any positive amount (billed and OOP) on medical care, use any outpatient visits or preventive care, or fill any prescriptions. This effect is strongest in the year of the diagnosis and decays slightly over time, but remains significant for five years following the health event.

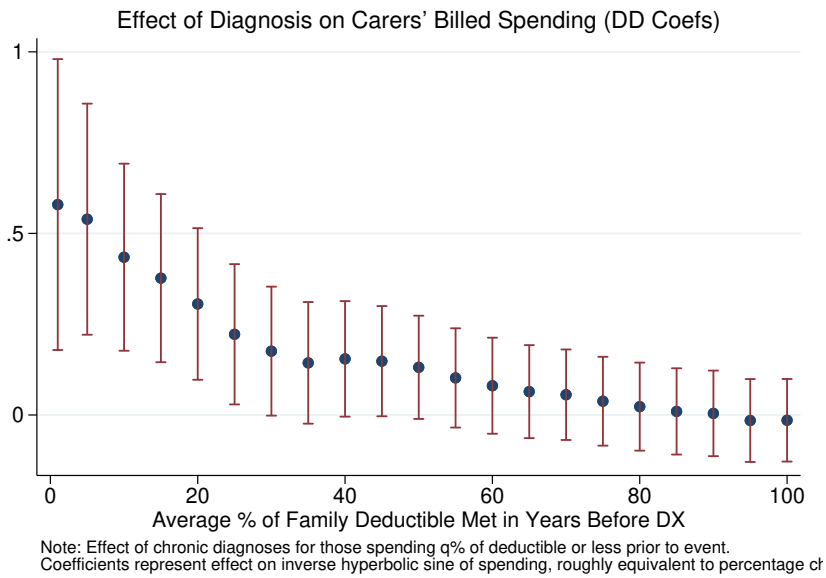


Figure A·7
Spending Responses Differ Based on Pre-Diagnosis Spending

	Year 0	Years 1–5 (average)
Any Billed Spending	1.54***	0.60***
	(0.08)	(0.13)
Any OOP Spending	2.62***	1.41***
	(0.11)	(0.18)
Any Outpatient Visits	2.20***	0.65***
	(0.09)	(0.15)
Any Preventive Care	3.23***	0.90***
	(0.15)	(0.22)
Any Prescription Fills	4.74***	2.45***
	(0.41)	(0.53)

A.2.4 Intra-familial relationships

For example, while a diabetes diagnosis is most likely to affect adult household members with similar lifestyles to the original diagnosed individual,[1] a mental health diagnosis may have a stronger genetic component. Hence, households where an adult was diagnosed with diabetes may choose to screen other adults, such as spouses, while

[1]The vast majority of diabetes diagnoses in my sample are for Type 2 Diabetes Mellitus, which generally affects adults and risk of which is increased or decreased based on specific lifestyle choices, such as diet and exercise. The same is not as true for Type 1 DM diagnoses.

households where someone received a mental health diagnosis may choose to screen children or siblings of the affected individual.

Screening *Diagnosis*	Hypertension *Any Chronic*	Diabetes *Diabetes*	Cholesterol *Diabetes*	High BMI *Diabetes*	Cancer *Cancer*	Depression *MDD/Bipolar*
$\text{Post}_t \times \text{Diagnosis}_f \times \text{Child}_j$	0.39***	-0.85***	-2.20***	-0.38**	**2.55***	**0.30**
	(0.03)	(0.21)	(0.29)	(0.12)	(0.43)	(0.10)
$\text{Post}_t \times \text{Diagnosis}_f \times \text{Parent}_j$	-0.34**	3.49*	3.73	1.73*	-1.90	-0.93***
	(0.11)	(1.71)	(2.26)	(0.70)	(2.49)	(0.13)
$\text{Post}_t \times \text{Diagnosis}_f \times \text{Spouse}_j$	-0.74***	2.54***	5.15***	1.03***	-3.33***	-0.62***
	(0.13)	(0.45)	(0.60)	(0.20)	(0.81)	(0.11)
$\text{Post}_t \times \text{Diagnosis}_f \times \text{Sibling}_j$	0.09	0.76	2.89	0.16	**1.56**	**0.68***
	(0.04)	(1.09)	(1.86)	(0.69)	(1.55)	(0.32)
Observations	4,039,602	3,680,725	3,680,725	3,680,725	3,671,064	3,724,608
Adjusted R^2	0.024	0.217	0.388	-0.025	0.473	0.117

Standard errors in parentheses
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Notes: Table shows results of a difference-in-differences estimation strategy highlighting the potentially differential effects of chronic illnesses on preventive care utilization by household relationships. The primary outcome variable in each column is a screening or new diagnosis, shown in the top row. The specific chronic illness used as the Diagnosis_f dummy is shown in the second row. Standard errors are clustered at the household level.

Table A.7
DDD Estimates: Disease-Specific Spending

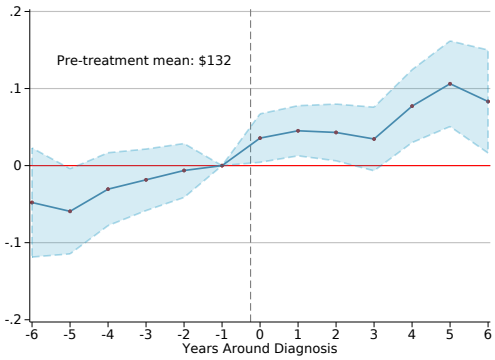
To assess these potentially heterogeneous effects, I utilize a simple difference-in-differences framework. In Table A.7, I present estimation results for the same six diagnosis/outcome pairs shown in Table 3. The dependent variable—either a screening or a new diagnosis—is shown in the top row, with the treatment variable—the chronic illness affecting the household—below in italics. I explore the potentially heterogeneous responses for four family relationships: parents, spouses, siblings, and children of the affected individual, with children as the reference group.

Throughout, I find consistent evidence that households respond by not only selecting screenings associated with the health events they experienced, but also selecting which individuals to screen based on their associated risk. New hypertension diagnoses following a chronic event are concentrated among children rather than parents and spouses, suggesting that households are identifying previously ignored risks among the previously lower-risk members of their household. Additionally, households affected with diabetes focus screenings on spouses more than on children, consistent with the lifestyle factors that affect diabetes risk. In contrast, households affected with chronic illnesses that communicate a greater level of genetic risk—cancer and mental health conditions—choose instead to screen children and siblings (in the case of mental health conditions) more than parents or spouses.

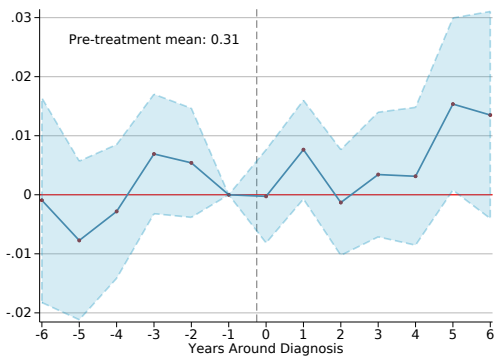
A.2.5 Low value care

Figure A·8
Chronic Diagnoses Increase Utilization of Low-Value Care

(a) Billed Low-Value Spending



(b) Low-Value Utilization (Rate/Person)



Notes: This figure shows estimated coefficients and 95% confidence intervals for the effect of major health events on the use of low-value services (see Appendix A.1 for definitions). In the first panel, the outcome is the inverse hyperbolic sine of billed spending. In the second panel, the outcome is the number of low-value services used per household member. Spending is measured in 2020 USD. Standard errors are clustered at the household level.

Figure A·8 presents estimates for the effect of new chronic diagnoses on the overall utilization of low-value services, including both total spending and overall utilization rates. Major health events are associated with a small increase in overall low-value spending of about 5 percent. In contrast, the average rate of service use among non-diagnosed household members does not change meaningfully following a diagnosis. Table A.8 depicts the event study regressions discussed in the text.

A.2.6 Plan choices

Finally, using the portion of my sample with identifiable plan choice information, I estimate the effect of chronic health events on household decisions to switch plans. Figure A·9 illustrates that affected households are less likely to switch insurance plans following their major health events relative to the general population. I observe both that plan switches do not become more likely overall (Panel (a)), and that even among active choosers, plan switches do not become higher-quality (proxied by the use of zero-deductible plans; see Panel (b)).

Category	All Pediatric		Adult Drugs		Adult Imaging		Adult Screening		Adult Surgery	
Outcome	Spending	Rate	Spending	Rate	Spending	Rate	Spending	Rate	Spending	Rate
DiD										
$\text{Post}_t \times \text{DX}_f$	0.05***	0.02***	-0.00	-0.00	0.03***	0.01***	0.10***	0.03***	-0.10***	-0.04***
	(0.017)	(0.003)	(0.000)	(0.000)	(0.013)	(0.002)	(0.014)	(0.005)	(0.012)	(0.002)
Adjusted R^2	0.192	0.228	0.143	0.259	0.123	0.141	0.163	0.151	0.230	0.255
Event Study										
$t-4$	-0.04**	-0.02*	0.01	0.00*	0.01	-0.00	-0.10***	-0.05***	0.09***	0.03***
	(0.014)	(0.008)	(0.003)	(0.002)	(0.016)	(0.005)	(0.021)	(0.011)	(0.012)	(0.004)
$t-3$	-0.02	-0.01	0.00	0.00	-0.01	-0.01	-0.03	-0.09	0.04***	0.02***
	(0.012)	(0.007)	(0.002)	(0.001)	(0.013)	(0.004)	(0.019)	(0.010)	(0.010)	(0.003)
$t-2$	-0.01	-0.01*	0.00	0.00	0.01	0.00	-0.02	0.00	0.01	0.01**
	(0.010)	(0.005)	(0.002)	(0.001)	(0.016)	(0.004)	(0.016)	(0.010)	(0.009)	(0.002)
$t-1$	–	–	–	–	–	–	–	–	–	–
t	0.02**	0.008	0.00	0.00	0.01	0.01	0.03*	0.008	-0.03***	-0.01***
	(0.009)	(0.004)	(0.002)	(0.001)	(0.010)	(0.003)	(0.015)	(0.008)	(0.008)	(0.002)
$t+1$	0.03***	0.01***	0.00	0.00	0.03***	0.01***	0.07***	0.04***	-0.07***	-0.02***
	(0.009)	(0.005)	(0.002)	(0.001)	(0.011)	(0.003)	(0.015)	(0.008)	(0.009)	(0.003)
$t+2$	0.04***	0.02***	-0.00	0.00	0.02*	0.01**	0.06***	0.02	-0.08***	-0.03***
	(0.010)	(0.005)	(0.002)	(0.00)	(0.012)	(0.003)	(0.016)	(0.009)	(0.011)	(0.003)
$t+3$	0.05***	0.02***	-0.00	-0.00	0.03**	0.02***	0.07***	0.03**	-0.11***	-0.05***
	(0.011)	(0.006)	(0.002)	(0.001)	(0.013)	(0.004)	(0.018)	(0.011)	(0.013)	(0.005)
$t+4$	0.04***	0.02***	0.00	0.00	0.06***	0.02***	0.10***	0.03*	-0.10***	-0.05***
	(0.013)	(0.007)	(0.003)	(0.002)	(0.016)	(0.005)	(0.021)	(0.012)	(0.016)	(0.005)
Adjusted R^2	0.192	0.228	0.143	0.259	0.123	0.141	0.163	0.151	0.230	0.255

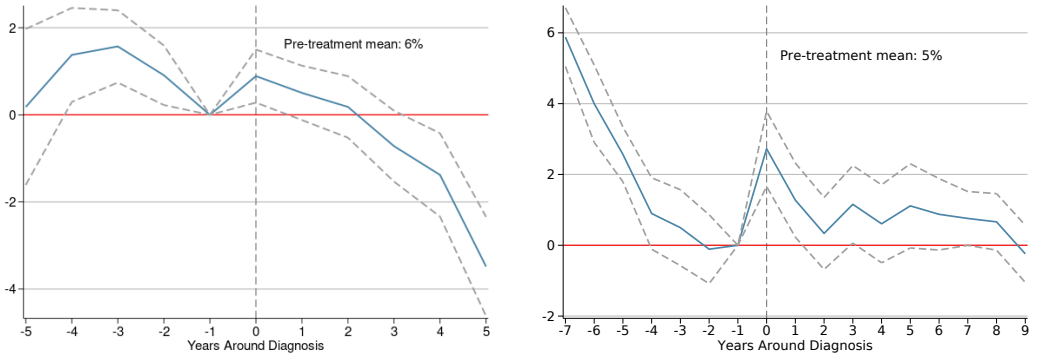
Notes: Table shows estimated difference-in-difference and event study regression coefficients for the effect of a new chronic diagnosis ($N = 1,538,161$). Two outcome variables are reported for each category: the inverse hyperbolic sine of billed spending and the number of low-value services used per household member. See Appendix A.1 for service definitions. Spending is measured in 2020 USD. Standard errors clustered at the household level are reported in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Table A.8

Estimated Effects of Chronic Illness on Low-Value Care Utilization, by Category

Figure A·9
Effect of Chronic Diagnoses on Health Plan Switching

(a) Pr(Any Health Plan Switch) **(b)** Pr(Switch to 0-Deductible—Any Switch)



Note: These figures assess the impact of major health events on plan switches. The outcome variables are a binary indicator for whether the household switched plans in the first panel, and whether they switched plans to a plan with zero-deductible in the second panel. The second panel restricts the sample to those who ever made an active plan choice. Standard errors are clustered at the household level.

