Essays on the Industrial Organization of Healthcare Markets

by

Cici McNamara

A dissertation submitted in partial fulfillment of
the requirements for the degree of

Doctor of Philosophy

(Economics)

at the

UNIVERSITY OF WISCONSIN–MADISON

2022

Date of final oral examination: 05/06/2022

The dissertation is approved by the following members of the Final Oral Committee:
Alan Sorensen, Professor, Economics
Kenneth Hendricks, Professor, Economics
Christopher Sullivan, Assistant Professor, Economics
John Mullahy, Professor, Population Health Sciences

Table of Contents

Abstract

Electronic health record (EHR) alerts can serve as nudges that remind physicians of relevant clinical guidelines and inform them when their patients are eligible for recommended services. This paper examines another possible mechanism by which EHR alerts might influence physician behavior, which is by validating new technologies. I analyze the effects of an EHR alert for colorectal cancer screening implemented at the primary care clinics of a large academic health system. This alert functioned both as a standard nudge and implicitly validated a new screening technology. I find that the implementation of the alert more than doubled the new technology's share of screenings and increased overall screening rates by 29%. I use a nested logit model to quantify how much the nudge and validation contribute to the increase in screening rates and find that the nudge accounts for 80% of the increase.

JEL No. D91, I12, I18, O33
Keywords: Preferences, Cancer, Health Behavior, Public Health, Choice of Technology

Abstract

This paper estimates the causal effect of the expansion of Colombia's national prescription drug formulary to include five new types of insulin on the healthcare utilization and costs of type I diabetics, and identifies the mechanism through which outpatient cost reductions are realized. We find that expanded coverage generates an increase in the cost of insulin for type I diabetics equal to 17% of their average baseline healthcare costs. At the same time, annual outpatient care utilization falls by 1.9 claims. We devise tests to explore the relative importance of two mechanisms by which the expansion may have lowered non-drug

healthcare utilization: spillovers from drug to non-drug spending and rationing of care. We find no evidence that the formulary expansion reduces the rate of complications from diabetes, and find substantial declines in non-drug costs even among the subset of diabetics with no scope for spillovers. We find large reductions in the utilization of discretionary care including diagnostic tests, but no such declines for the use of essential drugs, suggesting that rationing of care is the primary driver of observed cost savings.

JEL No. I100, I110, I130, I180
Keywords: Healthcare rationing, prescription drug formulary, diabetes, insulin, health insurance

Abstract

Healthcare providers have been shown to be responsive to the financial incentives introduced by capitation and fee-for-service contracts. In most settings, one of these contracts applies to all the services that a provider renders for the patients of a given insurer. In this paper, we examine the extent to which providers are responsive to contracts that vary at the service level and generate profit variation across patients. We develop a Bayesian learning model where primary care providers (PCP) have prior beliefs regarding a patient's profitability, which are a function of service contracts. These beliefs are updated based on the patient's use of services under the PCP's care. We apply this model to data on all pregnant women enrolled to Colombia's statutory health care system in 2011. We find that PCPs are more likely to refer to a different hospital patients whose insurers have lower expected reimbursement for obstetric care. We find no evidence that PCP referral decisions are updated based on the usage and profitability of a service conditional on the insurer, suggesting that the negotiation of contract types at the insurer level does not meaningfully change the way in which providers determine which patients to retain.

Keywords: Physician moral hazard; Capitation; Fee-for-service; Health insurance.
JEL codes: I12, I13, I18, D80.

1 Health Information Technology and Innovation Diffusion in Primary Care Settings

1.1 Introduction

Improvements in health information technology (HIT) have provided new ways to influence physician behavior. Perhaps no HIT is more widely used by physicians than the electronic health record (EHR), the digital version of a patient's chart which contains their medical and treatment histories, diagnoses, test results, and demographic information. Use of EHRs by office-based physicians more than doubled from 42 to 86 percent between 2008 and 2017 (Office of the National Coordinator for Health Information Technology, 2019). Efforts to use the EHR to affect physicians' decisions have largely manifested themselves in the EHR as *hard stops* and *nudges*. Hard stops such as computerized medication alerts warn physicians of unsafe practices and prevent them from continuing the patient's course of care until there is manual indication that the unsafe practice has been resolved. Nudges inform a physician that a practice is recommended but do not force compliance with the recommendation. Both hard stops and nudges are targeted alerts, drawing on information from the EHR to identify patients for whom the alert is relevant. In this paper, I ask whether there are other ways that the EHR might improve clinical quality aside from forcing or encouraging compliance with clinical guidelines.

Specifically, I ask whether the EHR can improve clinical quality by accelerating the adoption of new technologies.

I answer this question in the context of scheduling for colorectal cancer screening (CRC) at the primary care clinics of the University of Wisconsin Health System (UW Health), a large academic health system that serves over 700,000 patients each year. In the spring of 2010, the health system upgraded its EHR to implement an alert that, using information on patients' demographics and screening histories, informed primary care providers (PCPs) when a patient was eligible and due for CRC screening and listed the technologies by which the patient could be screened. One of the two technologies listed as part of this alert, virtual colonoscopy (VC), had not yet been endorsed by any official body or the health system. The alert therefore provided initial validation of VC and had the potential to act not only as a nudge, but also as a signal about the quality of a seldom-used new technology. The alert's function as a nudge had the potential to increase screening rates directly by increasing awareness of the CRC screening guideline and by prioritizing CRC screening during the PCP appointment. The alert's validation of VC also had the potential to increase screening rates by providing patients with a sanctioned and meaningfully different alternative to the dominant screening technology, optical colonoscopy. In this paper, I quantify the relative importance of these two effects, which sheds light on what the barriers to screening participation are and what policies are most effec-

tive at increasing participation.

I develop a nested logit model of screening participation and screening technology choices to estimate the effect of the alert on both screening rates and adoption of VC. I find that the validation of new technologies has only a small effect the extensive margin decision of whether to participate in screening despite meaningfully increasing the new technology's share of screenings. Put differently, screening technologies are much closer substitutes for one another than they are for the outside option of not participating in screening. I find that the implementation of the EHR alert increased VC's share of screenings by over 12 percentage points (p.p.) and screening rates by 0.5 p.p. The alert's validation of VC not only increased its attractiveness relative to the most commonly used modality, optical colonoscopy (OC), but also increased the expected utility of the modality choice. It is by this latter channel that the validation has scope to impact the extensive margin through what I refer to as the "validation effect." The alert also impacts the extensive margin directly through what I term a "nudge effect." I estimate a version of my nested logit model that allows both the nudge and validation effects of the alert to vary across patients according to their likelihood of participating in screening in the period preceding the implementation of the alert. Patients least likely to participate in the baseline period are older, non-white, and have multiple comorbidities. I find that both the nudge and validation effects are greater among patients with lower

baseline screening propensities. I use the model to disentangle the contributions of the nudge and validation effects to the increase in screening rates and find that the validation effect accounts for 20% of the increase and the nudge effect for the remaining 80%.

By simply informing physicians of a patient's eligibility for a preventive care service, EHR alerts satisfy the recently popularized definition of a nudge as "any aspect of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler and Sustein, 2008). There is a large economic literature that quantifies the effects of nudges on decisions ranging from household energy consumption (Allcott and Mullainathan, 2010) to job choice (Coffman and Kessler, 2017) to health insurance enrollment (Handel, 2013). There is also a large medical literature examining the effectiveness of both physician- and patient-directed nudges aimed at increasing adherence with preventive care guidelines. Several studies have looked at the effects of such interventions on rates of colon cancer screening specifically; see Dougherty et al. (2018) for a detailed summary. Results from randomized control trials of physician-directed reminders for CRC screening have found positive though in some cases insignificant effects (Levy et al., 2013; Sequist et al., 2009). Bai et al. (2021) examines the effectiveness of commitment devices, which are another tool of behavioral economics, at increasing participation in preventive care

and finds no impact on doctor visits or health outcomes. The alert considered in this paper is novel because it not only functioned as a standard nudge but also provided implicit validation of a modality for screening other than optical colonoscopy, the long-standing and widely dominant method of CRC screening. This paper's main contribution is in disentangling the marginal benefit of this novel element of the alert from the benefit generated by the nudge.

This paper is also related to the literature on the effects of HIT adoption. Papers looking at the impact of EHR adoption and alert utilization have found no significant effects on costs, mortality, or preventive care utilization (Agha, 2014; Rodrigues Llorian and Mason, 2021). My study differs from those in this literature by focusing on a setting in which the EHR had long been in use. Papers estimating the impact of HIT adoption have posited that there may be a learning curve associated with HIT adoption that prevents any positive impact in the short run. The learning curve for physicians associated with the EHR upgrade considered in this paper is likely far less steep than that faced by physicians transitioning from paper to electronic records. I can therefore reasonably expect effects to manifest in the short run. In addition, this paper focuses on the impact of an alert for a specific preventive care service on the utilization of that service among patients due according to official medical guidelines. It therefore speaks less to the general potential of HIT to improve quality across dimensions of healthcare and more to the power of

HIT to target information to those with the greatest potential benefit.

Lastly, this paper belongs to the health economics literature that examines the patterns and determinants of technology adoption in clinical settings. Crawford and Shum (2005) and Coscelli and Shum (2004) model physician's gradual adoption of a newly introduced antacid as a Bayesian learning process. Grigolon et al. (2021) quantifies the extent to which the level of social stigma faced by patients impacts their decision to participate in lung cancer treatment and by extension the adoption of innovative lung cancer treatment. The intervention considered here has the potential to increase the adoption of innovative modalities not only on the extensive margin as in Grigolon et al. (2021) but also on the intensive margin as impressions of VC improve. My paper contributes to this literature by showing how a health system's initial validation of a technology impacts its utilization. My finding that the health system's validation of VC doubled its utilization complements Chen et al. (2021), which shows that the 2016 validation of VC by the U.S. Preventive Services Task Force increased its monthly utilization among the privately insured from 0.4 to 0.6 procedures per 100,000.

Preventive care has the potential to reduce the rates of chronic diseases including cancer, heart disease, and diabetes. Despite the fact that chronic diseases are the leading cause of disability and mortality in the U.S., less than 8% of adults nationwide are up to date with all recommended preventive care services (Levine

et al., 2019). There is a pressing need to understand why patients do not participate in preventive care and what strategies are effective at increasing participation. Understanding how HIT can impact screening decisions will likely be of particular importance for CRC screening in the coming years, as official clinical guidelines for CRC screening were recently revised to extend the eligibility criteria to include persons between the ages of 45 and 50 (Lin et al., 2021). As the share of patients eligible for screening grows, the EHR becomes an increasingly cost effective tool relative to interventions whose costs scale with the number of due patients. This paper shows that the EHR can be used to speed the diffusion of new technologies, improving patients' choice sets and increasing screening participation.

1.2 Background & setting

Among all cancers, colorectal cancer (CRC) is third in incidence and mortality for both men and women in the U.S. (Siegel et al., 2019). The lifetime risk of acquiring CRC in the United States is about 4.2 percent and survival largely depends on the stage of the cancer at the time of diagnosis. Patients with localized disease at diagnosis have a 5-year survival rate of 90 percent. The same survival rate for patients diagnosed with cancer that has spread to regional lymph nodes is 71 percent, and that for patients diagnosed with distantly metastasized cancer is only 14 percent (NCI Surveillance, Epidemiology, and End Results Program, 2019). Incidence of

CRC is increasing in age, nearly half of all new cancers being diagnosed in patients between the ages of 65 and 84. Male sex and Black race are also associated with higher rates of CRC. While a family history of CRC that is not attributable to any known inherited syndromes is a well-established risk factor for CRC, in practice definitions of family history vary substantially across studies (Lin et al., 2021).

Unlike many cancers, there are multiple types of modalities that screen for CRC, including direct visualization, stool-based, serum-based, and urine-based tests (Lin et al., 2017). Optical colonoscopy (OC) is the most commonly used screening modality in the U.S. (de Moor et al., 2018; Zapka et al., 2012). OC is completed by using a flexible, lighted tube to screen the rectum and entire colon. Another much more recently developed direct visualization test is virtual colonoscopy (VC), also known as computed tomographic colonography (CTC). VC is performed by taking computed tomography images of the abdomen and pelvis and is therefore less invasive than OC. Screening technologies vary not only in terms of their invasiveness but also in terms of their specificity, sensitivity, and recommended frequency (Zauber et al., 2016). Table 1.1 summarizes these characteristics for the most commonly used screening modalities.

The screening recommendation of United States Preventive Services Task Force (USPSTF), an independent, volunteer group of national experts in prevention and evidence-based medicine established by Congress in 1984, shapes clinical prac-

tice and coverage decisions by determining the eligibility criteria and set of validated modalities for preventive care services (Agency for Healthcare Research and Quality, 2019; Siu et al., 2015). The task force's 2008 recommendation for CRC screening, presented in table 1.2, advised that adults between the age of 50 and 75 be screened for colon cancer. The task force did not include VC in its list of validated modalities, citing concerns about the frequency of extra-colonic findings and harms of radiation exposure during VC (US Preventive Services Task Force, 2008). In the years that followed, research showed the costs of diagnostic workup of these incidental findings to be quite low, within the range of 24-35 USD (Pickhardt et al., 2008). Additionally, the radiation-related cancer risk from VC was found to be more than offset by the number of colorectal cancers prevented (Berrington de González et al., 2011). In response to this new evidence, the task force added VC to its list of validated modalities in its 2016 recommendation (US Preventive Services Task Force, 2016).

Table 1.1: Summary of CRC screening modalities

Modality	**Description**	**Frequency**	**True negative rate**	**True positive rate**			
				polyps $\leqslant$ 5 mm	**polyps 6 to 9 mm**	**polyps $\geqslant$ 10 mm**	**colorectal cancer**
Optical colonoscopy	Long, thin, flexible, lighted tube used to screen rectum and entire colon.	10 years	0.86	0.75	0.85	0.95	0.95
Virtual colonoscopy	CT images of entire colon taken and analyzed.	5 years	0.88	0	0.57	0.84	0.84
Flexible sigmoidoscopy	Short, thin, flexible, lighted tube used to screen rectum and lower third of the colon.	5 years	0.87	0.75	0.85	0.95	0.95
Fecal occult blood test	Stool samples obtained by patient at home and analyzed in lab.	1 year	0.93	0.08	0.12	0.24	0.70

Sources: 1. Zauber A, Knudsen A, Rutter CM, et al. Evaluating the Benefits and Harms of Colorectal Cancer Screening Strategies: A Collaborative Modeling Approach. AHRQ Publication No. 14-05203-EF-2. Rockville, MD: Agency for Healthcare Research and Quality; October 2015.

2. Knudsen AB, Zauber AG, Rutter CM, et al. Estimation of Benefits, Burden, and Harms of Colorectal Cancer Screening Strategies: Modeling Study for the US Preventive Services Task Force. JAMA 2016; 315:2595.

Table 1.2: USPSTF 2008 Colorectal Cancer Screening Recommendation Summary

Population	**Recommendation**
Adults, beginning at age 50 years and continuing until age 75 years	The USPSTF recommends screening for colorectal cancer using fecal occult blood testing, sigmoidoscopy, or colonoscopy in adults, beginning at age 50 years and continuing until age 75 years. The risks and benefits of these screening methods vary.
Adults age 76 to 85 years	The USPSTF recommends against routine screening for colorectal cancer in adults 76 to 85 years of age. There may be considerations that support colorectal cancer screening in an individual patient.
Adults older than age 85 years	The USPSTF recommends against screening for colorectal cancer in adults older than age 85 years.

Source: U.S. Preventive Services Task Force (2008). Screening for Colorectal Cancer: U.S. Preventive Services Task Force Recommendation Statement *Annals of Internal Medicine* 149(9):627-638.

In the spring of 2010, when VC was an available but not yet officially validated modality, the primary care clinics of the University of Wisconsin Health System upgraded their EHR, implementing an alert for CRC screening in an effort to improve screening rates. The alert in question flagged patients that were eligible and due for screening. This was the second EHR alert to be implemented in this system, the first being for breast cancer screening. This alert not only notified the PCP that

the patient was due for screening but also informed the PCP that the patient could be screened by either VC or OC. All PCPs that were part of the UW Health system had open access to VC and most of the private insurance providers of UW Health patients covered VC during this period. There is no evidence that UW Health's inclusion of VC in the alerts was part of a wider effort to increase usage of VC. Figure 1.A1 shows the PubMed and Google search trends for the phrase "virtual colonoscopy" over the past fifteen years; there is no notable increase in interest in VC leading up to the time of the intervention in May of 2010.

1.3 Data

I employ retrospective EHR data for all patients managed by UW Health who were both eligible and due for CRC screening at some point between January 2009 and December 2011. A patient is considered managed by the health system if they have had at least two PCP visits within the health system in the past 36 months and at least one in the past 24 months (Wisconsin Collaborative for Healthcare Quality, 2011). Patients are considered *eligible* for CRC screening if they are between 50 and 75 years of age and have not had a total colectomy. Patients are considered *due* for screening if they have not been screened by OC in the past ten years, VC or flexible sigmoidoscopy in the past five years, or fecal occult blood test in the past year. I also observe visits to other health systems in Wisconsin by patients that

satisfy the definition of being currently managed by UW Health. While these data are too sparse to use in my main analyses, I use them to examine aggregate trends in modality shares in subsection 1.4. All other analyses including the summary statistics below use only observations corresponding to PCP visits at UW Health clinics.

I observe the dates of patients' PCP visits and CRC screenings, as well as the physician and clinic corresponding to these events. Note that for any given patient, I only observe visits and screenings that take place during the periods when a patient is both eligible and due. I therefore do not observe events for patients once they are beyond the age of 75. Neither do I observe visits or screenings that take place within ten years of an OC or five years of a VC. Patients who become due during my sample period are observable only after becoming due, and patients who get screened during my sample period drop out of the sample following their screening. During my sample period, the share of currently managed UW Health patients that were eligible for and up-to-date with CRC screening was approximately 70% (Wisconsin Collaborative for Healthcare Quality, 2021b). This implies that my sample of currently managed, eligible, and due patients accounts for around 30% of UW Health's eligible patients.

An observation in my data is a visit and my outcome of interest is whether each visit results in CRC screening. I do not observe the actual scheduling of screenings

and so assume that all CRC screenings that take place within 9 months of a PCP visit were scheduled during that visit.[1] It is likely the case that some of these visits are "follow-up visits" that are focused on the treatment of a condition established in the previous visit. In order to restrict my sample to the set of primary care visits during which a full assessment of the patient's preventive care needs was likely to have been performed, I drop follow-up visits – which I define as visits that occur within two weeks of a previous visit for the same patient – and assess whether the initial visit resulted in CRC screening.

For each visit, I observe several demographic characteristics of the patient, the name of the patient's insurance carrier, a set of 29 indicators for the patient's Elixhauser comorbidities, and the patient's Adjusted Clinical Group (ACG) score, which is a measure of expected healthcare costs (Johns Hopkins University Health Services Research & Development Center, 2008). I also observe characteristics of the patient's place of residence, including the fraction of individuals in the patient's zip code that have at least a high school education; the fraction of individuals in the patient's zip code that live below the poverty line; and the RUCA code of the patient's census tract, which is a classification based on population density, urbanization, and daily commuting patterns. Finally, I observe the specialty and sex of each physician. I drop from the sample visits corresponding to insurance carriers,

[1] Results throughout are robust to using either a 6 or 12 month threshold rather than a 9 month one.

comorbidities, and clinics that are only observed either in the period preceding or following the alert.

Table 1.3: Summary statistics of sample stratified by time of EHR alert implementation, Part 1

	Full sample period	Baseline	Post-EHR alerts	p-value
Outcomes				
Number of visits	93,602	49,175	44,427	
Result in any screening	1,514 (1.6%)	776 (1.6%)	738 (1.7%)	0.3
Screening by optical colonoscopy	1,039 (1.1%)	624 (1.3%)	415 (0.9%)	<0.001
Screening by virtual colonoscopy	409 (0.4%)	112 (0.2%)	297 (0.7%)	<0.001
PCP characteristics				
Number of visits per PCP	1,075 (753)	565 (426)	510 (337)	
Female	50,296 (53.7%)	25,932 (52.7%)	24,364 (54.8%)	<0.001
Specialty				<0.001
Internal medicine	62,863 (67.2%)	33,497 (68.1%)	29,366 (66.1%)	
Family medicine	29,530 (31.5%)	14,941 (30.4%)	14,589 (32.8%)	
Other	1,209 (1.3%)	737 (1.5%)	472 (1.1%)	
Patient characteristics				
Number of visits per patient	3.7 (2.3)	2.4 (1.3)	2.2 (1.2)	
Prior screening	420 (0.4%)	173 (0.4%)	247 (0.6%)	<0.001

Note: Cell values are presented as mean (standard deviation) for continuous measures, and n (%) for categorical measures. p-values for comparison of baseline and post-EHR alert periods computed using Wilcoxon rank-sum test for continuous variables and Pearson's chi-squared test for categorical and binary variables. All statistics are calculated across visits with the exception of number of visits per patient and number of visits per PCP, which are computed across patients and PCPs respectively.

Table 1.4: Summary statistics of sample stratified by time of EHR alert implementation, Part 2

	Full sample period	Baseline	Post-EHR alerts	p-value
Patient characteristics				
Age group				<0.001
[50,55]	21,224 (22.7%)	11,219 (22.8%)	10,005 (22.5%)	
(55,60]	21,720 (23.2%)	11,727 (23.8%)	9,993 (22.5%)	
(60,65]	19,193 (20.5%)	10,153 (20.6%)	9,040 (20.3%)	
(65,70]	17,873 (19.1%)	9,053 (18.4%)	8,820 (19.9%)	
(70,75]	13,592 (14.5%)	7,023 (14.3%)	6,569 (14.8%)	
Number of comorbidities				<0.001
None	67,593 (72.2%)	38,126 (77.5%)	29,467 (66.3%)	
One	17,007 (18.2%)	7,015 (14.3%)	9,992 (22.5%)	
Two	6,217 (6.6%)	2,696 (5.5%)	3,521 (7.9%)	
Three or more	2,785 (3.0%)	1,338 (2.7%)	1,447 (3.3%)	
Female	65,760 (70.3%)	34,448 (70.1%)	31,312 (70.5%)	0.2
Married	30,551 (32.6%)	16,602 (33.8%)	13,949 (31.4%)	<0.001
Private insurance	64,851 (69.3%)	34,477 (70.1%)	30,374 (68.4%)	<0.001
Non-white	4,953 (5.3%)	2,673 (5.4%)	2,280 (5.1%)	0.04
Zip code fraction $\geqslant$ H.S.	0.94 (0.03)	0.94 (0.03)	0.94 (0.03)	0.1
Zip code fraction in poverty	0.10 (0.09)	0.10 (0.09)	0.10 (0.09)	0.002

Note: Cell values are presented as mean (standard deviation) for continuous measures, and n (%) for categorical measures. p-values for comparison of baseline and post-EHR alert periods computed using Wilcoxon rank-sum test for continuous variables and Pearson's chi-squared test for categorical and binary variables. All statistics are calculated across visits with the exception of number of visits per patient and number of visits per PCP, which are computed across patients and PCPs respectively.

Summary statistics for my sample of visits are provided in tables 1.3 and 1.4. There are 25,641 unique patients, who have an average of 3.7 primary care visits during the sample period. Patients are on average 61 years of age. 5% of patients are under 51 years of age and therefore newly eligible for CRC screening. While VC was covered by nearly all of the private insurance carriers in my sample during this period, it was not covered by Medicaid or Medicare. 30% of the sample therefore had insurance coverage only for OC, while the remainder had coverage for both modalities. Over two-thirds of the sample is female, which is consistent with women being 33 percent more likely to visit the doctor than men (Centers for Disease Control and Prevention, 2001). Over two-thirds of visits correspond to patients with no comorbidities. The most common comorbidity is hypertension, the rate of which is 20%. Depression and diabetes are the next most common comorbidities, each with a rate of 6%. There are 87 unique PCPs, about half of which are female. Most PCPs specialize in internal medicine. Family medicine is the next most common specialty. These PCPs work across 16 unique clinics.

Turning to summary statistics for my outcomes of interest, I see that 1.6% of PCP visits for this sample result in screening within 9 months. There is no statistically significant difference in the overall screening rate between the periods preceding and following the implementation of the EHR alerts. There is dramatic substitution away from OC toward VC across the two periods: OC's share of screenings

falls from 84% in the baseline period to 58% after the alert's implementation. 96% of all screenings are completed by either OC or VC. For this reason, I focus on only these two modalities and omit from our analyses screening by all other modalities, including stool tests.

Healthcare quality with respect to the provision of preventive care is typically measured by the share of patients who are up-to-date with screening. As mentioned, this share was around 70% within the UW Health system during during this period, which was a bit above the national average of 66% (Centers for Disease Control and Prevention, 2021). These shares are markedly larger than the screening rate in table 1.3. One reason why the screening rate in this sample is so low is that I focus only on patients who are already due for screening. Patients who remain continually up-to-date with screening and never become due do not appear in my sample, and their screenings do not figure into the rate shown in table 1.3.

Another reason why my screening rate is so much lower than the share of patients up-to-date with screening stems from the fact that the former measure is computed at the visit level, whereas the latter is computed at the patient level. Also, while a person who is screened for CRC is included in the share of patients up-to-date with screening for up to ten years, patients who are screened in my sample disappear after the visit during which screening was scheduled. There is therefore persistence in the share of patients up-to-date with screening that is not exhibited

by my screening rate measure. Additionally, note that only 0.5% of patients in my sample have ever been screened before. This is consistent with the literature that has examined longitudinal predictors of screening participation, which has found that patients who have never participated in screening are less likely to be screened than patients who have been screened before (Murphy et al., 2014; Green et al., 2017). Finally, note that both my measure of previous screening as well as my measure of screening rates only account for screenings that were or are performed within the health system. The inclusion of patients who have been screened elsewhere and are not in fact due for screening will bias downward my screening measures. Limiting the sample to those who are currently managed by the health system helps mitigate but may not completely prevent the inclusion of such patients.

1.4 Preliminary evidence

In this subsection, I present evidence that the EHR alert implemented at UW Health primary care clinics in the spring of 2010 had an effect both on screening rates and technology choice. In figure 1.1, I plot the share of screenings between 2007 and 2011 that are completed by VC and OC. I stratify these trends by whether the PCP visit corresponding to the screening took place at a treated UW Health clinic, or another clinic belonging to the Wisconsin Collaborative for Healthcare Quality, an organization that collects data from and publishes performance reports on more than

325 health systems and over 65% of PCPs in Wisconsin (Wisconsin Collaborative for Healthcare Quality, 2021a). The vast majority of screenings prior to 2010 are completed by OC, and there is no obvious trend in technology shares. At the time of the EHR alert implementation in 2010, there is a dramatic increase in VC's share of screenings at UW Health clinics. There is no such change observable for other WCHQ clinics.

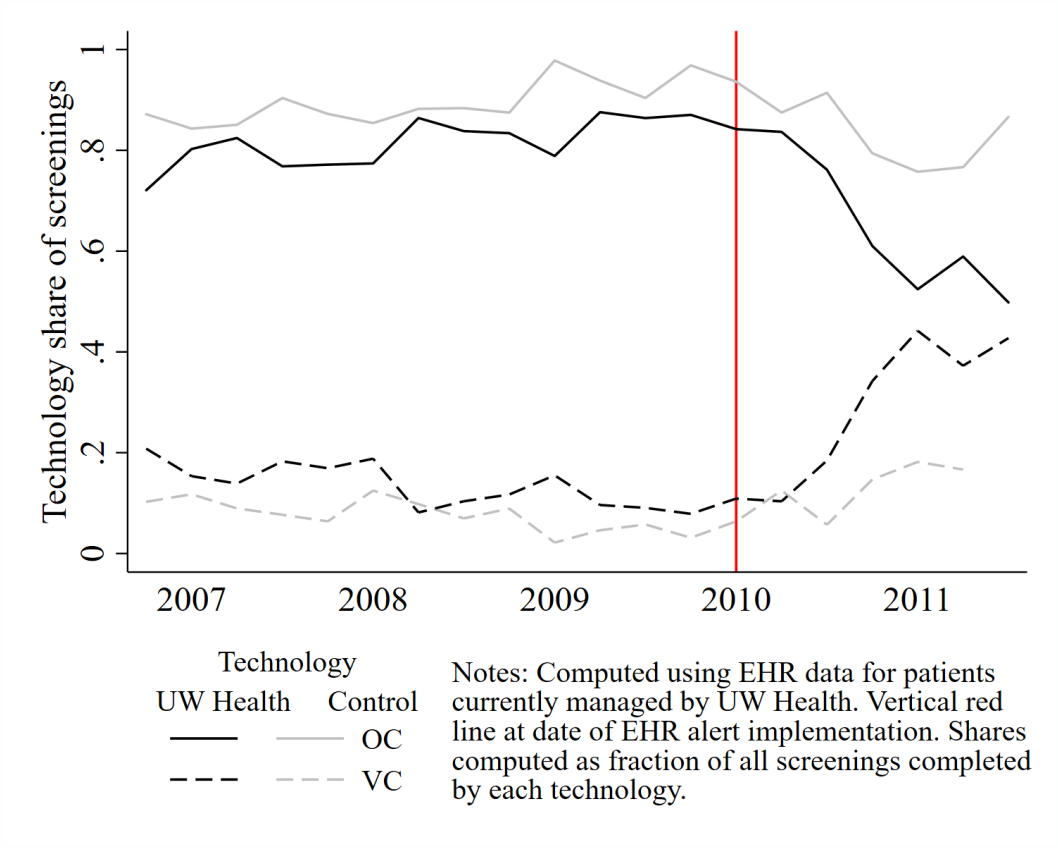
I more rigorously estimate the impact of the alerts on the intensive and extensive margins of the screening decision by estimating a probit model with sample selection. This model accounts for the fact that I only observe the modality choices of patients who participate in screening, and that patients who are unlikely to participate in screening may make different modality choices than those who are likely to participate. Let m_{ipt} be a binary indicator equalling one if patient i chooses to schedule screening by VC during her appointment with PCP p in period t and zero if they choose screening by OC. The probit model of the screening technology choice is given by

$$m_{ipt} = (\alpha + \beta_{\tau_t} + z_{it}'\gamma + \eta_p + \varepsilon_{ipt} > 0). \tag{1}$$

τ_t are quarter fixed effects, where the quarter prior to the EHR alert implementation is the reference category. z_{it} are time-varying patient characteristics including age group, number of comorbidities, insurance type, sex, race, marital status, ACG

score, and an indicator for whether the patient has ever been screened before. η_p are PCP fixed effects, and $\varepsilon \sim N(0,1)$.