A.3 Additional Reduced Form Results

A.3.1 Solving the utility maximization problem

In the final choice stage of the model, households choose medical spending m_{it}^* based on the realization of their acute shocks $\{\lambda_{it}, m_{ft}^{\text{CH}}\}$ and their type parameters $\{p_{it}, \omega\}$. Their expected utility is given by

$$\begin{aligned} u_{it}(m_{it}) = p\left[(\alpha_1 m_{it} + \alpha_2 m_{ft}^{\text{CH}} - \lambda_{it}) - \frac{1}{2\omega}(\alpha_1 m_{it} + \alpha_2 m_{ft}^{\text{CH}} - \lambda_{it})^2 - c_j(m_{it})\right] \\ + (1-p)\left[(m_{it} - \lambda_{it}) - \frac{1}{2\omega}(m_{it} - \lambda_{it})^2 - c_j(m_{it})\right] + \varepsilon_{ijt}. \end{aligned} \tag{A.1}$$

Ignoring the idiosyncratic shock ε_{ijt}, the first order condition for utility maximization implies that optimal spending is given by:

$$m_{it}^* = \frac{1}{1 + p_{it}(\alpha_1 - 1)}\left[\lambda_{it} + \omega(1 - c_j'(m_{it}) + p_{it}\left((\alpha_1 - 1)\omega - \alpha_2 m_{ft}^{\text{CH}}\right)\right]. \tag{A.2}$$

Without the expected utility framework or allowing for state-dependent utility across states, this reduces to the typical solution of $m^*_{it} = \lambda_{it} + \omega(1 - c'_j(m_{it})$. Here, $c'_j(m_{it})$ depends on the optimal level of spending, with $c' = 1$ when households choose a level of spending below the deductible, and then declining to $c' = c < 1$ when OOP spending is between the deductible and the OOP max, and $c' = 0$ otherwise. The piecewise linear structure of the cost-sharing scheme does not yield a closed form solution for m^*_{it}, but rather implies a discrete set of possible solutions that must be evaluated.

A.3.2 Alternate interpretations of p

The evidence presented in Section 3 of the main text suggests that health events generate spending responses as household members reevaluate their health risks. This leads to the simple interpretation of the dynamic learning parameter p_{it} as a probability of an adverse health event occurring. However, to the extent that other informational effects affect spending choices in ways that are separate from health risk information, moral hazard effects, or salience effects, these effects may "load" onto the estimated p_{it} parameter, affecting its interpretation. These informational effects may include physician relationship building, increased comfort obtaining care covered by an insurer, or other, more general health information effects, which alter consumer *preferences* for health care rather than their *beliefs* about risk.

The transition probability parameter p_{it} can therefore be interpreted, in part, as an adjustment to consumer preferences for care in addition to risk beliefs. Consider equation A.1. If we assume that $\alpha_1 \approx 1$, as estimated in Section 5 of the text, the equation reduces to:

$$u_{it}(m_{it}) = m_{it} - \lambda_{it} - c_j(m_{it}) + p_{it}\alpha_2 m^{\text{CH}}_{ft} - \frac{p_{it}}{2\omega}(m_{it} + \alpha_2 m^{\text{CH}}_{ft} - \lambda_{it})^2 - \frac{1 - p_{it}}{2\omega}(m_{it} - \lambda_{it})^2. \tag{A.3}$$

Hence, p_{it} can be construed, together with the estimated parameter α_2, to be representative of the preference weight individuals place on chronic care, relative to all non-chronic care. In this setting, the informational effect of health shocks increases individual preferences for chronic care.

A.4 Additional Reduced Form Results

A.4.1 Estimation algorithm

I estimate the model described in Section 4 of the text using a maximum likelihood approach similar to Train (2009) and Revelt and Train (1998), with the appropriate extension to a discrete/continuous multi-stage choice model as discussed in Dubin and McFadden (1984). My estimation approach is similar to other models like mine, including Marone and Sabety (2021). I estimate the parameter values θ that maximize the probability density of households' observed total healthcare spending conditional on their plan choices. The estimation is done in R version 4.0.3, following the best practices laid out in Conlon and Gortmaker (2020).

My model allows for individuals to have three type-specific dimensions of unobservable heterogeneity, in addition to the typical Type 1 Extreme Value idiosyncratic shock (which can be integrated out analytically): individual health states, individual beliefs about health risks, and household risk aversion. I therefore must numerically integrate over the three dimensions $\beta_{ft} = (p_{it}, \mu_{\lambda,i}, \psi_{ft}) \in \theta$. Given a guess of θ, I use Gaussian quadrature with 27 support points (three in each dimension) to simulate underlying consumer types, yielding simulated points $\{\beta_{fts}(\theta)\}_s$ and weights W_s.

For each simulation draw s, I can then calculate the conditional density at individuals' observed total healthcare spending and the probability of households' observed plan choices.

Household spending

Given data on realized choices m_{it}, I construct the distribution of healthcare spending for each individual-year implied by the model and guess of parameters θ. Based on underlying consumer types β_{fts}, I construct individual-level parameters for health states $(\mu_{\lambda,i}, \sigma_{\lambda,i}, \kappa_i)$ based on the parameters β_{fts} and the distributions outlined in Section 4.3.1 of the text.

The model predicts that given an acute-chronic health state $(\lambda_{it}, m_{ft}^{\text{CH}})$, households choose total healthcare spending m by trading off the benefit of healthcare utilization with its out-of-pocket cost, as discussed above. Given that m_{ft}^{CH} does not have individual parameters to be estimated (as these values are drawn from an empirical distribution), inverting the expression in equation 18 of the text yields the health state realization λ_{its} that would have given rise to observed spending m_{it} given m_{ft}^{CH}.

Given that observed spending is truncated from below at 0, there are two possibilities for the conditional pdf:

$$f_m(m_{it}|c_{jt}, \beta_{fts}, \theta) = \begin{cases} \mathbf{\Phi}\left(\frac{\log(\kappa_i)-\mu_{\lambda,i}}{\sigma_{\lambda,i}}\right) & m_{it} = 0 \\ \mathbf{\Phi}'\left(\frac{\log(\lambda_{its})-\mu_{\lambda,i}}{\sigma_{\lambda,i}}\right) & m_{it} > 0, \end{cases} \tag{A.4}$$

where $\mathbf{\Phi}(\cdot)$ is the standard normal cumulative distribution function. In practice, there are iterations where the implied pdf is zero; hence, in order to rationalize the data for any parameter guess, I use a convolution of f_m with a uniform distribution over the range [-1e-75, 1e-75], as done by Marone and Sabety (2021).

Plan choices

I next calculate choice probabilities for each available health insurance plan. Given θ and β_{fts}, I numerically integrate over the joint distribution of acute and chronic health care shocks using $D = 10$ support points in each dimension. The support points for the chronic health care shocks are chosen uniformly across the empirical distribution with the empirical pdf used in calculating the associated weights. For the acute health shocks, support points are calculated over the lognormal distribution as:

$$\lambda_{itsd} = \exp\left(\mu_{is} + \sigma_{is} Z_d\right) + \kappa_{is}, \tag{A.5}$$

where Z_d is the appropriate Gaussian quadrature vector of points (with corresponding weights W_d). The utility maximization framework discussed above (Equation 18 in the text) is then used to calculate the optimal spending levels given individual and household shocks and the underlying parameter p_{it}. Expected utility for each support point is calculated as in equation 9 of the text and summed (with weights) over all 100 points.[2] Choice probabilities for a plan j are then given by the standard logit formula

$$L_{ftjs} = \frac{\exp(U_{ftjs}/\sigma_\epsilon)}{\sum_{i \in \mathcal{J}_{ft}} \exp(U_{ftis}/\sigma_\epsilon)}. \tag{A.6}$$

[2]In practice, to speed up estimation, I ignore points with associated weights smaller than 1e-5.

Likelihood function

Based on the choice probabilities and conditional density functions for observed spending, the likelihood function is approximated by

$$LL_f = \sum_{j=1}^{J} d_{fjt} \sum_{s=1}^{S} W_s \prod_{t=1}^{T} f_m(m_{it}|c_{jt}, \beta_{fts}, \theta) L_{ftjs}, \tag{A.7}$$

where d_{fjt} is an indicator variable equal to one if household f chose plan j at time t and zero otherwise. The log-likelihood function to be maximized is therefore the sum over households:

$$LL(\theta) = \sum_{f=1}^{F} \log(LL_f). \tag{A.8}$$

A.4.2 Additional parameters

Table A.9 includes additional structural parameters not discussed in the text. These are reported only for the preferred specification of interest (column 3 in Table 5 of the text).

Figure A·10 illustrates the estimated percentage changes in welfare from new health information, the corresponding result to Figure 8 in the text.

Figure A·11 illustrates heterogeneity in household characteristics and the value of new health information.

	(1)	(2)	(3)
Panel A: Mean-shifters			
Initial Probabilities			
Intercept	0.00	-9.91	-10.11
Age	-0.11	1.00	0.48
Age^2	0.32	0.34	0.33
Female	-6.94	5.00	0.50
Individual risk score	-5.12	-1.63	-0.88
Any PE in family	3.01	4.25	0.53
Acute Health Shocks			
Intercept	–	5.00	5.00
Age	–	0.09	0.11
Age^2	–	-0.14	-0.14
Female	–	0.49	0.77
Type	–	-0.59	0.30
Initial Risk Aversion			
Intercept	7.14	10.00	4.68
Family size	-0.10	-7.75	-0.10
Average family age	-0.75	9.27	1.93
Average family risk score	-1.51	-9.87	-4.93
Panel B: Other Parameters			
σ^2_κ (acute health shifter, variance)	–	0.02	10.56
ω (moral hazard shifter)	249.36	146.60	250.00
η (switching costs)	40.13	34.34	23.13
Beliefs Evolve	Yes	Yes	Yes
Acute Shock Heterogeneity		Yes	Yes
Risk Aversion Evolves			Yes

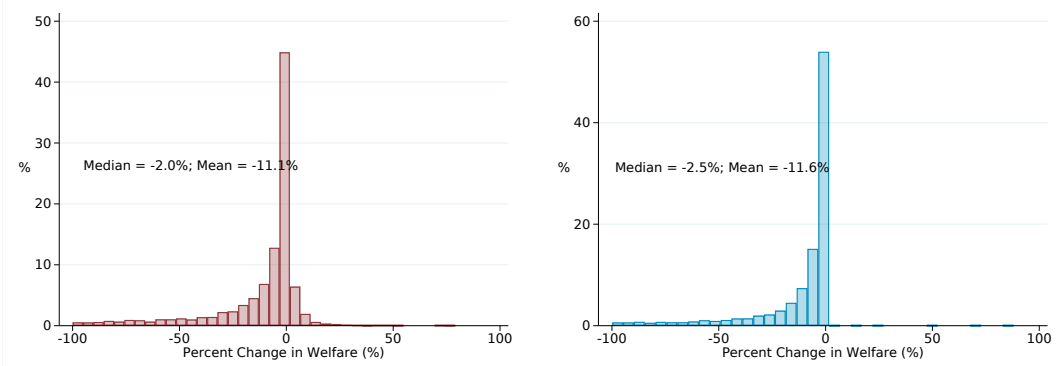
Table A.9

Estimated Type Mean Shifting Parameters

Notes: See Table 1.5 for structural parameters of interest.

Figure A·10
Percentage Changes in Household Welfare Following Health Information

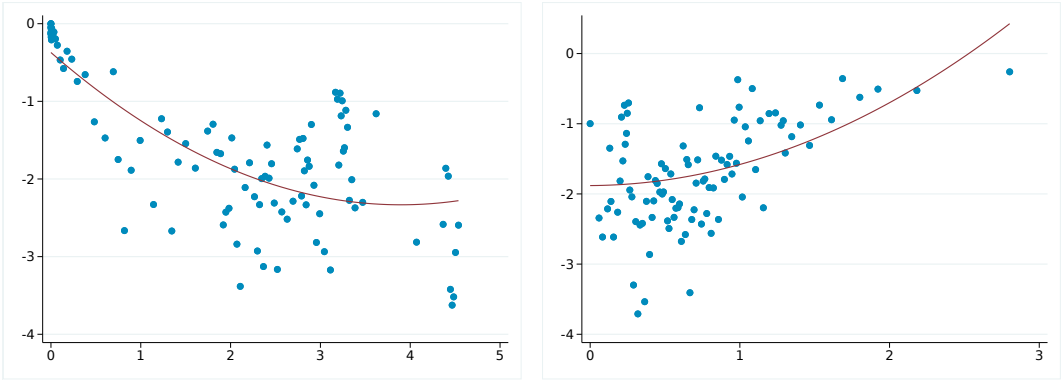
(a) Full Welfare Effect **(b)** Health Information Effect Only



Notes: Figures show estimated percentage changes in household willingness to pay associated with major health events. The panel on the left shows differences in the case of a full response to a new diagnosis, including adjustments to risk aversion and moral hazard effects; the panel on the right shows only differences arising from adjustments to household risk assessments. Welfare effects are calculated in the year of the diagnosis relative to a benchmark in which no information is transmitted.

Figure A·11
Heterogeneity in Household Characteristics and WTP for Health Information

(a) Household Risk Aversion, $\overline{\psi}$ **(b)** Household Risk Scores



Notes: Figures show binscatters depicting the association between pre-diagnosis household health characteristics on the x-axis and the estimated welfare effects of receiving health risk information on the y-axis. Household characteristics include (a) average household risk aversion and (b) average household risk scores (calculated using the Johns Hopkins ACG System). Welfare effects are calculated in the year of the diagnosis relative to a benchmark in which no information is transmitted; see Figure 1·8 for details. Binscatters are constructed using 100 bins and a quadratic fit line.

Appendix B

Appendix to Chapter 2

B.1 Proof of Theorem 1

Proof of Theorem 1. Suppose that we observe a panel $\{Y_{it}\}$ of outcomes for individuals $i = 1, 2, ..., N$ and periods $t = 1, 2, ..., T$. Additionally, suppose that there is a true binary treatment status $D_{it} \in \{0, 1\}$ for all i and t—for consistency with the event study framework, $D = 1$ is an absorbing state, so that $D_{is} = 1$ implies that $D_{it} = 1$ for all $t \geq s$. Additionally, there is a binary instrument $Z_{it} \in \{0, 1\}$ and two mis-measured treatments T^a, T^b satisfying Assumptions 1 and 2 above.

The event study is estimated as follows (using the notation from Sun and Abraham (2021)): for each individual, define the time of first treatment as $E_i = \min\{t : D_{it} = 1\}$ and the related treatment-time dummy variables $D_{it}^{\ell} = \mathbb{1}\{t - E_i = \ell\}$. The regression equation is as in 3.23 with D_{it}^{ℓ} in place of the indicator variables.

As the true treatment is unobserved, we are interested in the local cohort-specific ATE of a transformed regression of $T^j Y$ on T^j for $T^j \in \{T^a, T^b\}$. That is, we are interested in the vector $\vec{\gamma}$ resulting from estimation of:

$$T_{it}^j Y_{it} = \alpha_i + \tau_t + \sum_{\ell=-K}^{-2} \gamma_\ell T_{it,j}^{\ell} + \sum_{\ell=0}^{L} \gamma_\ell T_{it,j}^{\ell} + \epsilon_{it}$$

In moving from a typical panel data analysis to an event study approach, we move from a single treatment T_{it}^j to a vector of dummy variables $\{T_{it}^{\ell,j}\}_\ell$. In the setting where the fully dynamic equation is not estimated, there are $L + K$ dummy variables to be concerned with (as we drop the period $\ell = -1$ to avoid any collinearity problems). We therefore similarly expand the set of instruments from Z_{it} to $\{Z_{it}^{\ell}\}$, again of size $L + K$. This vector of instrument dummies is created in the same way the vector of treatment dummies, and described in Section 2.3.2. By expanding to

multiple instruments, however, each coefficient in a two-stage regression will be given by a weighted average of LATEs, as discussed in Angrist and Imbens (1995). That is, for each resulting coefficient γ_ℓ on any dummy $T^\ell_{it,j}$:

$$\gamma_\ell = \sum_{k=1}^{L+K} \beta_k \frac{\text{Cov}(T^j_{it} Y_{it}, Z_j)}{\text{Cov}(T^\ell_{it,j}, Z_j)},$$

where the weights β_k are in the interval $[0, 1]$ for all k and satisfy $\sum_k \beta_k = 1$. As in Calvi et al. (2019), define $q = \frac{p_1}{p_1 - p_0}$. Using their Theorem 1 and the result above:

$$\begin{aligned}
\gamma_\ell &= \sum_{j=1}^{L+K} \beta_j \frac{\text{Cov}(Y_{it} T_{it}, Z_j)}{\text{Cov}(T^\ell_{it}, Z_j)} \\
&= \sum_{j=1}^{L+K} \beta_j \lambda_j \\
&= \sum_{j=1}^{L+K} \beta_j \mathbb{E}[q_j Y_1 + (1 - q_j) Y_0 | C] \\
&= \sum_{j=1}^{L+K} \beta_j \mathbb{E}[q Y_1 + (1 - q) Y_0 | C] \text{ (as each } q_j = q) \\
&= \mathbb{E}\left[\left(\sum_{j=1}^{L+K} \beta_j \right) q Y_1 + \left(\sum_{j=1}^{L+K} \beta_j \right) (1 - q) Y_0 | C \right] \\
&= \mathbb{E}[q Y_1 + (1 - q) Y_0 | C] \text{ (as weights sum to 1).}
\end{aligned}$$

Hence, for any time period ℓ, Calvi et al.'s Theorem 1 applies. One can therefore use two mismeasured treatments T^a and T^b with the same properties as in their paper (so that $p^a_0 = p^b_1 = 0$) and construct the local cohort average treatment effect:

$$\rho_\ell = \mathbb{E}\left[Y^e_{i,t+\ell} - Y^\infty_{i,t+\ell} | E_i = e, C \right] = \hat{\gamma}^a_\ell - \hat{\gamma}^b_\ell$$

■

B.2 Details of the Travel Cost Algorithm

The travel cost algorithm is used to infer the earliest date each medical professional was exposed to a continuing education event targetting either FBT use or olanzapine prescribing in eating disorder treatments.

The first step in the algorithm is to assign a location to each specialist. While MarketScan does not have specifically geotagged locations for their physicians, they do have information on the Metropolitan Statistical Area in which the main enrollee on each insurance plan resides. Hence, each claim in a physician's treatment profile is tagged to one of these MSAs. By taking the bulk of patients seen in a given month and taking a geocentric average of their home MSAs (taking as each patient's location the global midpoint of their MSA), I can assign a specific location to each specialist-month observation. To avoid large errors, I discard specialist-month observations that treat patients from larger than a 100-mile radius.

Once a specific location has been assigned to a specialist-month, I can compute the travel costs between therapists and a given conference. The algorithm allows a specialist to travel to the conference either by car directly, or by any network of flights. To estimate driving time and costs, I allow for different average driving speeds in urban areas and on freeways/interstates, and estimate the price of gas using data from the United States Bureau of Labor Statistics (BLS). To estimate flying time and costs, I incorporate data on airport locations, flight availability, and airfare from the United States Bureau of Transportation. Then, I construct a network between a specialist's origin point (their home location) and their destination (the conference location) that allows them to (i) drive to any of the 5 airports closest to their home, (ii) take any network of flights from that airport to any of the 5 airports closest to their destination, and (iii) drive from that airport to the conference location. Once this network is completed, I assume that travelers will choose the cheapest option (in terms of both airfare and opportunity cost of travel).

Opportunity cost of time is calculated using BLS wage data. I estimate this opportunity cost in two ways: assigning each specialist in my data set the same hourly wage (the average median wage for all specialist groups in the sample), or assigning each specialist group their own median wage. For example, psychiatrists would be assigned a median wage of $105.95 per hour while mental health clinicians would be assign a median wage of $21.46 per hour. Under the second method, travel

costs are reported as a percentage of each specialist's average salary, to keep units consistent.

While there is clearly a large amount of variation in specialist wages, this variation appears to be negatively correlated with true attendance. That is, those with the lowest median wages (e.g., treatment center workers, therapists) have a greater incentive to attend conferences on eating disorder treatments than general practice doctors or psychiatrists, who treat a larger range of diagnoses. However, when adjusting for different salaries, I am implicitly making the costs of travel (gas, airfare, etc.) less inhibitive for those with higher salaries, so the algorithm may be more likely to predict treatment for those who are *less* incentivized to truly attend.

As discussed in the paper, once this algorithm is complete, each specialist-conference pair is assigned a travel cost $c \in \mathbb{R}$. For the "unnormalized" option where each specialist is assigned a flat salary, this is a monetary measure in 2016 U.S. dollars; on the other hand, when different salaries are assigned to different specialists, this is measured as each specialist's travel costs in salary hours. From this continuous measure, I form the dichotomous prediction of treatment and control groups using an artificial cutoff $\eta \in [0, 1]$, where any specialist with travel cost at or below the value $F(\eta)$ is considered to have attended the conference.

B.3 Additional Results

As mentioned previously, the results shown in the paper are robust to several iterations of estimation. I re-estimated the results with a continuous dependent variable instead of a binary one, and with a travel cost that was normalized to be in terms of each therapist's estimated salary (instead of a pure monetary measure). These sensitivity results are shown in Figure B·1 for all 12 regression coefficients of the dynamic treatment effect. Note that throughout, I include the figures only for family-based therapies; the results are similar for olanzapine prescriptions.

Additionally, the results are re-estimated with various thresholds used in assigning treatment/control status. Each treatment measure assumes that those therapists with travel costs in the lowest η-percentile of all travel costs for a given conference attended it, and were thus treated. Each specification consists of a treatment measure using η as the percentile, and the other using $1 - \eta$. For example, the specification of choice uses the "strict" treatment measure as those whose travel costs are in the bottom 2%, while the "liberal" one applies treatment to those in the bottom 98%. I also test specifications where $\eta \in \{1, 2, 5, 10, 15\}$. The resulting variation in estimated coefficients is illustrated in Figure B·2.