Figure 1.1: Aggregate trends in screening technology shares



The screening technology choice is only observed conditional on screening participation. Let s_{ipt} be a binary indicator equalling one if patient i chooses to schedule screening during her appointment with PCP p in period t and zero otherwise. The selection equation is given by

$$s_{ipt} = (\phi + \delta_{\tau_t} + x_{it}'\gamma + \zeta_p + \omega_{ipt} > 0). \tag{2}$$

Here, ζ_p are physician fixed effects. $\omega \sim N(0,1)$, and $\text{corr}(\varepsilon, \omega) = \rho$.

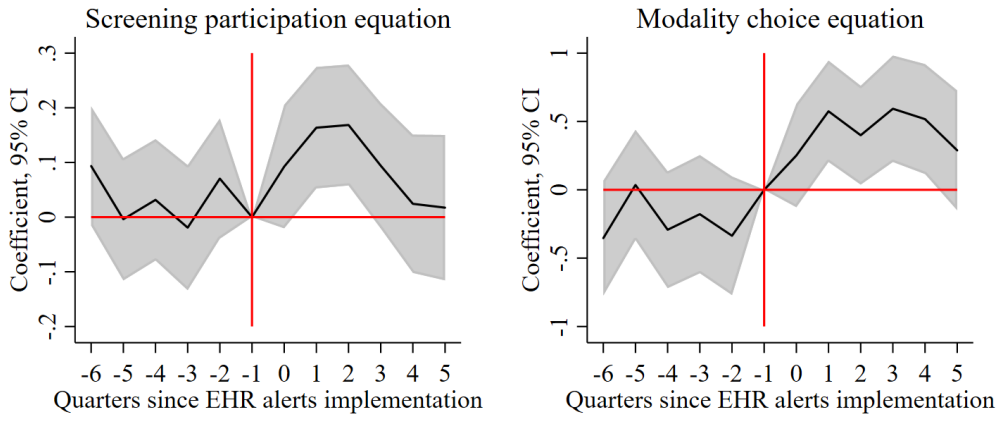
x_{it} contains all of the characteristics included in z_{it}, as well as two covariates which have been shown to influence the screening participation decision but not the technology choice decision. These covariates will act as instruments for selection into screening. The first is an indicator for whether the patient is newly eligible for screening; i.e. whether they are between 50 and 51 years of age. During the first five years of eligibility, rates of screening initiation have been shown to be highest in the first year of eligibility and to fall thereafter (Fedewa et al., 2017; Wernli et al., 2014). This is likely a result of both CRC awareness campaigns highlighting the age 50 eligibility threshold and of the fact that physicians can rightfully be assume that any patient who is 50 years old has never been screened for CRC and is therefore due for screening. The second is the patient's Rural-Urban Commuting Area (RUCA) code, which captures the degree of commuting done by residents of each Census tract as well as the rurality of the tract. RUCA codes are therefore a good measure of the transportation costs associated with screening completion, which have been shown to be a major barrier to participation (Jones et al., 2010).

I estimate this model by maximum likelihood, clustering standard errors at the clinic level. This design supports causal inference if there are sufficient observations

to control for underlying trends in my outcomes of interest, and if there are no concurrent events that may confound my estimation of the effect of the alert. I will examine pre-trends below and know of no confounding events that took place during my sample period. I do not employ the non-UW Health WCHQ clinics in my analyses, as the rate of missing data is too high and the number of observations too low to allow reliable estimation of the covariates of interest.

I plot the estimated quarters-since-alert fixed effects from equations 1 and 2 in the left and right hand panels of figure 1.2 respectively. Looking first at the estimates for the screening participation selection equation, I see no evidence of a pre-trend in screening rates in the quarters preceding the alert and a statistically significant increase in screening rates in the year following its implementation. Screening rates are approximately 0.5 p.p (29%) higher in the post-alert period than they were in the period before the alert's implementation. This increase does not last, screening rates seeming to return to their pre-alert level by the end of the sample period.

Figure 1.2: Quarters-since-EHR alert estimates from probit model with sample selection



Wald test of independent equations ($\rho = 0$): $\chi^2_2 = 1.9$, p-value = 0.2
F test of joint insignificance of instruments: $\chi^2 = 250.1$, p-value < 0.001
Notes: Black lines illustrate the estimated coefficents for the quarter-year fixed effects of a probit model of screening technology choice with sample selection. 95% confidence intervals displayed as gray shaded area. Lefthand panel shows estimates from selection equation for participating in screening. Righthand panel shows estimates from the probit equation for choosing to be screened by VC conditional on participating in CRC screening. Horizontal red lines at zero, and vertical red lines at quarter prior to implementation of EHR alert.

Turning to the estimates for the screening technology choice model in the right hand panel of figure 1.2, I see that there is little evidence of a pre-trend in VC utilization in the quarters preceding the alert's implementation. I also see that there is a dramatic increase in VC utilization following the implementation of the EHR alert, its share of screenings doubling from 12% to 24%. Estimation results for the remainder of the covariates included in this model are presented in table 1.A1.

I estimate a correlation between the structural error terms of $\rho = 0.3$. The χ^2 of a Wald test of independent equations is 1.77, and I cannot reject the null hypothesis that there is no correlation between these error terms. These results provide lim-

ited evidence of selection into screening and help justify my forthcoming modeling assumption that random components of the screening participation and modality choice decisions are independent. A test for joint significance of the instruments for selection into screening yields a χ^2 of 250 and strongly rejects the null hypothesis that they have no effect on the screening decision.

Having presented evidence that the EHR alert had a causal effect on screening rates and technology choice, I now turn to quantifying how much the validation of VC contributed to the increase in screening rates. I do so by developing a model of screening participation and technology choice that takes into account how the technology choice decision influences the screening participation decision. Note that while the nudge effect generates substitution away from non-participation toward both OC and VC, the validation effect generates substitution away from both non-participation and OC toward VC. This pattern alone allows us to estimate the relative size of each effect without modeling the screening decision. However, this reduced form approach will not allow us to say anything how improvements to the screening choice set other than the validation of VC can be expected to impact screening participation. For instance, if I find that the validation effect explains little of the increase in screening rates, I will not be able to say whether this is because the validation did not constitute a meaningful improvement to the choice set or because patients do not consider the attractiveness of available screening alter-

natives when deciding whether to participate in screening. This would greatly limit our understanding of how best to increase screening participation. The model that I develop will allow us both to disentangle the nudge and validation effects and also to say how future improvements in screening technologies are likely to affect the extensive margin.

1.5 Model

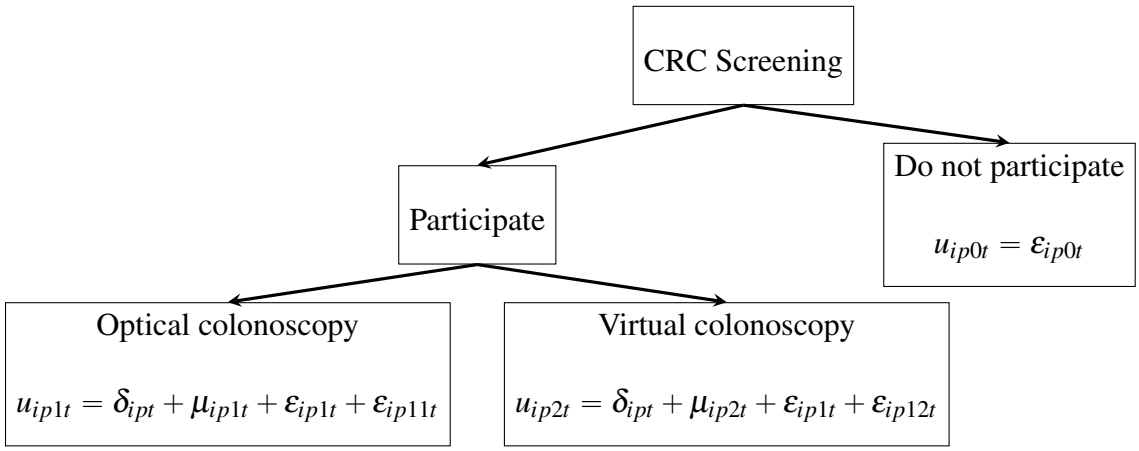
I develop a two-level nested logit model of the screening participation and technology decisions. Patient i has a visit with her primary care doctor p in period t. During this visit, the patient and her PCP jointly make the extensive margin decision as to whether to schedule colorectal cancer screening; i.e. they choose between $g = 0, 1$. Conditional on choosing to participate in screening, they make the intensive margin decision between two screening technologies: OC, indexed by $j = 1$, and VC, indexed by $j = 2$.

The nesting tree and the indirect utilities associated with each terminal decision are depicted in figure 1.3. The utility associated with each of the alternatives within the treatment nest contains the term δ_{ipt}, the empirical specification for which is

$$\delta_{ipt} = \alpha^{EX} + \mathbf{x}'_{it}\beta^{EX} + \tau^{EX}\mathbb{1}Post_t + \zeta_p^{EX}.$$

Here, $\mathbf{x}_{it}$ is a vector of time-varying patient characteristics including age group, number of comorbidities, insurance type, sex, race, marital status, ACG score, an indicator for whether the patient has ever been screened before, an indicator for being newly eligible for screening, characteristics of the patient's zip code, and the RUCA code of the patient's census tract. $\mathbb{1}Post_t$ is a binary indicator for whether the EHR alert had been implemented. ζ_p^{EX} are PCP fixed effects.

Figure 1.3: Nesting tree for model of screening choice



μ_{ipjt} is the portion of the indirect utility that is technology-specific. The empirical specification of μ_{ipjt} is

$$\mu_{ipjt} = \begin{cases} 0 & \text{if } j = 1 \\ \alpha^{IN} + \mathbf{z}_{it}'\beta^{IN} + \tau^{IN}\mathbb{1}Post_t + \zeta_p^{IN} & \text{if } j = 2 \end{cases}$$

$\mathbf{z}_{it}$ is a vector of time-varying demographic variables and measures of the patient's

comorbidity burden, and ζ_p^{IN} are PCP fixed effects.

Turning now to the random components of the utilities, I assume that the timing of the realization of these components is as follows: patient i observes ε_{ipgt}, makes her screening participation decision, observes ε_{ipgjt}, and then makes her modality choice decision. I assume that all random error terms ε_{ipgt} and ε_{ipgjt} are independently and identically type 1 extreme value distributed, the former with scale parameter μ^g and the latter with scale parameter μ^j.[2] I assume that each patient receives a new ε draw at each visit. Let the parameter $\sigma \in [0,1]$ be the correlation between ε_{ip11t} and ε_{ip12t}, which is determined by the ratio of the scale parameters μ^g and μ^j.

Conditional on choosing to participate in screening, patients decide which of the available alternatives to be screened by. The bottom-level choice probability of choosing to be screened by technology j conditional on $g = 1$ is

$$s_{ipjt|g=1} = \frac{\exp\{\mu_{ipjt}/(1-\sigma)\}}{1+\exp\{\mu_{ipjt}/(1-\sigma)\}}. \tag{3}$$

Let I_{ipt} be a measure of the expected aggregate utility or "inclusive value" of all the alternatives in the screening nest:

[2]The likelihood function of this model is equivalent to one in which all random elements of utility are observed simultaneously, but in which ε_{ip1t} follows a Cardell distribution. The choice probabilities and parameters estimated under a model without sequential choice are therefore the same as in the model presented here.

$$
\begin{aligned}
I_{ipt} &= \mathbb{E}[\max\left(u_{i11t}, u_{i12t}\right)] \\
&= \log\left[1 + \exp\{\mu_{ip2t}/(1-\sigma)\}\right].
\end{aligned}
$$

The top-level choice probability for choosing to participate in screening is

$$
\begin{aligned}
s_{ip1t} &= Pr\left[\mathbb{E}\left[\max\left(u_{i11t}, u_{i12t}\right)\right] \geqslant \varepsilon_{i0t}\right] \\
&= \frac{\exp\{\delta_{ipt} + (1-\sigma)I_{ipt}\}}{1 + \exp\{\delta_{ipt} + (1-\sigma)I_{ipt}\}}
\end{aligned}
$$

while that for choosing not to participate in screening is

$$
s_{ip0t} = \frac{1}{1 + \exp\{\delta_{ipt} + (1-\sigma)I_{ipt}\}}.
$$

I assume that the number of alternatives in the screening nest remains constant over the sample period. Note that the degree to which the utility associated with an individual's best screening alternative affects the screening participation decision depends on σ. In the case where $\sigma = 1$, the modality decision is deterministic and all observationally identical visits that result in screening also result in screening by the technology with the highest systematic utility. In this case, there is no

utility premium from the screening technology choice. As σ approaches zero, the expected value of the maximum of the two random error terms within the screening participation nest increases, and so does the magnitude of the inclusive value's effect on the extensive margin decision. These substitution patterns imply that the effect of changes in the screening alternatives on the screening participation decision is decreasing in σ; i.e.

$$\left| \frac{\partial s_{ip0t}}{\partial I_{ipt}} \right| = \frac{(1-\sigma)}{(1+\exp(\delta_{ipjt}+(1-\sigma)I_{ipt}))^2}$$

is decreasing in σ.

1.6 Identification & estimation

It is useful to consider what variation in the data allows us to identify τ^{EX}, τ^{IN}, and σ, the parameters that quantify the alert's effect on the extensive and intensive margins as well as the mechanisms by which the extensive margin effect is realized. τ^{EX} and τ^{IN} are both coefficients on $\mathbb{1}Post$, the binary indicator for the EHR alert being in place. τ^{EX} parameterizes $\mathbb{1}Post$ in the empirical specification for the portion of the systematic utility associated with participating in screening, δ_{ipt}, whereas τ^{IN} parameterizes $\mathbb{1}Post$ on the technology-specific portion of the systematic utility, μ_{ipjt}. σ is identified off of the correlation between the expected aggregate utility of

participating in screening, I_{ipt}, and the likelihood of participating in screening. The greater this correlation, the smaller the implied value of σ. This is consistent with the intuition laid out in subsection 1.5 about how σ determines the extent to which the modality choice decision impacts the participation decision.

τ^{IN} is identified off of variation in each technology's share of screenings across the pre- and post-alerts periods. The greater the substitution between the technologies across periods, the greater the magnitude of τ^{IN}. I_{ipt} is an increasing function of τ^{IN}. If τ^{IN} is nonzero, then there will be variation in I_{ipt} across the pre- and post-alerts periods conditional on all other covariates. τ^{EX} is identified off of whatever variation in screening rates across periods remains after controlling for the inclusive value given the estimated value of σ.

There are 110 parameters to estimate in the bottom level, and 117 to estimate in the top level. Because the number of parameters to estimate is large, I estimate the model by sequential maximum likelihood with some loss of efficiency relative to estimating the full model in a single step. I bootstrap the standard errors of the top level using 1,000 bootstrap replications. I do not restrict $\sigma \in [0, 1]$, though an estimated σ outside of this range is inconsistent with random utility theory.

1.7 Results

The results of estimating the nested logit model presented in subsection 1.5 are presented in table 1.5. The point estimate of σ is close to zero, which implies that there is little correlation in the structural error terms on OC and VC. Stated differently, this means that each modality does not compete any more with the other than it does with the outside option. This estimate of σ also implies a large marginal effect of the inclusive value on the extensive margin.

Table 1.5: Estimation results for nested logit model

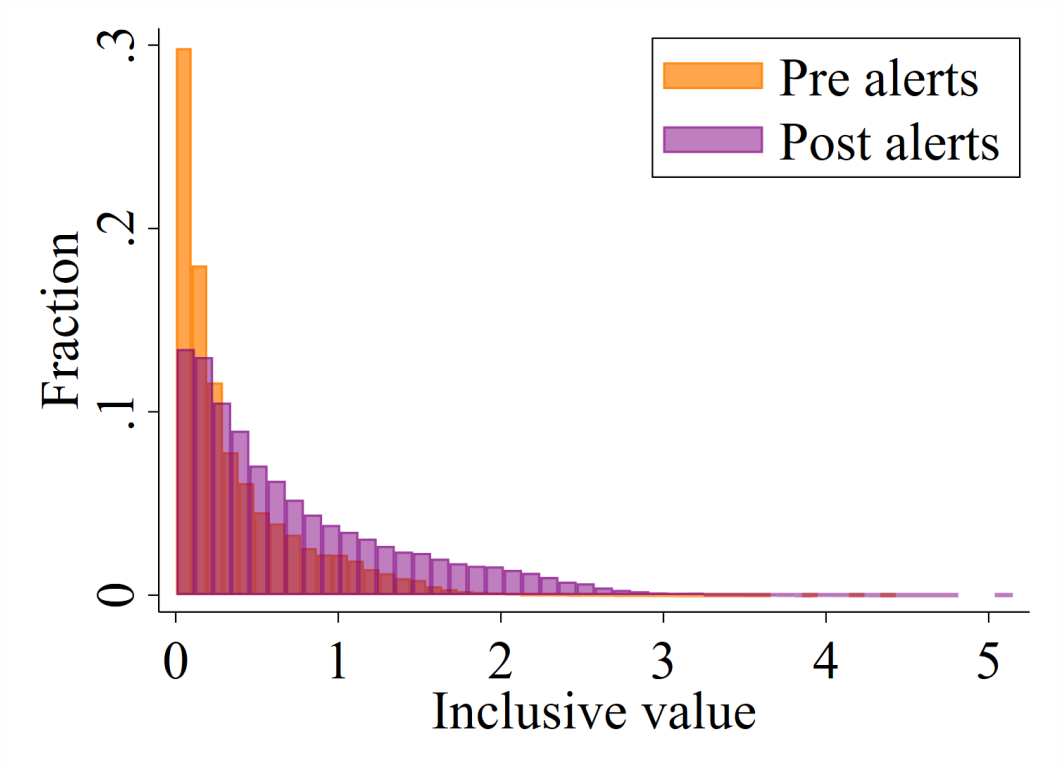
	(1) Any screen			(2) Virtual colonoscopy		
	Coef.	S.E.	M.E. $\times 100$	Coef.	S.E.	M.E. $\times 100$
Inclusive value, $(1-\sigma)$	0.94	(0.21)	1.32			
Post EHR alerts, τ^{EX}/τ^{IN}	0.10	(0.14)	0.15	1.16	(0.32)	15.20
Patient characteristics, β^{EX}/β^{IN}						
Prior screening	2.76	(0.18)	3.91	1.40	(0.24)	18.33
Age [55, 60]	-0.99	(0.12)	-1.89	0.66	(0.20)	9.01
Age [60, 65]	-1.03	(0.14)	-1.94	0.52	(0.22)	6.89
Age [65, 70]	-1.39	(0.18)	-2.30	0.21	(0.39)	2.64
Age [70, 75]	-2.56	(0.27)	-2.93	0.71	(0.57)	9.63
One comorbidity	-0.54	(0.12)	-0.69	0.32	(0.22)	4.38
Two comorbidities	-0.59	(0.21)	-0.74	-0.61	(0.49)	-7.20
Three or more comorbidities	-1.66	(0.59)	-1.42	1.73	(0.90)	26.42
Non-white	-0.20	(0.12)	-0.27	-0.32	(0.36)	-3.98
Married	0.06	(0.08)	0.08	0.17	(0.18)	2.19
Private insurance	0.14	(0.16)	0.18	0.72	(0.40)	8.63
Female	-2.38	(0.16)	-4.80	1.37	(0.22)	19.97
ACG score	-0.53	(0.13)	-0.75	-0.01	(0.23)	-0.18
Zip code fraction $\geqslant$ H.S.	1.58	(1.17)	2.23	1.09	(2.56)	14.26
Zip code fraction in poverty	0.04	(0.41)	0.05	1.44	(0.86)	18.87
Newly eligible	1.28	(0.10)	1.81			
RUCA code 2	-0.01	(0.09)	-0.02			
RUCA code 3	0.05	(0.27)	0.07			
RUCA code 4	0.13	(0.20)	0.19			
Observations		93,602			1,448	
Pseudo R^2		0.24			0.30	

Notes: Coefficients, standard errors, and marginal effects multiplied by 100 are presented in the first, second, and third columns of each model. Screening participation model (1) estimated on full sample. Technology choice model (2) estimated on the sample of visits that result in screening. Omitted categories are age [50, 55], zero comorbidities, and RUCA code 1. Year, month, and provider fixed effects are included but not shown. Standard errors in parentheses.

The effect of the EHR alert on the intensive margin is captured by the estimate

of τ^{IN} in column 2. The implementation of the alert increases VC's share of screenings by 15%. This large intensive margin effect generates a sizable increase in the inclusive value of the modality decision. I illustrate this increase in figure 1.4. I see that there is meaningful within-period variation in the inclusive value, as well as a rightward shift of the distribution following the implementation of the alert that increases the within-period mean by 0.4. Given that my estimate of the marginal effect of the inclusive value on the extensive margin is 1.32, the increase in the inclusive value generated by the alerts implies an increase in screening rates of 0.5 p.p., which is consistent with my results from subsection 1.4. These results imply that the entirety of the increase in screening rates generated by the alert was realized through the alert's validation of VC, and that the alert's function as a nudge had little to no impact on screening rates. These results do not take into account potentially important heterogeneity across patients with respect to both the degree of substitution toward VC and changes in screening rates. Meaningful heterogeneity in either of may change our conclusions regarding the source of increased participation, as I will show in subsection 1.8.

Figure 1.4: Within period distribution in inclusive value



Turning to the estimates for the patient characteristics, I see that patients who participated in screening prior to the beginning of the sample period are much more likely to participate during the sample period and to be screened by VC. The likelihood of screening is decreasing monotonically with age and with the patient's number of comorbidities. There is no clear relationship between age or comorbidity burden and modality choice. There is no evidence of a racial bias in screening rates. Females are less likely to participate in screening, though they are more likely to be screened by VC conditional on participation. Consistent with the discussion

in subsection 1.4, newly eligible patients are more likely to participate in screening.

1.8 Heterogeneous effects

There are good reasons to expect meaningful heterogeneity in the effectiveness of the alert across patients with different baseline screening propensities. The more chronic conditions a patient has, the more concerns they likely have to discuss with the physician during the primary care appointment. There is a sizable literature documenting the crowding out generated by these competing demands and the behavior of PCPs under time constraints (Tai-Seale et al., 2007; Tai-Seale and McGuire, 2012; Yarnall et al., 2003). In keeping with this literature, my results in table 1.5 show that patients with more comorbidities are less likely to participate in screening. These patients have greater potential to benefit (in terms of increased CRC screening rates) from the alert's function as a nudge and its prioritization of CRC screening among their set of competing demands. Patients with lower baseline screening propensities also have greater potential to benefit from the alert's validation of VC, as they likely have lower systematic utility for OC. These patients would see the largest increase in the inclusive value of screening as a result of VC's validation.

I examine how the effect of the alert varies across patients according to their baseline screening propensities by modifying my nested logit model to allow the

effect of the alerts on the extensive and intensive margins to vary across this dimension. I generate baseline screening propensities by estimating the nested logit model outlined in subsection 1.5 on only the pre-alert data, omitting from my estimation τ^{EX} and τ^{IN}. The estimation results for this model are presented in table 1.A2. The patients least likely to be screened in the baseline period are older, non-white, and have multiple comorbidites. I generate predicted baseline probabilities of participating in screening using these estimates and group observations based on which quartile of the predicted baseline screening propensity they fall into. The distribution of baseline screening propensities is illustrated in figure 1.A2. I then estimate the nested logit model again on the full sample, interacting the post-alert indicator with a separate indicator for each of the baseline propensity quartiles. In particular, I modify the systematic utilities from subsection 1.5 to take the form

$$\delta_{ipt} = \alpha^{EX} + \mathbf{x}_{it}'\beta^{EX} + \tau_{\theta_Q}^{EX}\mathbb{1}Post_t + \zeta_p^{EX}$$

and

$$\mu_{ipjt} = \begin{cases} 0 & \text{if } j = 1 \\ \alpha^{IN} + \mathbf{z}_{it}'\beta^{IN} + \tau_{\theta_Q}^{IN}\mathbb{1}Post_t + \zeta_p^{IN} & \text{if } j = 2 \end{cases}$$

where $Q \in \{1,2,3,4\}$ denotes a quartile of the baseline propensity distribution.

The grouping of patients according to baseline screening propensities is an example of what Abadie et al. (2018) calls endogenous stratification. This type of disaggregation has the potential to bias upwards the treatment effects of low propensity individuals and bias downward the treatment effects of high propensity individuals. I correct for this bias by implementing the leave-one-out estimator proposed in Abadie et al. (2018) when estimating the screening participation model on the pre-alert data.

I present the estimation results for the nested logit model with heterogeneity in table 1.6. The estimate of σ here is significantly higher than that estimated in the results without heterogeneity, implying a meaningful amount of correlation between the random errors of the screening alternatives. I see that the alert's function as a nudge increased screening rates among the bottom three quartiles of the baseline propensity distribution. There is no nudge effect for those most likely to be screened in the baseline period. In column 2 of table 1.6, I see that the alert's validation of VC increased its share of screenings across the baseline propensity distribution. Patients with the lowest likelihood of participating see the greatest increase in VC utilization.

Table 1.6: Estimation results for nested logit model with heterogeneity

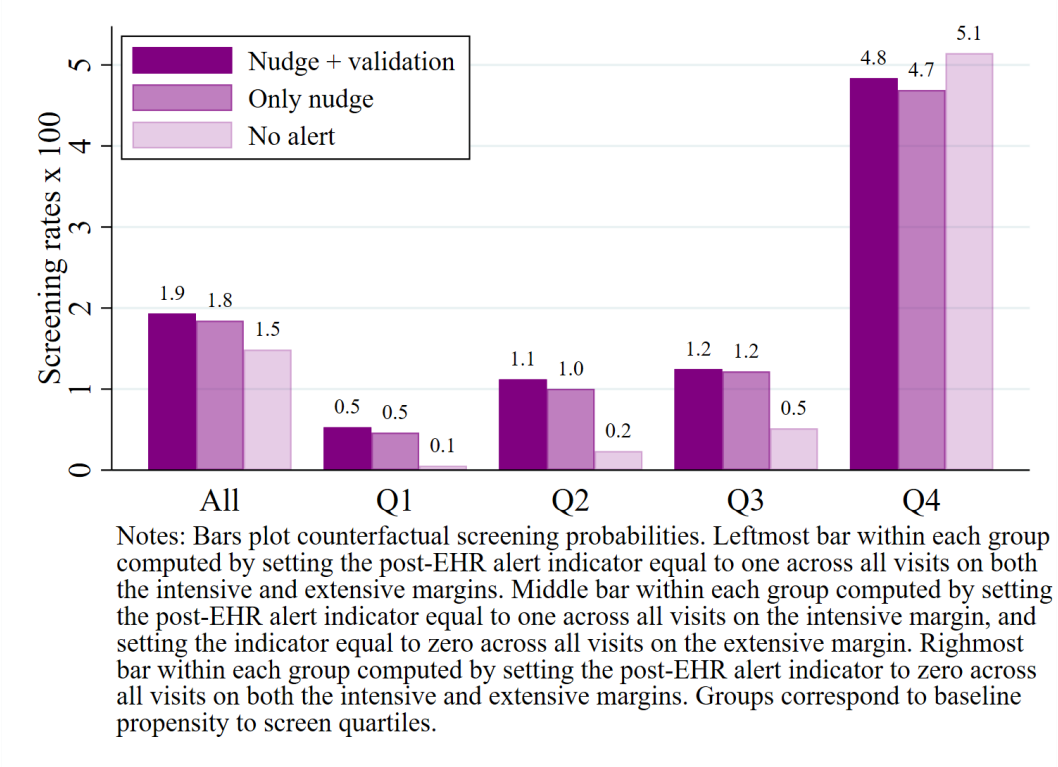
	(1) Any screen			(2) Virtual colonoscopy		
	Coef.	S.E.	M.E. $\times 100$	Coef.	S.E.	M.E. $\times 100$
Inclusive value, $(1-\sigma)$	0.37	(0.17)	0.46			
Post EHR alerts, $\tau^{EX}_{\theta_Q}/\tau^{IN}_{\theta_Q}$						
1st quartile	2.16	(0.26)	0.39	2.10	(0.58)	34.62
2nd quartile	1.46	(0.22)	0.71	1.72	(0.46)	27.13
3rd quartile	0.87	(0.17)	0.69	0.67	(0.41)	10.59
4th quartile	-0.11	(0.14)	-0.45	1.13	(0.33)	14.47
Observations		93,602			1,448	
Pseudo R^2		0.25			0.31	

Notes: Screening participation model (1) estimated on full sample. Technology choice model (2) estimated on the sample of visits that result in screening. Standard errors for screening participation model (1) bootstrapped. Patient characteristics as well as year, month, and provider fixed effects are included but not shown. Standard errors in parentheses.

Unlike the results without heterogeneity presented in subsection 1.7, the results with heterogeneity presented here indicate that there was potential for both the nudge and the validation effects of the alerts to impact the extensive margin. It is not clear from the estimates in table 1.6 alone how much each of these effects contributes to the increase in screening rates. I can quantify each effect's relative contribution to changes on the extensive margin by sequentially shutting down each effect and comparing the change in predicted screening rates at each step. In particular, I first compute screening rates setting the post-EHR alert indicator equal to one on both the intensive and extensive margins across all visits. Under this

regime, both functions of the alert are active. I then turn off the validation effect of the nudge by setting the post-EHR alert indicator equal to zero on the intensive margin; i.e. I set τ^{IN} equal to zero. I compute the inclusive value and screening rates under this counterfactual regime where the alert functioned only as a nudge. Lastly, I set the post-EHR alert indicator equal to zero on both the intensive and extensive margins, effectively shutting down both functions of the alert.

Figure 1.5: Contribution of nudge and signal channels to increase in screening rates



Notes: Bars plot counterfactual screening probabilities. Leftmost bar within each group computed by setting the post-EHR alert indicator equal to one across all visits on both the intensive and extensive margins. Middle bar within each group computed by setting the post-EHR alert indicator equal to one across all visits on the intensive margin, and setting the indicator equal to zero across all visits on the extensive margin. Righmost bar within each group computed by setting the post-EHR alert indicator to zero across all visits on both the intensive and extensive margins. Groups correspond to baseline propensity to screen quartiles.