The results are generally quite consistent—if anything, models estimated with more stringent treatment thresholds (smaller η) appear to detect larger estimates, but have larger standard errors as well. While future work may elaborate on the optimal decision of treatment threshold to balance the trade off inherent in its selection, this figure provides sufficient evidence that the choice of threshold contributes little to the overall result.

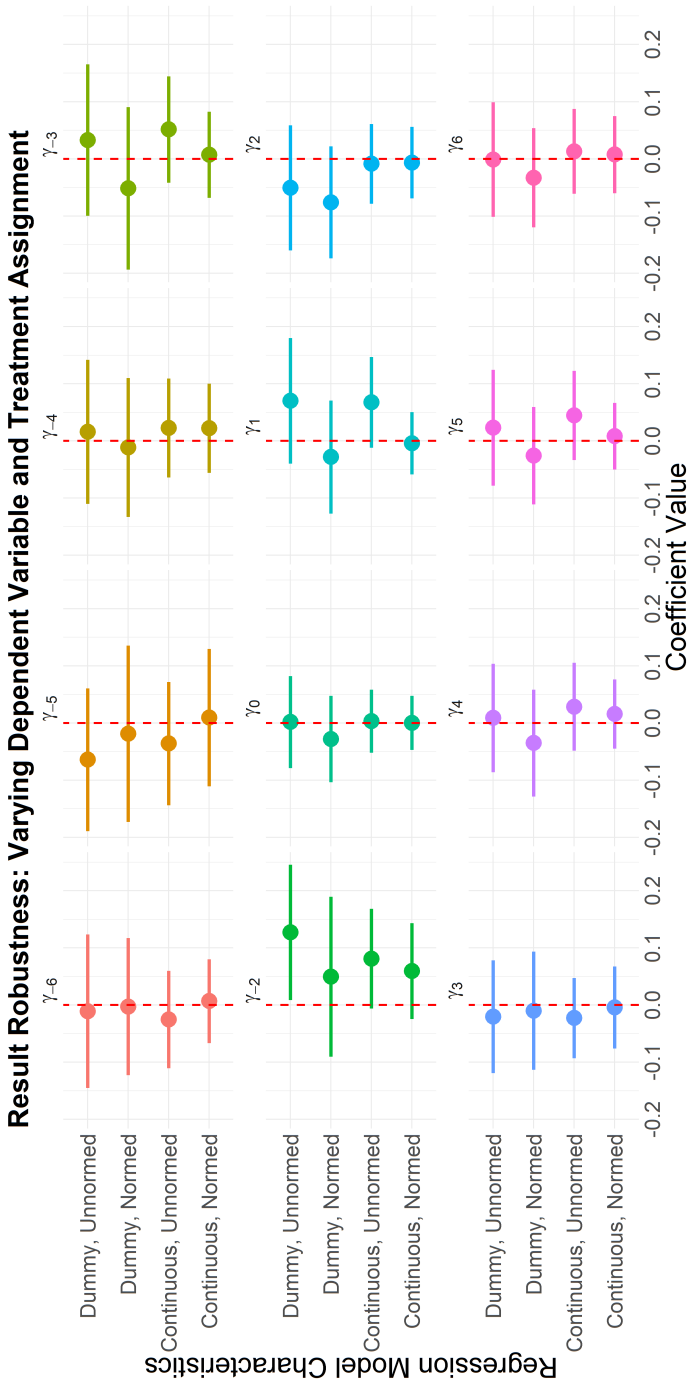


Figure B·1
Response to Family-Based Therapy Conferences, Sensitivity by Estimation Strategy

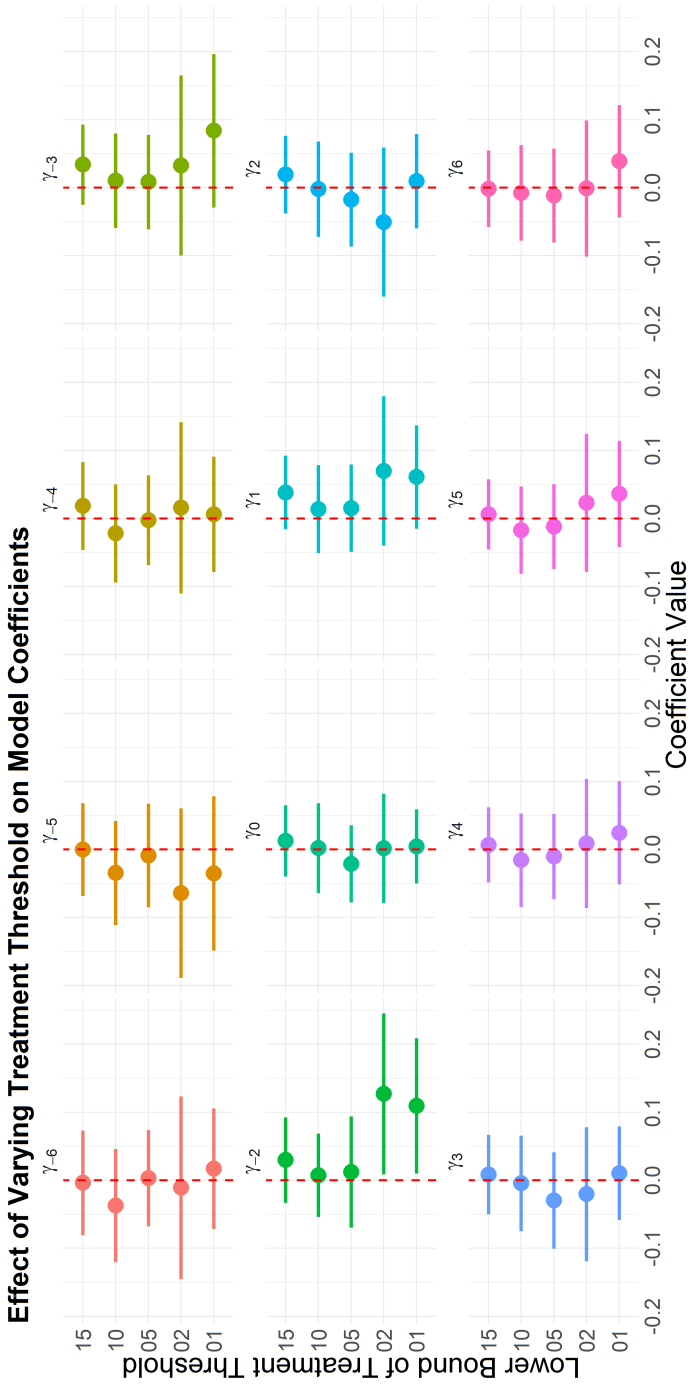


Figure B·2
Response to Family-Based Therapy Conferences, Sensitivity by Treatment Thresholds

Appendix C

Appendix to Chapter 3

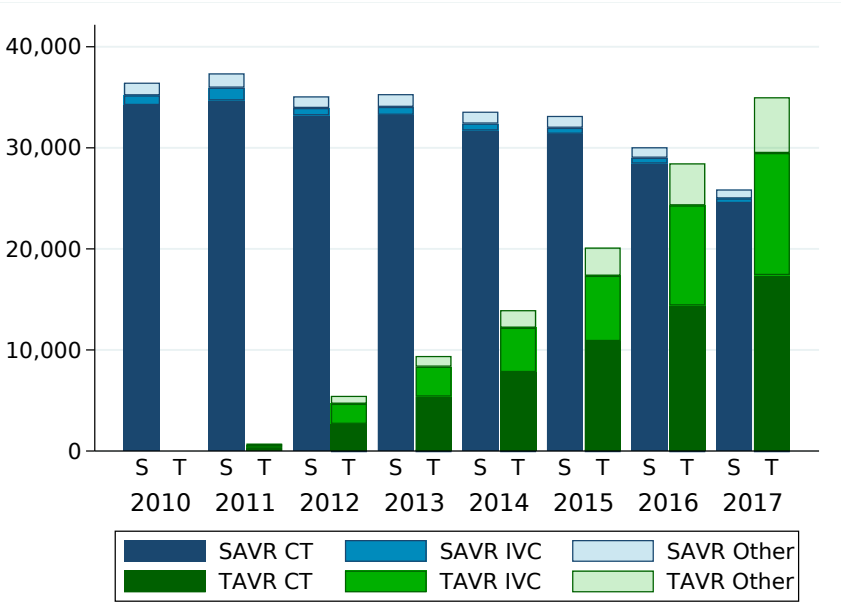
	Cardiothoracic Surgeons			Interventional Cardiologists		
	All Surgeries	TAVR	SAVR	All Surgeries	TAVR	SAVR
2010	85.97%	–	85.97%	7.76%	–	7.76%
2011	84.03%	41.39%	84.97%	9.69%	51.11%	8.78%
2012	81.06%	46.68%	87.45%	12.82%	45.61%	6.72%
2013	80.50%	56.38%	88.21%	13.73%	37.95%	5.99%
2014	77.18%	54.48%	88.60%	17.03%	39.61%	5.68%
2015	72.87%	51.53%	88.64%	21.02%	42.17%	5.38%
2016	66.88%	48.39%	89.04%	27.14%	45.50%	5.12%
2017	61.42%	46.32%	88.82%	31.87%	46.78%	4.83%

Table C.1

Role of Cardiologists in Aortic Stenosis Procedures, 2010–2017

Notes: Each cell represents the fraction of the surgical type performed by the type of medical professional in a given year. Sample is limited to only TAVR/SAVR procedures performed either by a Cardiothoracic surgeon or an interventional cardiologist. Cardiothoracic surgeons are those whose primary specialty is listed as "cardiac surgery", "thoracic surgery", or "general surgery"; interventional cardiologists are those whose primary specialty is listed as "interventional cardiology", "cardiology", or "cardiovascular disease".

Figure C·1
Timeline of TAVR Adoption



Notes: Figure shows diffusion of TAVR procedures among different cardiac surgeon specialties over time. Total volume of surgical valve replacements (SAVR and TAVR, labelled as "S" and "T" on the x-axis) for the full U.S. Medicare population are shown, with a breakdown of surgeon specialty. Cardiothoracic surgeons ("CT") are those whose primary specialty is listed as "cardiac surgery", "thoracic surgery", or "general surgery"; interventional cardiologists ("IVC") are those whose primary specialty is listed as "interventional cardiology", "cardiology", or "cardiovascular disease". Other surgeons include those with specialties outside of these fields (e.g., internal medicine) who also performed the procedures over time.