The results of this exercise are illustrated in figure 1.5. Across all patients, shutting down the validation effect is associated with a 0.09 p.p. (5%) decline in

screening rates relative to the scenario in which both the nudge and validation effects are active. Shutting down the nudge effect results in an additional decrease of 0.36 p.p., an 18% fall relative to the scenario in which both effects are active. These results imply that the nudge effect is the source of the majority of the increase in screening rates across patients, accounting for 80% of the 0.45 p.p. increase that the alert generated. This pattern holds for the bottom three quartiles of the baseline propensity distribution, for which the nudge effect accounts for between 86% and 96% of the total increase in screening rates generated by the alert. There is no statistically significant nudge effect for the fourth quartile, and the increase in the average of the inclusive value for members of the fourth quartile is not large enough to generate a meaningful extensive margin effect. The alert's only impact on the fourth quartile is substitution of VC for OC. Note that this result is consistent with our event study results from subsection 1.4, which showed the screening rate returning to its original level by the end of the sample period but persistence in increased utilization of VC.

I reject the model without heterogeneity, the results for which were presented in subsection 1.7, in favor of the model with heterogeneity presented here. The χ^2 of a likelihood ratio test for the two models is 173, and the associated p-value is < 0.001. In both models and across all groups, I find there is substantial substitution away from OC toward VC, which is manifested in economically significant

estimates of τ^{IN}. The increase in the overall utility of the screening choice set is tied to the degree of substitution from OC to VC through τ^{IN}. In the model without heterogeneity, the average increase in screening rates is pulled down by the fourth quartile of the baseline propensity distribution. This limits the amount of variation in screening rates across periods that remains after controlling for the increase in the inclusive value. The specification with heterogeneity allows the size of the nudge effect to capture the remaining variation within quartiles, which is nonzero for the bottom three.

1.9 Conclusion

In this paper, I look at the impact of an EHR alert for colorectal cancer screening. The alert I study is novel in the sense that it not only reminded physicians to screen their patients but also provided the first official validation of a new screening modality, virtual colonoscopy. This paper has three main findings. The first is that the alert accelerated the adoption of VC, more than doubling its utilization. This result suggests that whether a technology has been officially sanctioned is a salient product characteristic for decision makers in healthcare settings and that such decision makers are responsive to new information about a technology's validation status.

Second, I find that the EHR alert reduced disparities in CRC screening, yielding

the greatest benefit to older, non-white patients with multiple comorbidities who were the least likely to be screened in the baseline period. The alert's effect as a nudge is greater among these patients, which is consistent with their having more competing demands to address during primary care visits and more to gain from the prioritization of screening. The validation effect is also greater among these patients, which is consistent with their having a lower systematic utility for OC. These findings highlights the importance of designing interventions that address the multiple reasons why patients do not participate in preventive care and the potential of such interventions to reduce health inequities.

Finally, I find that the alert's function as a nudge did more to increase screening rates than its validation of VC. After accounting for heterogeneous effects, I find that substitution between VC and OC is significantly greater than substitution between either modality and the outside option, implying that marginal improvements in screening technologies do not have a great impact on the extensive margin. The importance of the nudge effect is consistent with many explanations for why patients do not participate in screening, of which competing demands and a lack of awareness of preventive care guidelines are just two. If competing demands is the most important barrier to screening participation, then this raises the question of what other types of care are crowded out when the EHR alert prioritizes screening, as well as doubts as to whether this intervention was welfare enhancing. If a

lack of awareness of preventive care guidelines is the major barrier, then this calls into question why the alert was not more effective at increasing screening rates and highlights the gap between guideline awareness and adherence recently explored in (Abaluck et al., 2020). Overall, these findings suggest that a widely available health information technology can be used to increase participation in preventive care among the millions of adults who may be missing the chance to find colon cancer early when treatment works best.

1.10 Appendix

Table 1.A1: Estimation results for probit model with sample selection

	(1) Any Screen		(2) Virtual colonoscopy	
Prior screening	1.70	(0.08)	1.14	(0.26)
Age [55, 60]	-0.32	(0.03)	0.20	(0.15)
Age [60, 65]	-0.37	(0.03)	0.14	(0.17)
Age [65, 70]	-0.54	(0.05)	-0.07	(0.25)
Age [70, 75]	-0.94	(0.07)	0.09	(0.38)
Private insurance	0.08	(0.04)	0.43	(0.21)
Female	-0.81	(0.03)	0.53	(0.22)
Comorbidities: 1	-0.19	(0.03)	0.16	(0.12)
Comorbidities: 2	-0.29	(0.07)	-0.38	(0.28)
Comorbidities: 3	-0.39	(0.13)	0.82	(0.57)
Non-white	-0.11	(0.05)	-0.16	(0.20)
Married	0.03	(0.02)	0.11	(0.09)
ACG score	-0.22	(0.04)	-0.06	(0.12)
Zip code fraction $\geqslant$ H.S.	1.00	(0.51)	0.98	(1.35)
Zip code fraction in poverty	0.20	(0.15)	0.87	(0.44)
Newly eligible	0.59	(0.03)		
RUCA Code: 2	-0.01	(0.03)		
RUCA Code: 3	0.05	(0.12)		
RUCA Code: 4	0.07	(0.07)		
Constant	-2.83	(0.51)	-3.76	(1.40)
ρ	0.31	(0.23)		
Observations	93,602			

Notes: Omitted categories are age [50, 55], zero comorbidities, and RUCA code 1. Quarters-since-EHR alert implementation and provider fixed effects are included but not shown. Standard errors in parentheses.

Table 1.A2: Estimation results for nested logit model using pre-alert data

	(1) Any screen			(2) Virtual colonoscopy		
	Coef.	S.E.	M.E. $\times 100$	Coef.	S.E.	M.E. $\times 100$
Inclusive value, $(1-\sigma)$	0.66	(1.82)	0.87			
Patient characteristics, β^{EX}/β^{IN}						
Prior screening	2.99	(0.50)	3.95	1.15	(0.31)	13.62
Age [55, 60]	-1.03	(0.12)	-1.99			
Age [60, 65]	-1.00	(0.12)	-1.96			
Age [65, 70]	-2.22	(0.29)	-2.90			
Age [70, 75]	-3.62	(0.45)	-3.23			
Age $\geqslant 65$				0.68	(0.71)	9.98
One comorbidity	-0.32	(0.13)	-0.41	-0.15	(0.33)	-1.75
Two comorbidities	-0.83	(0.30)	-0.87			
Three or more comorbidities	-2.03	(0.75)	-1.43			
Two or more comorbidities				0.38	(0.60)	5.07
Non-white	-0.41	(0.20)	-0.47	0.20	(0.48)	2.57
Married	0.28	(0.13)	0.35	0.33	(0.30)	3.63
Private insurance	0.14	(0.30)	0.17	0.98	(0.78)	8.69
Female	-2.84	(0.13)	-4.51	0.21	(0.33)	2.64
ACG score	-0.60	(0.14)	-0.79	0.16	(0.34)	1.88
Zip code fraction $\geqslant$ H.S.	-0.04	(1.64)	-0.06	-1.31	(3.22)	-15.56
Zip code fraction in poverty	0.54	(0.62)	0.72	1.22	(1.15)	14.46
Newly eligible	1.43	(0.11)	1.89			
RUCA code 2	0.08	(0.12)	0.10			
RUCA code 3	0.30	(0.37)	0.44			
RUCA code 4	-0.40	(0.28)	-0.45			
PCP characteristics, β^{EX}/β^{IN}						
Female				-0.03	(0.27)	-0.32
Family Medicine specialty				0.10	(0.23)	1.15
Observations		49,175			736	
Pseudo R^2		0.32			0.04	

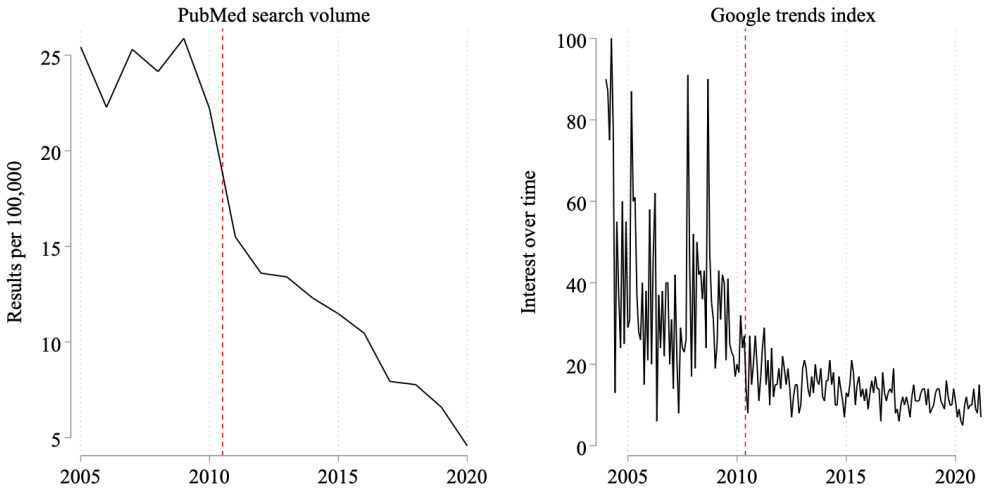
Notes: Coefficients, standard errors, and marginal effects multiplied by 100 are presented in the first, second, and third column of model. Screening participation model (1) estimated on sample of visits that took place prior to implementation of alert. Technology choice model (2) estimated on the sample of visits that result in screening and took place prior to alert. Omitted categories are age [50, 65], zero comorbidities, and RUCA code 1. Year and month fixed effects are included but not shown. Provider fixed effects included in model (1). Standard errors in parentheses.

Table 1.A3: Estimation results for nested logit model with heterogeneity cont.

	(1) Any screen			(2) Virtual colonoscopy		
Prior screening	3.31	(0.20)	4.13	1.50	(0.25)	19.48
Age [55, 60]	-1.04	(0.10)	-1.79	0.59	(0.20)	8.02
Age [60, 65]	-1.11	(0.12)	-1.86	0.43	(0.23)	5.73
Age [65, 70]	-1.68	(0.15)	-2.32	0.00	(0.41)	0.06
Age [70, 75]	-3.21	(0.24)	-2.85	0.21	(0.61)	2.72
Comorbidities: 1	-0.55	(0.10)	-0.62	0.30	(0.21)	4.12
Comorbidities: 2	-0.84	(0.19)	-0.85	-0.78	(0.50)	-8.90
Comorbidities: 3	-1.75	(0.42)	-1.30	1.45	(0.96)	21.79
Non-white	-0.31	(0.12)	-0.35	-0.40	(0.37)	-4.95
Married	0.12	(0.07)	0.15	0.20	(0.18)	2.61
Private	0.29	(0.14)	0.33	0.83	(0.42)	9.72
Female	-2.66	(0.15)	-4.39	1.29	(0.25)	18.49
ACG score	-0.65	(0.11)	-0.81	-0.08	(0.23)	-1.06
Zip code fraction >= H.S.	1.65	(1.18)	2.07	1.05	(2.56)	13.58
Zip code fraction in poverty	0.39	(0.35)	0.50	1.55	(0.87)	20.04
Newly eligible	1.39	(0.10)	1.73			
RUCA Code: 2	0.01	(0.08)	0.00			
RUCA Code: 3	0.08	(0.27)	0.10			
RUCA Code: 4	0.05	(0.20)	0.06			
Observations	93,602			1,448		
Pseudo R^2	0.25			0.31		

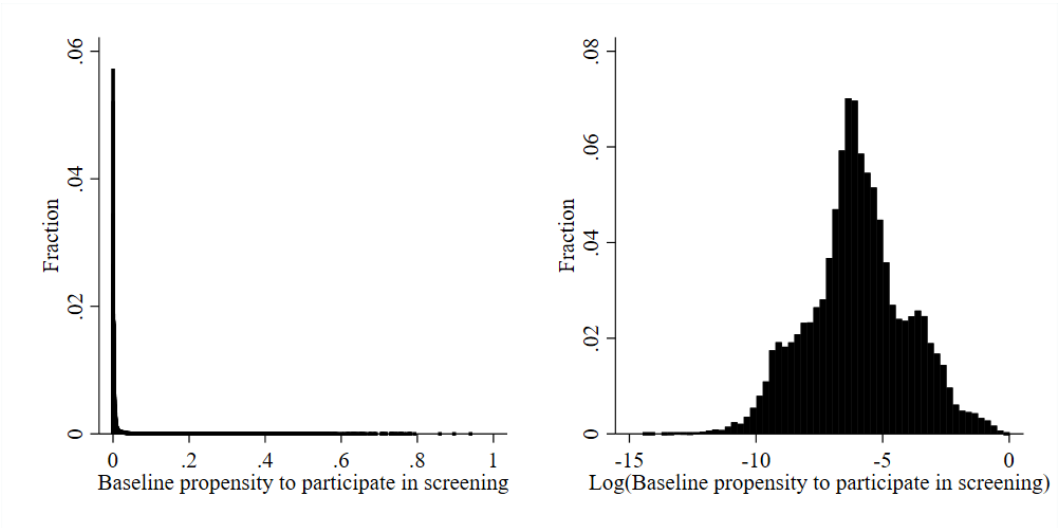
Notes: Screening participation model (1) estimated on full sample. Technology choice model (2) estimated on the sample of visits that result in screening. Year, month, and provider fixed effects are included but not shown. Standard errors in parentheses.

Figure 1.A1: General and academic interest in "virtual colonoscopy"



Note: Vertical line at date of EHR alert implementation. Lefthand panel made with PubMed by Year at http://esperr.github.io/pubmed-by-year

Figure 1.A2: Distribution of predicted baseline propensities to participate in CRC screening



2 The Impact of a National Formulary Expansion on Diabetics

2.1 Introduction

Prescription drug formularies are an element of health insurance plan design that determine coverage and coinsurance rates for medications. Formularies are an important mechanism for healthcare cost containment that can increase the bargaining power of insurance companies in their negotiations with pharmaceutical manufacturers over the price of prescription drugs. The complexity of insurance plan design in settings such as the U.S. healthcare market and insurers' discretion over the elements of that design make it difficult to assess how changes in prescription drug coverage impact consumers. Evaluating changes in formulary design when such changes are endogenous choices of the insurer usually requires a structural approach, as well as information on all other plan characteristics. While expanded coverage might increase prescription drug costs, spillovers from drug to non-drug spending raise the possibility that adding drugs to a formulary might decrease healthcare costs overall. For instance, Tamblyn et al. (2001) show that the rate of adverse health outcomes and emergency room visits increase among poor and elderly individuals following an increase in cost sharing for essential prescription drugs. Understanding how formulary design impacts healthcare costs and utiliza-

tion is of great concern to countries like Canada, Mexico, Japan, Colombia, and the U.S. where prescription drug spending comprises more than 10% of total healthcare costs (OECD, 2020). Also of keen interest is how insurers' ability to respond to exogenous changes in formulary design impacts enrollees.

In this paper, we examine the impact of an exogenous expansion of Colombia's national prescription drug formulary at the end of 2011. The national formulary in Colombia is designed by the government and determines the set of covered medications and their level of cost sharing for all insurers. The formulary is part of a wider benefits package that covers inpatient and outpatient care and which private health insurers are obliged to offer to all of their enrollees. An estimated 96.6% of Colombians are covered by the nation's universal healthcare system (OECD, 2015). The government regulates premiums and cost sharing, but service and drug prices are determined through bilateral bargaining between insurers and pharmaceutical companies. The government also regulates drug prices by setting price ceilings according to the degree of competition in each drug class. Before 2012, the formulary covered 673 medications. At the end of 2011, it was expanded to include 736 medications as part of a broader healthcare reform which unified the income-based insurance plans. We describe the Colombian healthcare system and this reform in more detail in section 2.2.