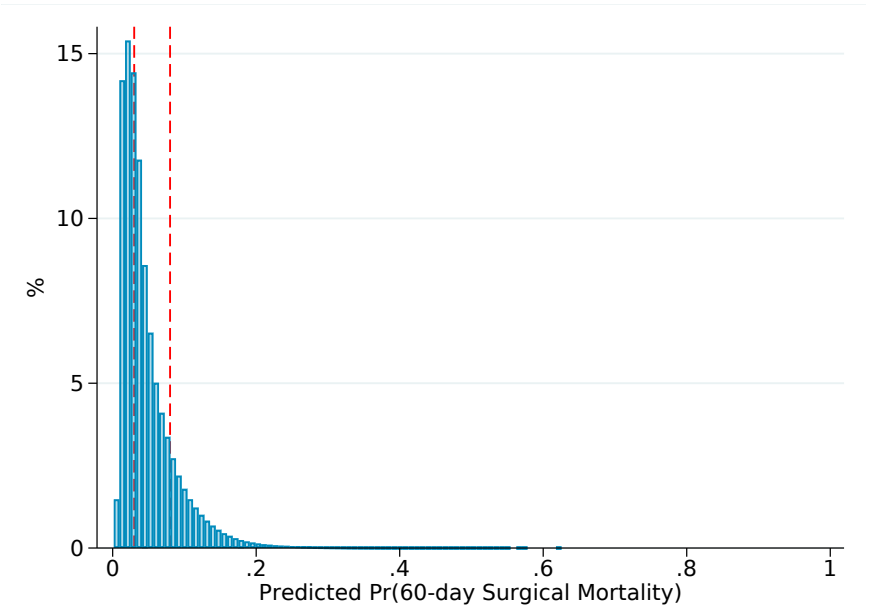


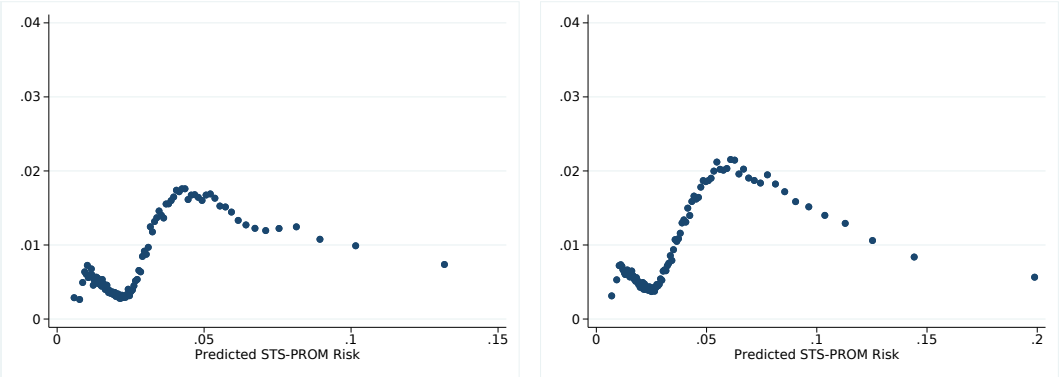
Figure C·2
Predicted Patient Risk of Surgical Mortality (STS-PROM)

	30-day surgical mortality
Any previous surgery	-0.0211
	(-0.71)
# of previous surgeries	-0.0895
	(-4.63)
Previous bypass	-0.859
	(-37.41)
Previous valve replacement	-0.662
	(-34.02)
Previous PCI	-0.699
	(-35.98)
Patient age	0.0309
	(51.76)
Female	0.0446
	(4.03)
Black	0.186
	(9.92)
Hispanic	0.0582
	(1.16)
Other Minority Race	0.0500
	(1.71)
# of Chronic Conditions	-0.0557
	(-17.38)
CC: CHF	1.156
	(81.22)
CC: Diabetes	0.177
	(14.30)
CC: Hypertension	-0.450
	(-20.84)
CC: Stroke	0.377
	(23.95)
CC: AMI	0.844
	(60.13)
CC: COPD	0.305
	(24.64)
Income Quintile 1	0.0890
	(4.56)
Income Quintile 2	0.0474
	(2.61)
Income Quintile 3	0.0237
	(1.46)
Income Quintile 4	0.0235
	(1.59)
Observations	714,400

Table C.2
STS-PROM Logistic Regression Coefficients

Figure C·3
Likelihood of Surgical Intervention Before/After TAVR Adoption

(a) Prior to Adoption (2010) **(b)** Adoption Years 0–2 (relative time)

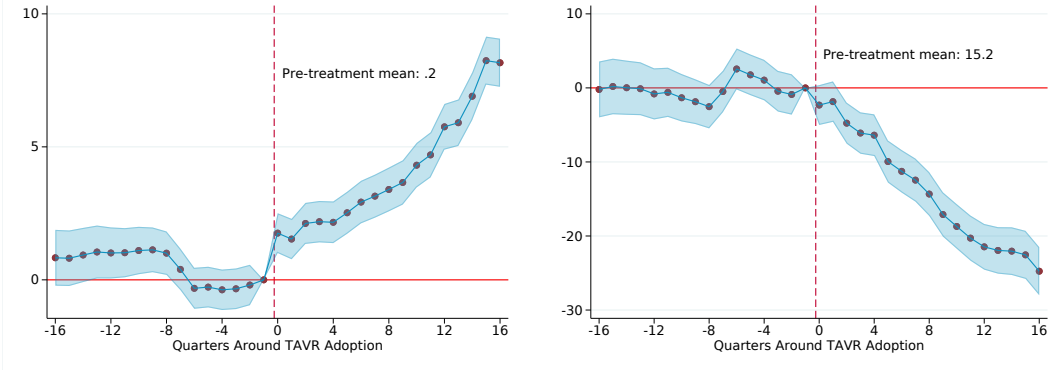


Notes: Figure shows estimated likelihood of an individual patient receiving valve replacement surgery (TAVR or SAVR) by estimated risk (based on STS-PROM score). Panel (a) shows relationship in year prior to TAVR approval (2010), while Panel (b) shows relationship in commuting zones during the first three years of TAVR adoption.

Figure C·4

Effect of TAVR Adoption on Total IVC Surgical Volumes, Commuting Zone Level

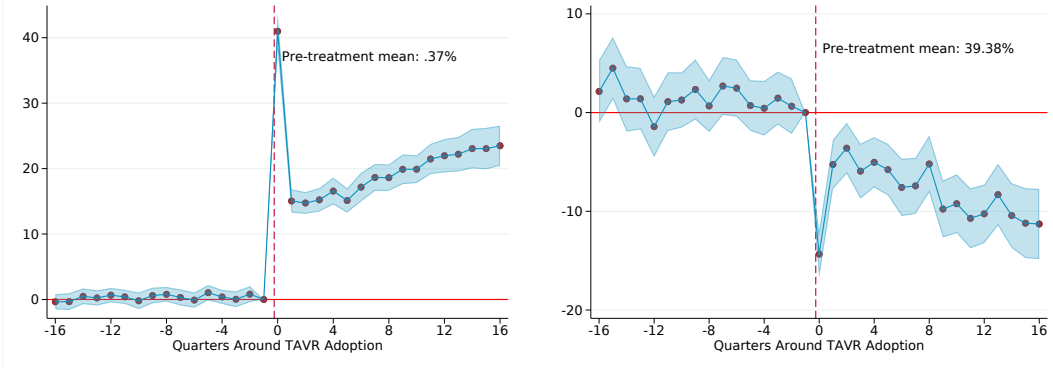
(a) SAVR/TAVR Surgeries **(b)** PCIs



Notes: Figure shows estimated impact of TAVR adoption on the total volume of surgical interventions performed by IVCs. Panel (a) shows the effect on all SAVR/TAVR surgeries, and panel (b) shows the effect on PCI procedures. Interventional cardiologists who perform fewer than 10 inpatient surgeries per year are dropped from estimation, and standard errors are clustered at the commuting zone level. *Abbreviations:* IVC = Interventional Cardiologist

Figure C·5
Effect of TAVR Adoption on Interventional Cardiologist Treatment Shares

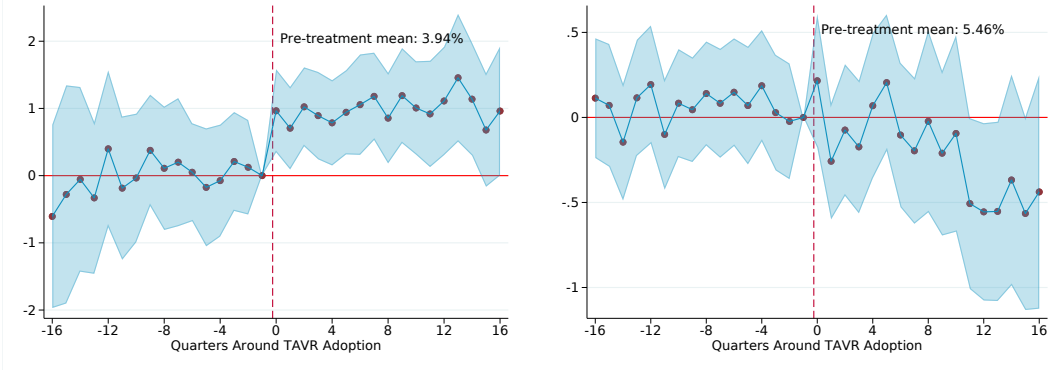
(a) SAVR/TAVR Surgeries **(b)** PCIs



Notes: Figure shows estimated impact of TAVR adoption on treatment decisions made by interventional cardiologists. The outcome variable in each panel is the total volume of each procedure performed by an interventional cardiologist; panel (a) shows the effect of TAVR on the use of all valve replacement surgeries, while panel (b) shows its effect on the use of PCIs. Interventional cardiologists who perform fewer than 10 inpatient surgeries per year are dropped from estimation, and standard errors are clustered at the commuting zone level.

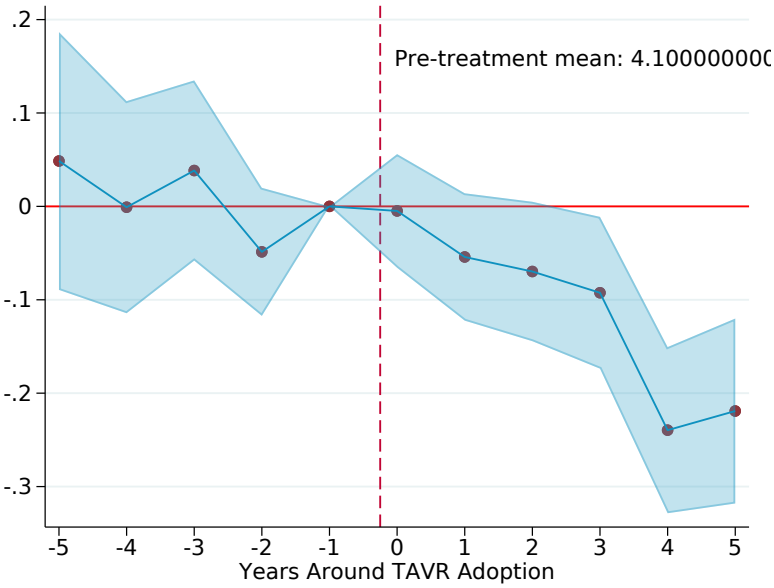
Figure C·6
Effect of TAVR Adoption on IVC Treatment Decisions (Risk)

(a) SAVR/TAVR Surgeries **(b)** PCI Interventions



Notes: Standard errors are clustered at the commuting zone level.

Figure C·7
Effect of TAVR Adoption on Total IVC Surgical Volumes, Commuting Zone Level



Notes: Figure shows estimated impact of TAVR adoption on the likelihood that a patient will receive any surgical intervention, including all SAVR, TAVR, and PCI procedures regardless of provider type. Patient pool is restricted to patients with appropriate cardiac symptoms who have not previously received surgery. Standard errors are clustered at the commuting zone level.

Bibliography

Abadie, A. (2002). Bootstrap tests for distributional treatment effects in instrumental variable models. Journal of the American Statistical Association, 97(457):284–292.

Abaluck, J. and Compiani, G. (2020). A method to estimate discrete choice models that is robust to consumer search. Technical Report w26849, National Bureau of Economic Research, Cambridge, MA.

Abaluck, J. and Gruber, J. (2011). Choice inconsistencies among the elderly: Evidence from plan choice in the medicare part d program. American Economic Review, 101(4):1180–1210.

Abaluck, J. and Gruber, J. (2016a). Evolving choice inconsistencies in choice of prescription drug insurance. American Economic Review, 106(8):2145–2184.

Abaluck, J. and Gruber, J. (2016b). Improving the quality of choices in health insurance markets. Technical Report w22917, National Bureau of Economic Research, Cambridge, MA.

Ackerberg, D. A. and Gowrisankaran, G. (2006). Quantifying equilibrium network externalities in the ACH banking industry. The RAND Journal of Economics, 37(3):738–761.