Regulation of cost sharing and coverage schedules in Colombia alleviates the

endogeneity of formulary design and allows us to isolate the impact of the formulary's expansion. We are able to separate the effect of the formulary expansion from the effect of plans' unification by focusing on a particular type of drug and its users, namely, insulin and diabetics. The formulary expansion affected all enrollees to the healthcare system, but focusing on expanded coverage of insulin - which is taken exclusively by diabetics - allows us to build a control group from the non-diabetic population. Since there is no generic insulin,[1] our focus also allows us to examine the impact of an increase in branded drug competition on healthcare utilization and costs. This is in contrast to the bulk of literature examining drug pricing and entry, which has focused on the effects of generic drug entry (Tenn and Wendling, 2014; Regan, 2008; Reiffen and Ward, 2005; Scott Morton, 2005). Diabetes is an increasingly prevalent chronic condition in many countries, especially the United States, where over 1 in 10 individuals had diabetes in 2018 and the total direct estimated costs of diagnosed diabetes increased from $188 billion in 2012 to $237 billion in 2017 (CDC, 2020). Research on how formulary design affects the costs associated with diabetes is therefore of particular interest to policymakers.

We use a difference-in-differences approach to identify the causal effect of the expanded coverage of insulin. The treatment group is comprised of individuals with

[1] Because insulins are biologically based rather than molecularly based, they are too complicated to replicate exactly. Insulins that are relatively close substitutes for one another are called biosimilars.

type I diabetes and the control group is constructed by exactly matching diabetics to non-diabetics based on demographics, comorbidity profiles, and insurance carrier. Using granular claims data, we are able to analyze the impact of expanded insulin coverage on the utilization and cost of several types of healthcare. We find that annual consumption of insulin increased by 2 claims (28%) among type I diabetics as a result of expanded insulin coverage. Because we do not observe out-of-pocket insulin purchases, we cannot quantify how much of the observed increase in insulin consumption is attributable to moral hazard, and how much is the mechanical result of out-of-pocket payments becoming covered claims. Substitution toward newly covered, relatively more expensive insulins increases insulin costs by over half a million Colombian pesos or 17% of baseline total healthcare costs for type I diabetics. At the same time, outpatient care and non-insulin prescription drug utilization as well as hospitalization rates for type I diabetics all decline as a result of expanded coverage of insulin. We look for evidence of two mechanisms which might generate this increase in insulin and decline in non-drug spending: spillovers and rationing of care. The spillovers hypothesis posits that increased drug spending can result in lower non-drug costs as patients are matched to their optimal prescription and their health status is improved. Empirical evidence on spillovers include Lavetti and Simon (2018) who find that Medicare Advantage plans capitalize on spillovers once open enrollment has closed by lowering coinsurance rates for drugs that are

associated with lower non-drug spending. Tamblyn et al. (2001) show that reductions in the utilization of essential drugs result in increased rates of adverse health events in elderly individuals and welfare recipients. In our setting, increased access to insulin, an essential drug for type I diabetics, has the potential to decrease the use of non-drug treatment of adverse health events among this population.

The rationing of care hypothesis posits that in the short run, insurers respond to increases in prescription coverage and costs by limiting the amount of discretionary care that newly more costly patients receive. In the long run, this rationing can result in lower enrollment from these more costly individuals, acting as a mechanism for selection.[2] The main contribution of this paper is providing evidence consistent with insurers engaging in rationing of care in response to changing selection incentives in a setting where they have no control over premiums or cost sharing. Risk selection through premium setting is well studied (Akerlof, 1970; Handel et al., 2015; Hackmann et al., 2015). There is also an impressive literature examining how insurers alter plan design in response to regulatory changes and risk selection incentives in government coordinated health insurance markets. Andersen (2017) finds that in the United States, insurers respond to drug coverage requirements under the Affordable Care Act (ACA) by placing both marginal

[2]If rationing care is a well-known phenomenon, we might worry about rationing generating disenrollment in the short run as well. The best way to account for that correlation would be to focus on the subsample of patients that are continuously enrolled. Unfortunately we have no data on enrollment spell lengths for our analysis period.

and inframarginal drugs on higher formulary tiers, or subjecting them to utilization management. Geruso et al. (2019) also find that insurers in the ACA Exchanges increase cost sharing and utilize non-price barriers, such as prior authorization for drugs, for patients who are predictably unprofitable conditional on risk adjustment. Whether insurers in tightly regulated health insurance markets like Colombia's are able to respond to changes in risk selection incentives is still an outstanding empirical question. We show that in a setting where plan design is heavily regulated and risk adjustment is coarse, selection is still possible through rationing of care. We devise tests to show that spillovers from drug to non-drug spending are unlikely to be the source of cost savings, and provide evidence that these cost reductions stem from the targeted rationing of discretionary outpatient healthcare.

Our paper is also related to the literature studying the effects of prescription drug coverage and insulin consumption on diabetics' healthcare utilization and costs. In recent work, Américo and Rocha (2020) evaluate the impact of a policy implemented in Brazil that made subsidized pharmaceutical drugs available at retail pharmacies, focusing on the spillover effects of the policy. They find that the increase in cost sharing for and availability of prescription drugs reduced the hospitalization rate of diabetics by 3.6 percent. This paper complements Américo and Rocha (2020) by examining how a different element of formulary design – the extensive margin decision of whether a drug is covered at all – impacts healthcare utilization

by type I diabetics. In testing for rationing of care as a response to changing selection incentives, this paper is also related to the literature on rationing of healthcare, the vast majority of which is both theoretical and specific to the use of waiting time as a tool for engaging in general rather than targeted rationing (Cullis and Propper, 2000; Fabbri and Monfardini, 2009). We draw on insights from Ellis and McGuire (2007) and Frank et al. (2000) to examine how utilization of discretionary and non-discretionary types of healthcare decline for type I diabetics as a result of expanded insulin coverage. Consistent with the rationing hypothesis, we show that utilization of lab tests, which are a type of diagnostic and preventive care that is predictably used by type I diabetics, declines across the board.

This paper both provides anecdotal evidence of the mechanisms by which rationing is achieved, and quantifies insurers' success at achieving rationing targeted at a perfectly observable, newly unprofitable set of enrollees. Quantifying the degree to which the formulary expansion increased access to essential drugs and reduced access to discretionary care among the population of type I diabetics provides valuable insights into the efficacy of health insurance market regulations and the welfare impacts of expanding prescription drug coverage.

2.2 Background

2.3 Colombia's universal health insurance system & formulary expansion

Colombia's universal health insurance system was established in 1993 with Law 100. All individuals are divided into one of the two regimes - the Contributory Regime (CR) and the Subsidized Regime (SR) - that make up the health insurance system. The CR covers individuals above a monthly income threshold. The 51% of the population eligible to join the CR contribute 12% of their monthly income to the system. The remaining 49% of the population below the income threshold are part of the SR. Before 2012, the national plan offered by the CR covered different services, procedures and medications than that offered by the SR. At the beginning of that year the plans' benefits were equalized. Individuals choose from amongst a set of private insurers with which to enroll and access the national plan. To deliver the services covered under the national plan, insurers contract with healthcare providers to create a network. Out-of-network claims are not covered. The sample of enrollees used in our analysis belong only to the CR.

The cost sharing rules in the CR are determined by the government and indexed to the enrollee's monthly income. For individuals earning less than two times the monthly minimum wage (MMW), the coinsurance rate is 11.5% of the price of the

health claim, the copay is 11.7% of the daily minimum wage, and the maximum out-of-pocket expenditure in a year equals 57.5% of the MMW. For those whose monthly income is between 2 and 5 times the MMW, the coinsurance rate is 17.3% of the health claim price, the copay is 46.1% of the daily minimum wage, and the maximum out-of-pocket expenditure in a year equals 230% of the MMW. Finally, for enrollees whose income exceeds 5 times the MMW, the coinsurance rate equals 23%, the copay is 121.5% of the daily minimum wage, and the maximum out-of-pocket expenditure in a year is 460% of the MMW. There are no deductibles in the Colombian system, so copays and coinsurance rates always apply. These cost sharing percentages have remained fixed since the establishment of Colombia's healthcare system and their absolute levels only vary with changes in the minimum wage. Individuals are required to report their income monthly so that their plan can apply the appropriate cost sharing rules to them.

Insurers are not allowed to charge premiums through the national plan. Instead, they are reimbursed by the government every year with capitated payments that are risk adjusted for age, sex, and location. Transfers for year t are approximately calculated as the present value of the average healthcare cost of a given risk pool using the data of all claims reimbursed by insurers during year $t-2$. The capitated payments replace premiums, so that other than the monthly contribution to the CR, enrollment in the national insurance plan is free. The strict regulation of cost

sharing and benefits means that insurers compete in terms of quality and provider networks (Giedion and Uribe, 2009). As in the US, insurers and providers bargain freely over the price of health services, devices, and medications. Private insurers are also allowed to offer complementary insurance plans, for which they can determine cost sharing rules and premiums. However, consumers can only access these complementary packages once they are enrolled in the national plan. The data used in this paper are the cross sections of health claims made by all enrollees to the CR through the national insurance plan in 2011 and 2013.

At the end of 2011, the Colombian Ministry of Health implemented a reform that unified the contributory and subsidized systems' insurance plans and expanded the national prescription drug formulary by 63 drugs.[3] Most drug inclusions were for treatment of mental health conditions (all of which were previously uncovered), insulin, antibiotics, and chemotherapy. The wider benefits package was also expanded to cover complex procedures like open breast biopsy, laparoscopy ovary cystectomy, and colored doppler echocardiogram. Although there have been studies that measure the overall impact of the unification of the CR and SR (Riascos and Camelo, 2014), there is less work on the impact of the formulary expansion. The exogeneity of variables that in other countries would be chosen by the insurer allows us to isolate the impact of the fomulary expansion on healthcare costs and

[3]Established in decree 029 of 2011.

utilization. This is relevant not only for Colombia, where the formulary continues to be modified, but also for countries where the scope and role of national health insurance continues to be debated.

While many of the tools that insurers in other settings use to engage in risk selection are not available to insurers in the CR, there is still scope for insurers to target healthier enrollees through "illicit formularies," whereby insurers deny the provision of or payment for certain medical services or medicines. It is perhaps in part because of these illicit formularies that, since the establishment of its universal health system, Colombia has become the most litigious country in Latin America with respect to lawsuits concerning the refusal of treatments, exams, and pharmaceuticals by insurers. In 2013, 115,147 of such lawsuits were filed in Colombia (Lamprea and Garcia, 2016). We provide anecdotal and empirical evidence of the use of illicit formularies in section 2.8.

2.4 Diabetes & insulin

Insulin is a hormone produced by the pancreas that allows glucose from food to enter a person's cells and controls their blood sugar. A person has diabetes if they do not produce enough insulin or their body does not use insulin well. Type I diabetics produce no insulin, and must take insulin every day in order to stay alive. Type II diabetics produce some insulin, but less as the disease progresses. Type II diabetics

can control their blood sugar with diet and exercise or oral medications such as metformin, but others will require insulin (FDA, 2020). In the short run, failure to control blood sugar can result in hypoglycemia. In the long run, uncontrolled blood sugar can result in kidney disease, heart disease, nerve damage, and several other adverse health outcomes (NIH, 2020).

Prior to its expansion in 2012, two types of insulin were covered by the Colombian formulary: regular and NPH.[4] Regular insulin is a type of bolus insulin. Bolus insulins are fast-acting, taking effect and wearing off more quickly. Bolus insulin is usually taken shortly before mealtimes to provide immediate blood sugar control. NPH belongs to the class of basal insulins, which are longer-lasting and provide blood sugar control throughout the day. These two types of insulin can be consumed together to better manage the disease. With the expansion of the formulary in 2012, the number of insulins covered increased from two to seven, providing more options for patients to exploit the complementarities between types of insulin. The five newly added insulins included three additional bolus insulins and two additional basal insulins.[5] The characteristics of these insulin types are summarized in table 2.1. Differentiation in the characteristics of insulin including onset time, peak time, and duration generate the potential for increased insulin coverage to allow diabetics to be better matched to an insulin regimen and generate spillovers. Note

[4] ATC codes A10AB01 and A10AC01, respectively.
[5] ATC codes A10AB05, A10AB06, A10AB04, A10AE05, and A10AE04.

that the average price of newly added insulins is anywhere from 5 to 12 times the price of NPH or regular insulin in 2013. This means that even modest amounts of substitution towards the newly added insulins can generate large increases in costs.

Table 2.1: Types of insulin

Insulin type (brand name)	Onset	Peak	Duration	Avg. price 2011	Avg. price 2013
Bolus (preprandial or mealtime) insulins					
Rapid-acting insulin analogues					
Insulin aspart (NovoRapid®)	9-20 min	1-1.5 hr	3-5 hr	—	79.2
Faster-acting insulin aspart (Fiasp®)	4 min	0.5-1.5 hr	3-5 hr		
Insulin glulisine (Apidra®)	10-15 min	1-1.5 hr	3.5-5 hr	—	86.2
Insulin lispro (Humalog®)	10-15 min	1-2 hr	3-4.75 hr	—	81.6
Short-acting insulins					
Insulin regular (Humulin®-R, Novolin®ge)	30 min	2-3 hr	6.5 hr	33.4	15.6
Insulin regular (Entuzity®(U-500))	15 min	4-8 hr	17-24 hr		
Basal Insulins					
Intermediate-acting					
Insulin NPH (Humulin®-N, Novolin®ge NPH)	1-3 hr	5-8 hr	Up to 18 hr	14.5	15.5
Long-lasting insulin					
Insulin detemir (Lemevir®)	90 min	N/A	16-24 hr	—	146.3
Insulin glargine U-100 (Lantus®)			24 hr	—	182.1
Insulin glargine U-300 (Toujeo®)			$>$ 30 hr		

Notes: Adapted from Canadian Journal of Diabetes, 2018-04-01, Volume 42, Pages S314-S314. Continuously covered insulins are regular and NPH. The rest of insulins in the table were added to the formulary by the end of 2011. Prices are in thousands of Colombian pesos.

2.5 Data

We use two samples of cross sectional health claims data from the CR from 2011 and 2013, which are one year pre- and post-policy respectively. For every enrollee, we observe basic demographic characteristics including sex, age, and municipality of residence. For every claim, we observe date of provision, service provided, service price, contract under which the claim is reimbursed, insurer, provider, and associated ICD-10 diagnosis code. Since patients must remain enrolled with their choice of insurer for at least a year, we do not observe patients switching insurers during either cross-section of our data.

We obtain each enrollee's set of diagnoses by grouping ICD-10 codes according to Alfonso et al. (2013) into the following conditions: genetic anomalies, arthritis, arthrosis, asthma, autoimmune disease, cancer, cardiovascular disease, type I diabetes, long-term pulmonary disease, renal disease, HIV-AIDS, transplant, tuberculosis, and epilepsy. Note that because diagnoses are defined using claims, our definition does not include individuals who have a diagnosis but who do not file any claims associated with its treatment. We believe that there is limited scope for this type of measurement error for the aforementioned diagnoses, which are chronic conditions that require treatment. For other diagnoses, such as type II diabetes, which can be treated with or without the use of prescription drugs, measurement error of this type is likely. Age is categorized into the following groups used by

the government for the risk adjustment formula: 19-44, 45-54, 55-59, 60-64, 65-69, 70-74, 75+. We do not observe each enrollee's monthly income, but rather an aggregate income measure that is collinear with combinations of sex, age group, and municipality, so we cannot separately identify the impact of differences in cost sharing across income groups on healthcare utilization. We collapse the claims level data to the patient-year level and build measures of utilization and cost by summing across each patient's claims within a year.