Adams, D. H., Popma, J. J., Reardon, M. J., Yakubov, S. J., Coselli, J. S., Deeb, G. M., Gleason, T. G., Buchbinder, M., Hermiller, J., Kleiman, N. S., Chetcuti, S., Heiser, J., Merhi, W., Zorn, G., Tadros, P., Robinson, N., Petrossian, G., Hughes, G. C., Harrison, J. K., Conte, J., Maini, B., Mumtaz, M., Chenoweth, S., and Oh, J. K. (2014). Transcatheter aortic-valve replacement with a self-expanding prosthesis. New England Journal of Medicine, 370(19):1790–1798.

Adler, N. E. and Rehkopf, D. H. (2008). U.S. disparities in health: Descriptions, causes, and mechanisms. Annual Review of Public Health, 29(1):235–252.

Aguirregabiria, V. and Jeon, J. (2020). Firms' beliefs and learning: Models, identification, and empirical evidence. Review of Industrial Organization, 56(2):203–235.

Agüero, J. M. and Beleche, T. (2017). Health shocks and their long-lasting impact on health behaviors: evidence from the 2009 H1N1 pandemic in mexico. Journal of Health Economics, 54:40–55.

Alalouf, M., Miller, S., and Wherry, L. (2019). What difference does a diagnosis make? Evidence from marginal patients. Technical Report w26363, National Bureau of Economic Research, Cambridge, MA.

Almond, D., Doyle, Jr., J. J., Kowalski, A. E., and Williams, H. (2010). Estimating marginal returns to medical care: Evidence from at-risk newborns. The Quarterly Journal of Economics, 125(2):591–634.

Altmejd, A., Barrios-Fernandez, A., Drljie, M., Goodman, J., Hurwitz, M., Kovac, D., Mulhern, C., Nielson, C., and Smith, J. (2021). O brother, where start thou? Sibling spillovers on college and major choice in four countries. The Quarterly Journal of Economics, 136(3):1831–1886.

American Psychiatric Association (2006). Treatment of patients with eating disorders, third edition. The American Journal of Psychiatry, 163(7 Suppl):4–54.

Angrist, J. D. and Imbens, G. W. (1995). Two-stage least squares estimation of average causal effects in models with variable treatment intensity. Journal of the American Statistical Association, 90(430):431–442.

Arcaya, M. C. and Figueroa, J. F. (2017). Emerging trends could exacerbate health inequities in the united states. Health Affairs, 36(6):992–998.

Arcelus, J., Mitchell, A. J., Wales, J., and Nielsen, S. (2011). Mortality rates in patients with anorexia nervosa and other eating disorders: a meta-analysis of 36 studies. Archives of General Psychiatry, 68(7):724–731.

Archibong, B. and Annan, F. (2021). 'We are not guinea pigs': The effects of negative news on vaccine compliance. SSRN Electronic Journal.

Arieli, I., Babichenko, Y., Peretz, R., and Young, H. P. (2020). The speed of innovation diffusion in social networks. Econometrica, 88(2):569–594.

Arkfeld, C. K. (2021). The sick bias. The Yale Journal of Biology and Medicine, 94(3):517–519.

Arni, P., Dragone, D., Goette, L., and Ziebarth, N. R. (2021). Biased health perceptions and risky health behaviors—Theory and evidence. Journal of Health Economics, 76:102425.

Arrow, K. J. (1962). The economic implications of learning by doing. The Review of Economic Studies, 29(3):155–173.

Ash, A. S., Mick, E. O., Ellis, R. P., Kiefe, C. I., Allison, J. J., and Clark, M. A. (2017). Social determinants of health in managed care payment formulas. JAMA Internal Medicine, 177(10):1424–1430.

Ash, E., Chen, D. L., and Naidu, S. (2019). Ideas have consequences: The impact of law and economics on american justice. NBER Working Paper.

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

Bajaj, A., Pancholy, S., Sethi, A., and Rathor, P. (2017). Safety and feasibility of PCI in patients undergoing TAVR: A systematic review and meta-analysis. Heart & Lung, 46(2):92–99.

Bansal, P., Bertels, S., Ewart, T., MacConnachie, P., and O'Brien, J. (2012). Bridging the research-practice gap. Academy of Management Perspectives, 26(1):73–92.

Barseghyan, L., Molinari, F., O'Donoghue, T., and Teitelbaum, J. C. (2013). The nature of risk preferences: Evidence from insurance choices. The American Economic Review, 103(6):2499–2529.

Barseghyan, L., Molinari, F., O'Donoghue, T., and Teitelbaum, J. C. (2018). Estimating risk preferences in the field. Journal of Economic Literature, 56(2):501–564.

Bellemare, M. F. and Wichman, C. J. (2020). Elasticities and the inverse hyperbolic sine transformation. Oxford Bulletin of Economics and Statistics, 82(1):50–61.

Bentley, P. N., Wilson, A. G., Derwin, M. E., Scodellaro, R., and Jackson, R. E. (2002). Reliability of assigning correct current procedural terminology—4 E/M codes. Annals of Emergency Medicine, 40(3):269–274.

Bhatia, R. S., Levinson, W., Shortt, S., Pendrith, C., Fric-Shamji, E., Kallewaard, M., Peul, W., Veillard, J., Elshaug, A., Forde, I., and Kerr, E. A. (2015). Measuring the effect of Choosing Wisely: An integrated framework to assess campaign impact on low-value care. BMJ Quality & Safety, 24(8):523–531.

Biasi, B., Dahl, M. S., and Moser, P. (2019). Career effects of mental health. Working Paper.

Bleichrodt, H. and Eeckhoudt, L. (2006). Willingness to pay for reductions in health risks when probabilities are distorted. Health Economics, 15(2):211–214.

Borusyak, K. and Jaravel, X. (2021). Revisiting event study designs: Robust and efficient estimation. Working Paper.

Bouckaert, N., Gielen, A. C., and Van Ourti, T. (2020). It runs in the family – Influenza vaccination and spillover effects. Journal of Health Economics, 74:102386.

Browning, M., Chiappori, P.-A., and Weiss, Y. (2014). Economics of the Family. Cambridge University Press.

Bulik, C. M., Berkman, N. D., Brownley, K. A., Sedway, J. A., and Lohr, K. N. (2007). Anorexia nervosa treatment: A systematic review of randomized controlled trials. International Journal of Eating Disorders, 40(4):310–320.

Bundorf, K., Polyakova, M., and Tai-Seale, M. (2021a). How do humans interact with algorithms? Experimental evidence from health insurance. Technical Report w25976, National Bureau of Economic Research, Cambridge, MA.

Bundorf, M. K., DeMatteis, J., Miller, G., Polyakova, M., Streeter, J., and Wivagg, J. (2021b). Risk perceptions and protective behaviors: Evidence from the COVID-19 pandemic. Technical Report w28741, National Bureau of Economic Research, Cambridge, MA.

Burke, L. A. and Rau, B. (2010). The research-teaching gap in management. Academy of Management Learning & Education, 9(1):132–143.

Callaway, B. and Sant'Anna, P. (2018). Difference-in-differences with multiple time periods and an application on the minimum wage and employment. SSRN Electronic Journal.

Calvi, R., Lewbel, A., and Tommasi, D. (2019). Women's empowerment and family health: Estimating late with mismeasured treatment. Working Paper.

Cardon, J. H. and Hendel, I. (2001). Asymmetric information in health insurance: Evidence from the national medical expenditure survey. The RAND Journal of Economics, 32(3):408–427.

Case, A. and Deaton, A. (2015). Rising morbidity and mortality in midlife among white non-Hispanic Americans in the 21st century. Proceedings of the National Academy of Sciences, 112(49):15078–15083.

Casper, E. S. (2007). The theory of planned behavior applied to continuing education for mental health professionals. Psychiatric Services, 58(10):1324–1329.

Catalano, M. A., Rutkin, B., Kohn, N., Hartman, A., and Yu, P.-J. (2020). Does heart valve team risk assessment predict outcomes after transcatheter aortic valve replacement? The International Journal of Angiology : Official Publication of the International College of Angiology, Inc, 29(1):39–44.

Chaisemartin, C. d., D'Haultfoeuille, X., and Guyonvarch, Y. (2021). DID_multiplegt: Stata module to estimate sharp Difference-in-Difference designs with multiple groups and periods.

Chandra, A., Kakani, P., and Sacarny, A. (2020). Hospital allocation and racial disparities in health care. Technical Report w28018, National Bureau of Economic Research.

Chandra, A. and Staiger, D. O. (2007). Productivity spillovers in health care: Evidence from the treatment of heart attacks. Journal of Political Economy, 115(1):103–140.

Chen, M. (2021). Family of San Diego COVID-19 victim makes emotional vaccine plea. ABC10 News. Place: San Diego, CA.

Choudhry, N. K., Fletcher, R. H., and Soumerai, S. B. (2005). Systematic review: The relationship between clinical experience and quality of health care. Annals of Internal Medicine, 142(4):260.

Choudhry, N. K., Shrank, W. H., Levin, R. L., Lee, J. L., Jan, S. A., Brookhart, M. A., and Solomon, D. H. (2009). Measuring concurrent adherence to multiple related medications. The American journal of managed care, 15(7):457–464.

Christensen, C. M., Baumann, H., Ruggles, R., and Sadtler, T. M. (2006). Disruptive innovation for social change. Harvard business review, 84(12):94–101, 163.

Christensen, C. M., Grossman, J. H., and Hwang, J. (2009). The Innovator's Prescription: A Disruptive Solution for Health Care. McGraw Hill.

Christensen, C. M., Raynor, M. E., and McDonald, R. (2015). What is disruptive innovation? Harvard Business Review, 93(12):44–53.

Chua, K.-P., Schwartz, A. L., Volerman, A., Conti, R. M., and Huang, E. S. (2016). Use of low-value pediatric services among the commercially insured. Pediatrics, 138(6).

Church, E. A., Heath, O. J., Curran, V. R., Bethune, C., Callanan, T. S., and Cornish, P. A. (2010). Rural professionals' perceptions of interprofessional continuing education in mental health. Health and Social Care in the Community, 18(4):433–443.

Coburn, C. E. and Penuel, W. R. (2016). Research-practice partnerships in education: Outcomes, dynamics, and open questions. Educational Researcher, 45(1):48–54.

Cohen, A. and Einav, L. (2007). Estimating risk preferences from deductible choice. American Economic Review, 97(3):745–788.

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.

Conlon, C. and Gortmaker, J. (2020). Best practices for differentiated products demand estimation with PyBLP. The RAND Journal of Economics, 51(4):1108–1161.

Couturier, J., Kimber, M., and Szatmari, P. (2013). Efficacy of family-based treatment for adolescents with eating disorders: A systematic review and meta-analysis. International Journal of Eating Disorders, 46(1):3–11.

Curran, V. R., Fleet, L., and Kirby, F. (2006). Factors influencing rural health care professionals' access to continuing professional education. Australian Journal of Rural Health, 14(2):51–55.

Cutler, D. M. and Huckman, R. S. (2003). Technological development and medical productivity: the diffusion of angioplasty in New York state. Journal of Health Economics, 22(2):187–217.

Cutler, D. M., Long, G., Berndt, E. R., Royer, J., Fournier, A.-A., Sasser, A., and Cremieux, P. (2007). The value of antihypertensive drugs: A perspective on medical innovation. Health Affairs, 26(1):97–110.

Cutler, D. M. and Zeckhauser, R. (2004). Extending the theory to meet the practice of insurance. Technical report, Brookings Institution Press.

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

Dare, C. and Eisler, I. (2000). A multi-family group day treatment programme for adolescent eating disorder. European Eating Disorders Review: The Professional Journal of the Eating Disorders Association, 8(1):4–18.

de Chaisemartin, C. and D'Haultfoeuille, X. (2019). Two-way fixed effects estimators with heterogeneous treatment effects. Technical Report Technical Report w25904, National Bureau of Economic Research, Cambridge, MA. Publication Title: NBER Working Paper 25904.

Depalo, D., Bhattacharya, J., Atella, V., and Belotti, F. (2019). When technological advance meets physician learning in drug prescribing. Technical Report w26202, National Bureau of Economic Research, Cambridge, MA.