We define treatment in a year as having been diagnosed with type I diabetes at any moment during that year and being at least 19 years old. We exclude type II diabetics from our analyses, as their decision to use medication to manage their diabetes determines whether we observe a diabetes diagnosis through their claims. In some cases, Type II diabetes can be managed with diet and exercise alone, without the use of any medication (CDC, 2021). The formulary expansion may have impacted this extensive margin decision of whether to use medication both by expanding the insulin choice set and through its impact on the price of insulin. This implies that treatment for type II diabetics is not exogenous conditional on observables, which is a requirement of the differences-in-differences approach we employ. We determine which patients are diabetic using the ICD-10 diagnoses that accompany their claims, so treated units who did not make a health claim are unobserved. We expect the number of unobserved treated individuals to be close to zero since

type I diabetics can be expected to make at least one claim in a year associated with diabetes management.

We use exact matching to create a control group that is similar to the treatment group of type I diabetics in terms of comorbidities (with the exception of diabetes), sex, age, insurer, and municipality. This choice of control group implies that the effect being estimated is that of the element of the formulary expansion that is relevant only to type I diabetics, namely expanded coverage of insulins. Using exact matching to create the comparison group has three advantages. First, the comparison group will resemble the treatment group, and will therefore be expected to respond to shocks in a similar way. This is important, since both the treatment and control groups were subject to universal elements of the healthcare reform in the post-policy period, including the plan unification. Second, by matching treated units in the pre-policy to those in the post-policy period and then matching treated units to controls separately for each year, we achieve common support on the distribution of the covariates across all four cells. This will keep us from making inferences about outcomes for treated individuals we don't observe in the data. Third, common support also allows us to relax the assumption that the effect of the policy is homogeneous across individuals.

Table 2.2: Balance table of type I diabetics and exactly-matched control units

	Control	Treated
Demographics		
Male (%)	43.08	43.18
Age, mean (sd)	62.99 (12.81)	63.02 (12.81)
Diagnoses (%)		
Arthrosis	2.28	2.32
Cardiovascular disease	73.61	73.43
Long term pulmonary disease	4.52	4.66
Renal disease	14.83	15.24
Insurer (%)		
A	1.78	1.79
B	7.75	7.74
C	1.81	1.81
D	4.18	4.18
E	5.66	5.61
F	9.78	9.75
G	1.73	1.74
H	12.60	12.62
I	12.69	12.73
J	7.08	7.05
K	0.04	0.04
L	0.00	0.00
M	0.31	0.31
Type of municipality (%)		
Metropolitan	73.47	73.05
Normal	26.35	26.75
Special	0.19	0.20
N	2,333,213	97,210

Notes: This table shows some summary statistics of treated and control units after $1-to-n$ exact matching on age, sex, comorbidities, type of municipality, and insurer. Summary statistics for control units are weighted by the inverse number of controls matched to each diabetic.

We perform one-to-many matching of diabetics to non-diabetics. Table 2.2 presents summary statistics of the demographic characteristics and diagnoses of the treatment and control groups. Statistics for control units are weighted by the inverse number of controls matched to each diabetic. 43% of our sample is male, and the average individual is 63 years old. The most common comorbidities are cardiovascular disease which is present in 73% of patients and renal disease which is present in 15% of patients. More than 12% of individuals are enrolled to each of insurers H and I. 73% of diabetics live in urban or metropolitan municipalities. The matched sample consists of 97,210 type I diabetics and 2,333,213 exactly matched controls.

Table 2.3: Utilization and cost for type I diabetics and exact match counterparts

	Control		Treated	
	2011	2013	2011	2013
Claims				
All claims	48.5 (46.9)	50.8 (50.1)	100.6 (71.3)	101.3 (75.6)
Outpatient claims	22.2 (27.2)	20.7 (25.9)	39.8 (41.2)	34.2 (37.7)
Inpatient claims	3.0 (11.1)	3.3 (11.4)	7.9 (21.0)	8.0 (20.5)
Prescription claims	23.4 (24.7)	26.8 (29.9)	52.9 (35.7)	59.0 (44.6)
Insulin claims	0.0 (0.4)	0.0 (0.4)	7.3 (4.9)	9.3 (5.9)
Procedure claims	10.8 (14.2)	10.8 (16.7)	20.8 (21.2)	18.1 (21.8)
Imaging claims	1.4 (2.4)	1.5 (2.5)	1.9 (3.2)	1.9 (3.2)
Lab claims	10.0 (16.1)	10.0 (15.6)	23.5 (28.7)	21.7 (26.5)
Office/consultation claims	7.3 (5.8)	7.1 (5.8)	10.6 (7.3)	10.4 (7.4)
Essential drugs claims	3.4 (5.5)	3.4 (6.1)	6.2 (7.2)	6.6 (8.3)
Costs (Million COP)				
All costs	1.80 (6.44)	1.68 (6.12)	3.44 (8.50)	3.84 (8.22)
Outpatient costs	0.92 (3.21)	0.77 (2.49)	1.50 (3.86)	1.38 (3.42)
Inpatient costs	0.67 (4.21)	0.59 (3.79)	1.37 (5.45)	1.26 (5.10)
Prescription costs	0.21 (1.80)	0.31 (2.68)	0.57 (1.96)	1.21 (2.72)
Insulin costs	0.00 (0.01)	0.00 (0.03)	0.14 (0.30)	0.64 (0.99)
Procedure costs	0.57 (2.60)	0.46 (1.97)	0.94 (3.09)	0.79 (2.77)
Imaging costs	0.12 (0.34)	0.12 (0.36)	0.18 (0.43)	0.19 (0.55)
Lab costs	0.12 (0.29)	0.13 (0.31)	0.30 (0.48)	0.30 (0.52)
Office/consultation costs	0.15 (0.23)	0.14 (0.25)	0.22 (0.21)	0.29 (0.51)
Essential drugs costs	0.04 (0.29)	0.05 (0.70)	0.09 (0.44)	0.11 (0.51)
Observations	1,065,674	1,267,539	41,911	55,299

Note: This table presents summary statistics of outcomes measures after 1-to-n exact matching on age, sex, comorbidites, type of municipality, and insurer. Summary statistics for control units are weighted by the inverse number of controls matched to each diabetic.

Table 2.3 summarizes healthcare utilization and costs respectively for each group in the pre- and post-policy periods. Outcomes for control units are weighted as they were in table 2.2. We define outpatient claims as those associated with a hospital length-of-stay (LOS) less than 1 day and inpatient claims as being associated with a

LOS of at least 1 day. Our measure of insulin consumption is the number of insulin claims. Almost all of these claims are for a concentration of 100 UL/ml. As a sanity check, in table 2.3 we note that while the average diabetic has 7.3 insulin claims in the pre-policy and 9.3 in the post-policy period, non-diabetics have virtually no insulin claims before or after the formulary expansion. We see that diabetics have a significantly higher number of claims for outpatient care, lab tests, and office visits than non-diabetics. Diabetics are also costlier than their matched counterparts both before and after the expansion as seen in the second panel of table 2.3.

2.6 Methodology

To estimate the impact of the formulary expansion on outcomes for type I diabetics, we employ a differences-in-differences identification strategy and a generalized linear modeling approach, summarized by the estimating equation:

$$G(\mathbb{E}[y_{it}]) = \alpha + \tau D_i * P_t + \delta D_i + \gamma P_t + \mathbf{x}_i'\beta. \tag{1}$$

Here, y_{it} is the outcome for patient i in year t; D_i is an indicator variable for patient i being a type I diabetic; P_t is an indicator variable for year t following the formulary expansion; and $\mathbf{x}_i$ is a vector of demographic characteristics including sex, age group, comorbidity dummies, insurer dummies, and municipality dummies. The

coefficient of interest is τ, which provides an estimate of the average treatment effect on the treated. G in equation 1 is the link function of the generalized linear model. For each of our outcomes, we specify a link function G and a distribution function F such that $y \sim F$. We assume that y follows a negative binomial distribution for specifications with counts of healthcare claims as outcomes and that y follows a gamma distribution for specifications using healthcare costs as outcomes.[6] Basu et al. (2004) show that the gamma regression model performs better than OLS on log transformed cost data. Alternatively, we could use a generalized beta distribution to choose nested distributions for each healthcare type as described in Jones et al. (2014), but we employ gamma regressions for clarity of exposition. For all specifications, we employ a log link function.

We use as outcomes six categories of healthcare claims and costs: all insulin, continuously covered insulin, outpatient, inpatient, prescriptions, and total. In the case of inpatient outcomes, which are zero for 75% of patient-years, we model y as the outcome of a two-part process. First, we model the probability that a patient is admitted to the hospital. Then, conditional on admission, we model either the number of inpatient claims made during the hospital stay or inpatient costs. We assume y follows a Bernoulli distribution in the first part of the model,[7] followed

[6] For all specifications with counts of claims as the outcome, we provide an estimate of the log-transformed overdispersion parameter, $\ln\alpha$. Likelihood ratio tests that α equals zero strongly reject the Poisson model in favor of the negative binomial across all specifications.

[7] In the two-part model for inpatient claims, we employ a logit link rather than log link in the first stage.

by a truncated negative binomial distribution in the second part of the model for inpatient claims, and a truncated gamma distribution in the second part of the model for inpatient costs. Insulin utilization and costs are subtracted from total and prescription utilization and costs respectively so that any declines from spillovers or rationing are not muted by increases in insulin consumption.

Our identifying assumptions are that the formulary expansion affected all insurance companies and enrollees so there is no selection into the policy, and that our definition of treatment as being type I diabetic is exogenous conditional on all other comorbidities and demographics. Because we use one-to-many matching, we weight our regressions using the weighting scheme described in Iacus et al. (2011b).

In the case of insulin, we note that outcomes for the control group of non-diabetics are mechanically zero across both periods. So, we estimate the effects on insulin outcomes using an interrupted-time-series identification strategy, estimating the following equation on the subsample of type I diabetics:

$$G(\mathbb{E}[y_{it}]) = \alpha + \tau P_t + \mathbf{x}_i'\beta. \tag{2}$$

Here, we assume total insulin utilization follows a truncated negative binomial distribution, truncated at zero, since for type I diabetics insulin consumption is strictly positive.

In equation (2), τ represents the average treatment effect on the treated. This

effect is causal because the formulary expansion is exogenously determined and because, to our knowledge, there were no other interventions in the health system during this period. Additionally, because we implement exact matching of type I diabetics across the pre- and post-policy periods, our results are not biased by changes in type I diabetics' enrollment patterns across time. While exact matching across years controls for changes in enrollment, our results might still be biased if insurers' networks are changing over time in a way that affects their ability to provide care to diabetics. In appendix table 2.A1, we show that the average number of reimbursed providers across municipalities does not change between 2011 and 2013 for the vast majority of insurers. It is therefore unlikely that insurers' ability to provide diabetes care through their network over time biases our estimate of expanded insulin coverage.

We provide evidence of parallel trends in pre-treatment outcomes for treated and control units in appendix 2.12. We see no pre-trend in total claims, outpatient claims, or prescription claims. There are no statistically significant differences in the total costs of type I diabetics and control units in the 9 months preceding the expansion. There is an uptake in inpatient admissions in the quarter prior to the formulary expansion, but no trend prior to this. We do not present the event study estimates as our main results as many of the kinds of healthcare we consider are not received on a quarterly basis, and measuring them so frequently results

in an overwhelming number of zeros. We therefore proceed by implementing the differences-in-differences framework laid out above.

2.7 Results

In table 2.4, we present the results of equations (1) and (2) using as dependent variable the annual utilization for various types of drug and non-drug healthcare utilization. Results are presented as average marginal effects. We turn first to the effect of expanded insulin coverage on insulin consumption by type I diabetics, the relevant outcomes for which are given in columns 1 and 2. These specifications are estimated on the subsample of type I diabetics, since for control units this outcome is by definition zero in both periods. Note that because our measure of insulin utilization is constructed using claims, our outcome does not capture out-of-pocket purchases of insulins not covered by the national formulary. If type I diabetics consumed uncovered insulins in the pre-policy period, then after the formulary expansion, these out-of-pocket purchases would become claims and generate an increase in insulin utilization as a result of the policy, even though insulin consumption patterns would not have changed. Column 1 shows that type I diabetics made 2 (28%) more insulin claims per year with the expansion. Column 2 shows that consumption of regular insulin and insulin NPH declined by 2 claims (27%). The positive impact of expanded insulin coverage on insulin consumption and the substitution it gener-

ated away from continuously covered insulins toward the newly covered insulins suggests that it was potentially welfare enhancing for diabetics who were newly able to consume a more optimal insulin regimen. These results are also consistent with the story of spillovers from drug to non-drug spending, which we explore further in section 2.8.

Table 2.4: Impact of expanded insulin coverage on healthcare utilization

	(1)	(2)	(3)	(4)	(5)		(6)
	All insulin	Continuously covered insulin	All (net insulin)	Outpatient	Inpatient		Prescriptions (net insulin)
Diabetic × policy			-2.82***	-1.91***	-0.04***	0.81***	-1.30***
			(0.26)	(0.11)	(0.00)	(0.15)	(0.18)
Diabetic			34.77***	12.91***	0.13***	5.04***	20.34***
			(0.20)	(0.08)	(0.00)	(0.11)	(0.14)
Policy	2.04***	-2.04***	0.90***	-2.26***	0.00***	-0.64***	2.65***
	(0.04)	(0.03)	(0.05)	(0.02)	(0.00)	(0.04)	(0.04)
Model	Zero-truncated negative binomial	Negative binomial	Negative binomial	Negative binomial	Logit	Negative binomial	Negative binomial
lnα	-1.11***	-0.37***	-0.59***	-0.59***		-0.22***	0.04***
	(0.01)	(0.01)	(0.00)	(0.00)		(0.00)	(0.00)
Observations	97210	97210	2430423	2430423	2430423		2430423

Notes: Cells contain average marginal effects (standard errors). Equation (2) estimated on sample of type I diabetics in columns 1 and 2. Equation (1) estimated on sample of type I diabetics and their exactly matched controls in columns 3-6. All models control for sex, age, comorbidities, municipality, and insurance carrier. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

In column 3, we see that expanded insulin coverage decreased total non-insulin healthcare utilization by type I diabetics by 2.8 claims. This decline in overall utilization is driven by reductions in outpatient and non-insulin prescription drug utilization as well as a fall in hospitalization rates. Outpatient utilization by type I diabetics declined by 1.9 claims, while non-insulin prescription drug utilization fell by 1.3 claims. The estimates of the two part model for inpatient utilization show that the hospitalization rate for type I diabetics decreased by 3.6 percentage points from a baseline rate of 24.9 as a result of the expanded coverage of insulin. Conditional on a hospitalization, inpatient utilization increased by 0.8 claims. These changes in the rate of hospitalizations and the number inpatient claims incurred conditional on admission together imply an overall decline in inpatient care utilization.