Dubin, J. A. and McFadden, D. L. (1984). An econometric analysis of residential electric appliance holdings and consumption. Econometrica, 52(2):345–362.

Dupas, P. (2011). Health behavior in developing countries. Annual Review of Economics, 3(1):425–449.

Dvir, D., Barbash, I. M., Ben-Dor, I., Okubagzi, P., Satler, L. F., Waksman, R., and Pichard, A. D. (2012). The development of transcatheter aortic valve replacement in the USA. Archives of Cardiovascular Diseases, 105(3):160–164.

Dysart, A. M. and Tomlin, G. S. (2002). Factors related to evidence-based practice among U.S. occupational therapy clinicians. American Journal of Occupational Therapy, 56(3):275–284.

D'Agostino, R. S., Jacobs, J. P., Badhwar, V., Fernandez, F. G., Paone, G., Wormuth, D. W., and Shahian, D. M. (2018). The society of thoracic surgeons adult cardiac surgery database: 2018 update on outcomes and quality. The Annals of Thoracic Surgery, 105(1):15–23.

Eichner, M. J. (1997). Medical expenditures and major risk health insurance. Doctoral Dissertation, Massachusetts Institute of Technology.

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

Eil, D. and Rao, J. M. (2011). The good news-bad news effect: Asymmetric processing of objective information about yourself. American Economic Journal: Microeconomics, 3(2):114–138.

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.

Ellis, R. P., Hsu, H. E., Siracuse, J. J., Walkey, A. J., Lasser, K., Jacobson, B., Andriola, C., Hoagland, A., Liu, Y., Song, C., Kuo, T.-C., and Ash, A. (2022). Diagnostic items: A new framework for disease surveillance, prediction, and risk adjustment. JAMA Health Forum.

Epstein, L. G., Noor, J., and Sandroni, A. (2010). Non-bayesian learning. The B.E. Journal of Theoretical Economics, 10(1).

Ericson, K. M., Kircher, P., Spinnewijn, J., and Starc, A. (2020). Inferring risk perceptions and preferences using choice from insurance menus: theory and evidence. The Economic Journal.

Eriksen, T. L. M., Gaulke, A., Skipper, N., and Svensson, J. (2021). The impact of childhood health shocks on parental labor supply. Journal of Health Economics.

Fadlon, I. and Nielsen, T. H. (2019). Family health behaviors. American Economic Review, 109(9):3162–3191.

Falk, V., Holm, P. J., Iung, B., Lancellotti, P., Lansac, E., Munoz, D. R., and Rosenhek, R. (2017). 2017 ESC/EACTS guidelines for the management of valvular heart disease. Kardiologia Polska (Polish Heart Journal), 76(1):1–62.

Finkelstein, A., Luttmer, E. F. P., and Notowidigdo, M. J. (2009). Approaches to estimating the health state dependence of the utility function. The American Economic Review, 99(2):116–121.

Finkelstein, A., Luttmer, E. F. P., and Notowidigdo, M. J. (2013). What good is wealth without health? The effect of health on the marginal utility of consumption. Journal of the European Economic Association, 11(suppl_1):221–258.

Finkelstein, A., Persson, P., Polyakova, M., and Shapiro, J. (2021). A taste of their own medicine: Guideline adherence and access to expertise. Working Paper, page 38.

Flament, M. F., Bissada, H., and Spettigue, W. (2012). Evidence-based pharmacotherapy of eating disorders. The International Journal of Neuropsychopharmacology, 15(02):189–207.

Garthwaite, C., Gross, T., and Notowidigdo, M. J. (2014). Public health insurance, labor supply, and employment lock. The Quarterly Journal of Economics, 129(2):653–696.

Giardinelli, C. (2021). 'Get the vaccine:' Oregon man pleads as 23-year-old wife fights for her life. KATU ABC2. Place: Medford, OR.

Gilardi, F. (2012). Transnational diffusion: Norms, ideas, and policies. In Carlsnaes, W., Risse, T., and Simmons, B., editors, Handbook of International Relations, Thousand Oaks, pages 453–477. Sage Publications.

Glanz, K., Rimer, B. K., and Viswanath, K. (2008). Health Behavior and Health Education: Theory, Research, and Practice. John Wiley & Sons, 4 edition.

Glasgow, R. E. and Emmons, K. M. (2007). How can we increase translation of research into practice? Types of evidence needed. Annual Review of Public Health, 28(1):413–433.

Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. Technical Report w25018, National Bureau of Economic Research, Cambridge, MA.

Goodman-Bacon, A., Goldring, T., and Nichols, A. (2019). BACONDECOMP: Stata module to perform a Bacon decomposition of difference-in-differences estimation.

Graham, A. K., Lattie, E. G., and Mohr, D. C. (2019). Experimental therapeutics for digital mental health. JAMA Psychiatry. Publication Title: JAMA Psychiatry.

Grimshaw, J. M., Shirran, L., Thomas, R., Mowatt, G., Fraser, C., Bero, L., Grilli, R., Harvey, E., Oxman, A., and O'Brien, M. A. (2001). Changing provider behavior: An overview of systematic reviews of interventions. Medical Care, 39(8):II–2.

Gross, T., Layton, T., and Prinz, D. (2020). The liquidity sensitivity of healthcare consumption: Evidence from social security payments. Technical Report w27977, National Bureau of Economic Research.

Gruber, J., Handel, B., Kina, S., and Kolstad, J. (2020). Managing intelligence: Skilled experts and AI in markets for complex products. Technical Report w27038, National Bureau of Economic Research, Cambridge, MA.

Hadley, J. (2007). Insurance coverage, medical care use, and short-term health changes following an unintentional injury or the onset of a chronic condition. JAMA, 297(10):1073–1084.

Handel, B. R. (2013). Adverse selection and inertia in health insurance markets: When nudging hurts. American Economic Review, 103(7):2643–2682.

Handel, B. R. and Kolstad, J. T. (2015). Health insurance for 'humans': Information frictions, plan choice, and consumer welfare. The American Economic Review, 105(8):2449–2500.

Harris, J. W. and Stöcker, H. (1998). Handbook of Mathematics and Computational Science. Springer Science & Business Media.

Hauser, D. and Bohren, J. A. (2021). Learning with heterogeneous misspecified models: Characterization and robustness. Econometrica, In Press.

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

Hodor, M. (2021). Family health spillovers: Evidence from the RAND health insurance experiment. Journal of Health Economics, 79:102505.

Holt, C. A. and Smith, A. M. (2009). An update on Bayesian updating. Journal of Economic Behavior & Organization, 69(2):125–134.

Hong, Y.-R., Jo, A., and Mainous, A. G. I. (2017). Up-to-date on preventive care services under affordable care act. Medical Care, 55(8):771–780.

Hu, Y. (2008). Identification and estimation of nonlinear models with misclassification error using instrumental variables: A general solution. Journal of Econometrics, 144(1):27–61.

Huckman, R. S. and Stern, A. D. (2022). The adoption and diffusion of medical technology: Evidence from cardiac procedures. Working Paper.

Hudson, S., Hull, P., and Liebersohn, J. (2017). Interpreting instrumented difference-in-differences. Working Paper.

Hugenholtz, N. I. R., de Croon, E. M., Smits, P. B., van Dijk, F. J. H., and Nieuwenhuijsen, K. (2008). Effectiveness of e-learning in continuing medical education for occupational physicians. Occupational Medicine, 58(5):370–372.

Hull, P. (2018). Estimating hospital quality with quasi-experimental data. Working Paper, Working Paper:31.

Iizuka, T., Nishiyama, K., Chen, B., and Eggleston, K. (2021). False alarm? Estimating the marginal value of health signals. Technical Report 23413, National Bureau of Economic Research, Cambridge, MA.

Imbens, G. W. and Angrist, J. D. (1994). Identification and estimation of local average treatment effects. Econometrica, 62:467–475.

Jensen, P. S., Bhatara, V. S., Vitiello, B., Hoagwood, K., Feil, M., and Burke, L. B. (1999). Psychoactive medication prescribing practices for u.s. children: Gaps between research and clinical practice. Journal of the American Academy of Child & Adolescent Psychiatry, 38(5):557–565.

Jiang, M., Hughes, D. R., and Wang, W. (2018). The effect of Medicare's annual wellness visit on preventive care for the elderly. Preventive Medicine, 116:126–133.

Johannesen, N. and Stolper, T. (2017). The deterrence effect of whistleblowing: An event study of leaked customer information from banks in tax havens. CESifo Working Paper No. 6784.

Kahneman, D. and Tversky, A. (1973). On the psychology of prediction. Psychological Review, 80(4):237–251.

Kapetanios, G. (2008). A bootstrap procedure for panel data sets with many cross-sectional units. The Econometrics Journal, 11(2):377–395.

Kazdin, A. E. (2011). Evidence-based treatment research: Advances, limitations, and next steps. American Psychologist, 66(8):685–698.

Kazdin, A. E. (2017). Addressing the treatment gap: A key challenge for extending evidence-based psychosocial interventions. Behaviour Research and Therapy, 88:7–18.

Kazdin, A. E. (2018). Implementation and evaluation of treatments for children and adolescents with conduct problems: Findings, challenges, and future directions. Psychotherapy Research, 28(1):3–17.

Kazdin, A. E. and Blase, S. L. (2011). Rebooting psychotherapy research and practice to reduce the burden of mental illness. Perspectives on Psychological Science, 6(1):21–37.

Kazdin, A. E., Fitzsimmons-Craft, E. E., and Wilfley, D. E. (2017). Addressing critical gaps in the treatment of eating disorders. International Journal of Eating Disorders, 50(3):170–189.

Ketcham, J. D., Lucarelli, C., Miravete, E. J., and Roebuck, M. C. (2012). Sinking, swimming, or learning to swim in Medicare Part D. American Economic Review, 102(6):2639–2673.

KFF (2020). 2020 Employer health benefits survey. Technical report, Henry J. Kaiser Family Foundation, San Francisco, CA.

Khan, A. A., Murtaza, G., Khalid, M. F., and Khattak, F. (2019). Risk stratification for transcatheter aortic valve replacement. Cardiology Research, 10(6):323–330.

Kim, H. B., Lee, S. A., and Lim, W. (2019). Knowing is not half the battle: Impacts of information from the National Health Screening Program in Korea. Journal of Health Economics, 65:1–14.

King, M. S., Sharp, L., and Lipsky, M. S. (2001). Accuracy of CPT evaluation and management coding by family physicians. Journal of American Board Family Practice, 14(3):184–192.

Kosmerly, S., Waller, G., and Robinson, A. L. (2015). Clinician adherence to guidelines in the delivery of family-based therapy for eating disorders: Clinician Adherence to Guidelines in the Delivery of FBT. International Journal of Eating Disorders, 48(2):223–229.

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.

Kreider, B. (2010). Regression coefficient identification decay in the presence of infrequent classification errors. The Review of Economics and Statistics, 92(4):1017–1023.

Lakdawalla, D., Malani, A., and Reif, J. (2017). The insurance value of medical innovation. Journal of Public Economics, 145:94–102.

Layton, T. J. (2017). Imperfect risk adjustment, risk preferences, and sorting in competitive health insurance markets. Journal of Health Economics, 56:259–280.

Le Grange, D., Lock, J., Loeb, K., and Nicholls, D. (2010). Academy for eating disorders position paper: The role of the family in eating disorders. International Journal of Eating Disorders, 43(1):1.

Leon, M. B., Smith, C. R., Mack, M. J., Makkar, R. R., Svensson, L. G., Kodali, S. K., Thourani, V. H., Tuzcu, E. M., Miller, D. C., Herrmann, H. C., Doshi, D., Cohen, D. J., Pichard, A. D., Kapadia, S., Dewey, T., Babaliaros, V., Szeto, W. Y., Williams, M. R., Kereiakes, D., Zajarias, A., Greason, K. L., Whisenant, B. K., Hodson, R. W., Moses, J. W., Trento, A., Brown, D. L., Fearon, W. F., Pibarot, P., Hahn, R. T., Jaber, W. A., Anderson, W. N., Alu, M. C., and Webb, J. G. (2016). Transcatheter or surgical aortic-valve replacement in intermediate-risk patients. New England Journal of Medicine, 374(17):1609–1620.

Lewbel, A. (2007a). Estimation of average treatment effects with misclassification. Econometrica, 75(2):537–551.

Lewbel, A. (2007b). A local generalized method of moments estimator. Economics Letters, 94(1):124–128.

Lieber, E. M. (2017). Does it pay to know prices in health care? American Economic Journal: Economic Policy, 9(1):154–179.

Lock, J., Couturier, J., Bryson, S., and Agras, S. (2006). Predictors of dropout and remission in family therapy for adolescent anorexia nervosa in a randomized clinical trial. International Journal of Eating Disorders, 39(8):639–647.

Lock, J. and Grange, D. L. (2015). Treatment Manual for Anorexia Nervosa, Second Edition: A Family-Based Approach. Guilford Publications, New York, NY, US, 2 edition.

Loeb, K., Lock, J., Greif, R., and Le Grange, D. (2012). Transdiagnostic theory and application of family-based treatment for youth with eating disorders. Cognitive and Behavioral Practice, 19(1):17–30.

Mack, M. J., Leon, M. B., Thourani, V. H., Makkar, R., Kodali, S. K., Russo, M., Kapadia, S. R., Malaisrie, S. C., Cohen, D. J., Pibarot, P., Leipsic, J., Hahn, R. T., Blanke, P., Williams, M. R., McCabe, J. M., Brown, D. L., Babaliaros, V., Goldman, S., Szeto, W. Y., Genereux, P., Pershad, A., Pocock, S. J., Alu, M. C., Webb, J. G., Smith, C. R., and PARTNER 3 Investigators (2019). Transcatheter aortic-valve replacement with a balloon-expandable valve in low-risk patients. The New England Journal of Medicine, 380(18):1695–1705.

Maglione, M. and Hu, e. a. (2011). Off-label use of atypical antipsychotics: An update. Effective Health Care Review, 43:437.

Mahajan, A. (2006). Identification and estimation of regression models with misclassification. Econometrica, 74(3):631–665.

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

Millimet, D. L. (2010). The elephant in the corner: A cautionary tale about measurement error in treatment effects models. IZA Discussion Paper No. 5140.

Mitchell, E. M. (2020). Concentration of health expenditures and selected characteristics of high spenders, U.S. civilian noninstitutionalized population, 2017. Statistical Brief #521. Agency for Healthcare Research and Quality (US), Rockville, MD.

Mitchison, D., Hay, P., Slewa-Younan, S., and Mond, J. (2014). The changing demographic profile of eating disorder behaviors in the community. BMC Public Health, 14(1):943.

Mullainathan, S. and Obermeyer, Z. (2021). Diagnosing physician error: a machine learning approach to low-value health care. The Quarterly Journal of Economics, qjab046.

Murphy, K. M. and Topel, R. H. (2006). The value of health and longevity. Journal of Political Economy, 114(5):871–904.

Nishimura, R. A., Otto, C. M., Bonow, R. O., Carabello, B. A., Erwin, J. P., Guyton, R. A., O'Gara, P. T., Ruiz, C. E., Skubas, N. J., Sorajja, P., Sundt, T. M., and Thomas, J. D. (2014). 2014 AHA/ACC guideline for the management of patients with valvular heart disease. Journal of the American College of Cardiology, 63(22):e57–e185.

Obermeyer, Z., Powers, B., Vogeli, C., and Mullainathan, S. (2019). Dissecting racial bias in an algorithm used to manage the health of populations. Science, 366(6464):447–453.

O'Brien, S. M., Shahian, D. M., Filardo, G., Ferraris, V. A., Haan, C. K., Rich, J. B., Normand, S.-L. T., DeLong, E. R., Shewan, C. M., Dokholyan, R. S., Peterson, E. D., Edwards, F. H., and Anderson, R. P. (2009). The society of thoracic surgeons 2008 cardiac surgery risk models: Part 2—isolated valve surgery. The Annals of Thoracic Surgery, 88(1, Supplement):S23–S42.

Olfson, M., Druss, B. G., and Marcus, S. C. (2015). Trends in mental health care among children and adolescents. New England Journal of Medicine, 372.21:2029–2038.

Olshansky, S. J., Antonucci, T., Berkman, L., Binstock, R. H., Boersch-Supan, A., Cacioppo, J. T., Carnes, B. A., Carstensen, L. L., Fried, L. P., Goldman, D. P., Jackson, J., Kohli, M., Rother, J., Zheng, Y., and Rowe, J. (2012). Differences in life expectancy due to race and educational differences are widening, and many may not catch up. Health Affairs, 31(8):1803–1813.

Ortoleva, P. (2012). Modeling the change of paradigm: Non-Bayesian reactions to unexpected news. American Economic Review, 102(6):2410–2436.

Oster, E., Shoulson, I., and Dorsey, E. R. (2013). Optimal expectations and limited medical testing: Evidence from Huntington's Disease. American Economic Review, 103(2):804–830.

O'Connor, P. J. (2006). Improving medication adherence: Challenges for physicians, payers, and policy makers. Archives of Internal Medicine, 166(17):1802–1804.

Paserman, M. D. (2008). Job search and hyperbolic discounting: Structural estimation and policy evaluation. The Economic Journal, 118(531):1418–1452.

Popma, J. J., Deeb, G. M., Yakubov, S. J., Mumtaz, M., Gada, H., O'Hair, D., Bajwa, T., Heiser, J. C., Merhi, W., Kleiman, N. S., Askew, J., Sorajja, P., Rovin, J., Chetcuti, S. J., Adams, D. H., Teirstein, P. S., Zorn, G. L., Forrest, J. K., Tchétché, D., Resar, J., Walton, A., Piazza, N., Ramlawi, B., Robinson, N., Petrossian, G., Gleason, T. G., Oh, J. K., Boulware, M. J., Qiao, H., Mugglin, A. S., Reardon, M. J., and Evolut Low Risk Trial Investigators (2019). Transcatheter aortic-valve replacement with a self-expanding valve in low-risk patients. The New England Journal of Medicine, 380(18):1706–1715.

Posey, L. L. and Thistle, P. D. (2021). Genetic testing and genetic discrimination: Public policy when insurance becomes "too expensive". Journal of Health Economics, 77.

Rabideau, B., Eisenberg, M. D., Reid, R., and Sood, N. (2021). Effects of employer-offered high-deductible plans on low-value spending in the privately insured population. Journal of Health Economics, 76.

Rabin, M. (2013). Incorporating limited rationality into economics. Journal of Economic Literature, 51(2):528–543.

Rane, H. and Salem, S. (2012). Social media, social movements and the diffusion of ideas in the Arab uprisings. The Journal of International Communication, 18(1):97–111.

Revelt, D. and Train, K. (1998). Mixed logit with repeated choices: Households' choices of appliance efficiency level. Review of Economics and Statistics, 80(4):647–657.

Rios-Avila, F. and Naqvi, A. (2021). DRDID: Doubly-robust differences estimators for stata.

Rountree, M. A. and Pomeroy, E. C. (2010). Bridging the gaps among social justice, research, and practice. Social Work, 55(4):293–295.

Ryan, S. P. and Tucker, C. (2012). Heterogeneity and the dynamics of technology adoption. Quantitative Marketing and Economics, 10(1):63–109.

Rynes, S. L., Colbert, A. E., and Brown, K. G. (2002). HR professionals' beliefs about effective human resource practices: correspondence between research and practice. Human Resource Management, 41(2):149–174.

Sabety, A. (2020). The value of service sector relationships in health care. Working Paper.

Salcedo, A. (2021). He put off getting vaccinated. Now, he's in the ICU pleading for others to avoid his mistake: 'I messed up'. Washington Post.

Sant'Anna, P. H. C. and Zhao, J. (2020). Doubly robust difference-in-differences estimators. Journal of Econometrics, 219(1):101–122.

Shafer, P. R., Hoagland, A., and Hsu, H. E. (2021). Trends in pediatric wellness visits with out-of-pocket costs before and after the Affordable Care Act. JAMA Network Open.

Slejko, J. F., Ho, M., Anderson, H. D., Nair, K. V., Sullivan, P. W., and Campbell, J. D. (2014). Adherence to statins in primary prevention: Yearly adherence changes and outcomes. Journal of Managed Care Pharmacy, 20(1):51–57.

Smith, C. R., Leon, M. B., Mack, M. J., Miller, D. C., Moses, J. W., Svensson, L. G., Tuzcu, E. M., Webb, J. G., Fontana, G. P., Makkar, R. R., Williams, M., Dewey, T., Kapadia, S., Babaliaros, V., Thourani, V. H., Corso, P., Pichard, A. D., Bavaria, J. E., Herrmann, H. C., Akin, J. J., Anderson, W. N., Wang, D., and Pocock, S. J. (2011). Transcatheter versus surgical aortic-valve replacement in high-risk patients. New England Journal of Medicine, 364(23):2187–2198.

Song, Z. (2021). Social learning and health care utilization: Evidence from spousal health shocks. Working Paper.

Spinnewijn, J. (2015). Unemployed but optimistic: Optimal insurance design with biased beliefs. Journal of the European Economic Association, 13(1):130–167.

Strohman, E. (2014). An examination of research to practice gaps in education: Whole language, learning modalities, and brain gym. PhD Thesis, Liberty University.

Sun, L. and Abraham, S. (2021). Estimating dynamic treatment effects in event studies with heterogeneous treatment effects. Journal of Econometrics, 225(2):175–199.

Søndergaard, L., Popma, J. J., Reardon, M. J., Van Mieghem, N. M., Deeb, G. M., Kodali, S., George, I., Williams, M. R., Yakubov, S. J., Kappetein, A. P., Serruys, P. W., Grube, E., Schiltgen, M. B., Chang, Y., Engstrøm, T., and SURTAVI Trial Investigators (2019). Comparison of a complete percutaneous versus surgical approach to aortic valve replacement and revascularization in patients at intermediate surgical risk: results from the randomized SURTAVI trial. Circulation.

Tong, S. T., Webel, B. K., Donahue, E. E., Richards, A., Sabo, R. T., Brooks, E. M., Lail Kashiri, P., Huffstetler, A. N., Santana, S., Harris, L. M., and Krist, A. H. (2021). Understanding the value of the wellness visit: A descriptive study. American Journal of Preventive Medicine.

Train, K. E. (2009). Discrete Choice Methods With Simulation. Cambridge University Press, Cambridge, UK, 2 edition.

Ura, T. (2018). Heterogeneous treatment effects with mismeasured endogenous treatment. Quantitative Economics, 9(3):1335–1370.

van Ryn, M. and Burke, J. (2000). The effect of patient race and socio-economic status on physicians' perceptions of patients. Social Science & Medicine, 50(6):813–828.

Wandersman, A., Duffy, J., Flaspohler, P., Noonan, R., Lubell, K., Stillman, L., Blachman, M., Dunville, R., and Saul, J. (2008). Bridging the gap between prevention research and practice: The interactive systems framework for dissemination and implementation. American Journal of Community Psychology, 41(3-4).

Wolfson, D., Santa, J., and Slass, L. (2014). Engaging physicians and consumers in conversations about treatment overuse and waste: A short history of the Choosing Wisely campaign. Academic Medicine, 89(7):990–995.

Xie, Y., Brand, J. E., and Jann, B. (2012). Estimating heterogeneous treatment effects with observational data. Sociological Methodology, 42(1):314–347.

Yanagi, T. (2018). Inference on local average treatment effects for misclassified treatment. Working Paper.

Young, H. P. (2009). Innovation diffusion in heterogeneous populations: Contagion, social influence, and social learning. American Economic Review, 99(5):1899–1924.

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.

CURRICULUM VITAE