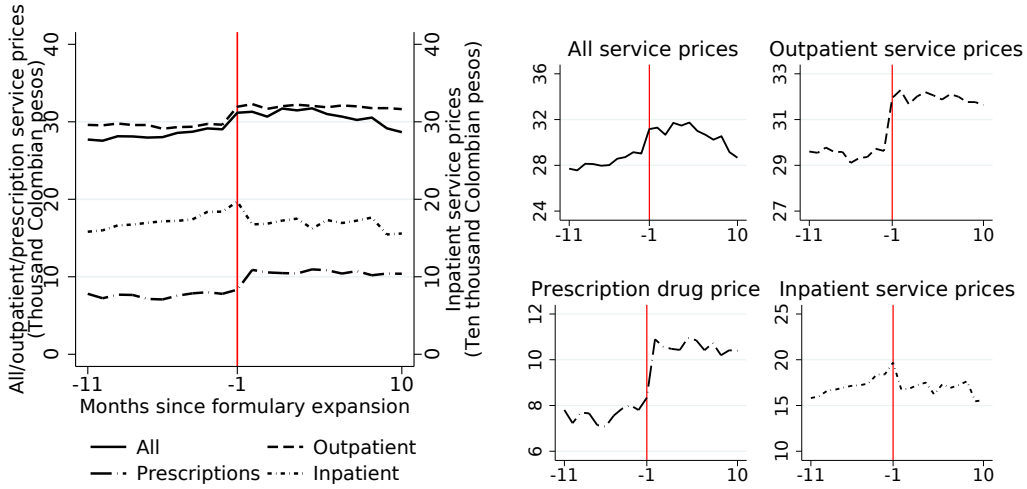
Table 2.5: Impact of expanded insulin coverage on healthcare costs

	(1)	(2)	(3)	(4)	(5)		(6)
	All insulin	Continuously covered insulin	All (net insulin)	Outpatient	Inpatient		Prescriptions (net insulin)
Diabetic × policy			0.04	0.05***	-0.04***	0.37***	-0.06***
			(0.03)	(0.01)	(0.00)	(0.10)	(0.01)
Diabetic			1.22***	0.45***	0.13***	0.84***	0.25***
			(0.03)	(0.01)	(0.00)	(0.07)	(0.01)
Policy	0.58***	-0.06***	-0.14***	-0.14***	0.00***	-0.22***	0.12***
	(0.01)	(0.00)	(0.02)	(0.01)	(0.00)	(0.03)	(0.01)
Model	Gamma	Gamma	Gamma	Gamma	Logit	Gamma	Gamma
Observations	97210	97210	2430423	2430423	2430423		2430423

Notes: Cells contain average marginal effects (standard errors). Equation (2) estimated on sample of type I diabetics in columns 1 and 2. Equation (1) estimated on sample of type I diabetics and their exactly matched controls in columns 3-6. All models control for sex, age, comorbidities, municipality, and insurance carrier. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.5 presents the results of equations (1) and (2) using costs for the same healthcare categories considered in table 2.4. Column 1 shows that the total cost of insulin consumption among type I diabetics increased by 0.58 million COP, a more than threefold increase relative to baseline insulin costs. There is no statistically significant effect of expanded insulin coverage on overall costs net of insulin. While expanded coverage of insulin decreased outpatient utilization, it increased outpatient costs. We plot the trends in average service prices in figure 2.1 to clarify how changes in utilization and service prices are separately affected by the expansion. We see that outpatient service prices increased with the formulary expansion, offsetting the reduction in costs generated from decreased utilization. Because the government's risk adjustment formula controls for only sex, age category, and type of municipality, all of which are covariates in our regressions, the estimated increase in insulin costs directly translates into changes in insurers' profits from coverage of type I diabetics, potentially altering baseline risk selection incentives. In the following section, we look for evidence that insurers respond to these changes in selection incentives by rationing discretionary care for type I diabetics and test competing hypotheses.

Figure 2.1: Average service price time trends for broad healthcare types



Notes: These figures show the trend in weighted average service prices for each type of healthcare category. Weights are computed using 2011 utilization. In particular, for healthcare type $x \in \{\text{All, Outpatient, Prescriptions, Inpatient}\}$ and period $y \in \{2011, 2013\}$, the average service price of healthcare type x in month m of year t is $AverageServicePrice_{xtm} = \sum_{s \in x} \frac{Claims_{st=2011}}{Claims_{xt=2011}} \times \frac{Cost_{stm}}{Claims_{stm}}$.

2.8 Mechanisms

In this section, we test for spillovers from drug to non-drug spending and provide evidence of insurers' rationing of discretionary healthcare provided to type I diabetics. Our tests will examine how proxies for the health status of diabetics and utilization of both discretionary and non-discretionary types of healthcare change with the increased availability of insulin. A theoretical framework of how drug availability impacts the relative effects of spillovers and rationing in the insurers' profit function can be found in appendix 2.13.

2.9 Testing for spillovers

Spillovers from drug to non-drug spending are generated when patients with specific diagnoses take up a drug that has the potential to prevent serious adverse health events. Because some of the diabetics in our sample change their choice of insulin following the expansion of insulin coverage, as seen in table 2.4, there is the potential for such spillovers. Our first test uses the subsample of diabetics to estimate equation (2) using logistic regression. We use as outcome variables indicators for being diagnosed with a complication associated with type I diabetes. We use only the subsample of treated units in these specifications as these diagnoses only apply to type I diabetics by definition and so will mechanically be zero for all non-treated units. The effect of the expanded coverage of insulin on complications from type I diabetes estimated by equation (2) is therefore identified using time series variation in the rate of such diagnoses and can be interpreted as causal for the same reasons laid out in section 2.6. We estimate the effect of expanded coverage on the rates of kidney complications; neurological complications; circulatory complications; other specified complications including diabetic arthropathy, skin complications, oral complications, hypoglycemia, and hyperglycemia; and unspecified complications due to diabetes.[8]

The results of estimating these specifications are presented in table 2.6. Results

[8] ICD10 codes used to create these indicators are E10.2, E10.4, E10.5, E10.6, and E10.8 respectively. Rates of ketoacidosis and ophthalmic complications are not sufficient for estimation.

are presented as average marginal effects times 100. The rates of all complications from diabetes rise as a result of the formulary expansion. We do not employ a balanced panel, so these results are not a mechanical byproduct of the accumulation of chronic conditions over time. Assuming that the rate of complications from diabetes is reflective of the patient's underlying health status, these results in general suggest that health status does not improve with the formulary expansion and constitute evidence against spillovers being the primary mechanism generating the reduction in non-drug care and costs observed in section 2.7.

Table 2.6: Impact of formulary expansion on rate of complications from diabetes

	(1) Kidney complications	(2) Neurological complications	(3) Circulatory complications	(4) Other specified complications	(5) Unspecified complications
Policy	1.49***	1.29***	0.64***	0.72**	1.88***
	(0.18)	(0.13)	(0.13)	(0.22)	(0.24)
Model	Logistic	Logistic	Logistic	Logistic	Logistic
Pseudo R^2	0.09	0.05	0.06	0.03	0.04
Observations	96753	96091	96557	96902	97022

Notes: Cells contain average marginal effects*100 (standard errors). Logistic regression estimation of equation (2) with binary indicator for diagnosis as dependent variable, on the sample of type I diabetics. All models control for sex, age, comorbidities, municipality, and insurance carrier. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

We perform another test for spillover effects which leverages the fact that diabetics who only ever consume continuously covered insulins have no scope for more optimal matching to a newly covered insulin and thus no scope for a subsequently improved health status. We construct a subsample of our data which includes only diabetics whose choice of insulin does not change as a result of the policy. If this subsample experiences declines in healthcare utilization and costs at least as large as those estimated for the full sample, then we will interpret this as evidence against the spillovers hypothesis. We create a sample of type I diabetics that have zero consumption of newly added insulins and for whom the amount of regular and NPH insulin consumed did not change. We use coarsened exact matching (CEM) as in Iacus et al. (2011a) to match diabetics in 2011 and diabetics in 2013 based on demographics, diagnoses, and their level of regular and NPH insulin claims, and estimate equation (2) using the resulting subsample. We also weight our regressions according to the weighting scheme outlined in Iacus et al. (2011a). 58% of all type I diabetics satisfied the sample selection criteria of having no change in insulin consumption and having an exact match counterpart. This test does not make use of control units, and exploits time-series variation in utilization by type I diabetics. Appendix table 2.A2 shows the characteristics of type I diabetics stratified by whether they are matched by CEM. Type I diabetics whose insulin consumption does not change and who are matched by CEM have similar demographic profiles

as those who are not matched, but have lower rates of all comorbidities and are more likely to live in more rural areas.

Table 2.7: Impact of expanded insulin coverage on healthcare utilization for subsample of type I diabetics with no change in insulin consumption

	(1)	(2)	(3)	(4)		(5)
	Insulin	All (net insulin)	Outpatient	Inpatient		Prescriptions (net insulin)
Policy	0.05	-9.27***	-10.17***	-0.07***	2.06***	1.45***
	(0.05)	(0.63)	(0.28)	(0.00)	(0.49)	(0.34)
Model	Zero-truncated negative binomial	Negative binomial	Negative binomial	Logit	Negative binomial	Negative binomial
$\ln\alpha$	-1.41***	-1.24***	-0.99***	-0.26***		-0.94***
	(0.01)	(0.01)	(0.01)	(0.01)		(0.01)
Observations	56537	56537	56537	56536		56537

Notes: Cells contain average marginal effects (standard errors). Equation (2) estimated on the sample of type I diabetics who have zero consumption of newly added insulins and similar consumption of continuously covered insulins before and after the formulary expansion. All models control for sex, age, comorbidities, municipality, and insurance carrier. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results for our second spillovers test are presented in table 2.7. The estimates for the policy coefficient τ capture the causal effect of the formulary expansion as a whole on utilization for this subsample of type I diabetics. As a sanity check of our matching strategy, we note that there is no statistically or economically significant change in insulin consumption as seen in column 1. We use our estimates from tables 2.4 and 2.7 to compute the average marginal effects of the formulary expansion on utilization for all type I diabetics and for the subsample of type I diabetics with no change in insulin consumption. These marginal effects are presented in table 2.8. With the exception of insulin consumption, the effect of the formulary expansion on healthcare utilization in each of the samples are all of the same sign. The formulary expansion's effect on type I diabetics whose insulin consumption does not change is in fact greater than for those whose insulin consumption does change.

Given that changing one's insulin prescription requires an appointment with a primary care physician, it is likely that diabetics who do not alter their insulin consumption in response to the formulary expansion have less contact with the healthcare system overall compared to their counterparts who consume the newly covered insulins. Selection of this kind could explain the lower rates of comorbidities among matched diabetics observed in appendix table 2.A2. While we might expect a lower baseline level of non-drug utilization for this subsample, its limited

scope for spillovers indicates that we would also expect a decline in non-drug utilization smaller in magnitude than the one estimated for the full sample. Instead, we find that the reduction in non-drug healthcare utilization for this subsample is larger than that estimated for the full analysis sample. We now turn to exploring a second mechanism - namely, rationing of care - that might explain the change in drug and non-drug consumption patterns that we observe as result of the formulary expansion.

Table 2.8: Impact of formulary expansion on healthcare utilization of all type I diabetics and those with no change in insulin consumption

	(1) Insulin	(2) All (net insulin)	(3) Outpatient	(4) Hospitalizations	(5) Prescriptions (net insulin)
All type I diabetics	1.99 (1.92, 2.06)	-3.53 (-4.36, -2.71)	-7.05 (-7.46, -6.64)	-0.04 (-0.05, -0.04)	2.71 (2.18, 3.25)
Type I diabetics with no change in insulin consumption	0.05 (-0.04, 0.14)	-9.27 (-10.5,-8.04)	-10.17 (-10.72, -9.62)	-0.07 -(0.08, -0.07)	1.45 (0.78, 2.12)

Notes: Cells contain average marginal effects (95% confidence intervals). Marginal effects of formulary expansion on type I diabetics computed using the estimates in tables 2.4 and 2.7 respectively.

2.10 Evidence of rationing

By making type I diabetics relatively more expensive than other enrollees with similar comorbidity profiles, the formulary expansion incentivized insurers to selectively avoid enrollment from type I diabetics by decreasing the quality of care offered to them. One way to do this is through rationing. Insurers in Colombia can ration care by limiting the provider network or provider choices made available to their enrollees, by requiring authorization for provision of certain services or procedures, or by steering physicians away from recommending certain treatments. Anecdotal evidence in local newspapers and magazines shows that healthcare rationing is a prevailing strategy used by insurers to contain costs. In 2014, *Semana* magazine conducted an investigation that revealed some of the most popular cost containment mechanisms used by insurers: notifying doctors periodically about the expenditures they generate, putting caps on per-patient spending, and denying requests by primary care physicians to refer patients to a specialist or provide expensive diagnostic services. Their investigation noted,

> "Although primary care physicians already have limited access to expensive diagnostic services like CT scans or MRIs, insurers also restrict the use of basic clinical services. In this email, the insurance company states that physicians need to start filing a formulary every time they request a Thyroid Stimulating Hormone (TSH) test for their patients.

> After evaluating every request, the insurer will notify physicians they believe are overprescribing this lab test."

In 2009, *El Colombiano* magazine published,

> "In a study conducted by the National University of Colombia, findings show that out of 458 people who visited their healthcare provider, 17% were denied a medical evaluation. Of those who were evaluated by the doctor, 24.9% were denied laboratory tests and 45% were denied other types of treatment including medications, surgeries, and medical equipment."

Our test for rationing of care is premised on the assumption that insurers will ration discretionary diagnostic services rather than essential healthcare. We construct three measures of discretionary healthcare services - claims for imaging, lab tests, and office visits - as well as a measure of utilization of essential drugs as defined in Tamblyn et al. (2001). The authors define essential drugs as those that "prevent deterioration in health or prolong life and would not likely be prescribed in the absence of a definitive diagnosis." Examples of essential drugs include insulin, inhaled steroids, and beta blockers. We do not include in our measure of essential drug utilization drug classes that were expanded as part of the policy. The full list of essential drugs as well as those that are included in our measure can be found

in table 2.A4 in the appendix. We employ the same differences-in-differences empirical specification as in equation (1) to estimate the impact of the formulary on diagnostic outpatient care and essential drug use.

Table 2.9: Impact of expanded insulin coverage on outpatient & essential drug utilization

	(1) Imaging	(2) Labs	(3) Office vists/ consultations	(4) Essential drugs
Diabetic × policy	-0.09***	-0.94***	0.11**	0.28***
	(0.07)	(0.08)	(0.03)	(0.04)
Diabetic	0.45***	9.43***	2.68***	2.77***
	(0.01)	(0.06)	(0.03)	(0.03)
Policy	0.05***	-0.09***	-0.18***	-0.40***
	(0.00)	(0.02)	(0.01)	(0.01)
Model	Negative binomial	Negative binomial	Negative binomial	Negative binomial
$\ln\alpha$	0.32***	0.27***	-0.99***	1.09***
	(0.00)	(0.00)	(0.00)	(0.00)
Observations	2430423	2430423	2430423	2430423

Notes: Cells contain average marginal effects (standard errors). Equation (1) estimated on the sample of type-I diabetics and their exactly matched controls. All models control for sex, age, comorbidities, municipality, and insurance carrier. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

The results of estimating these specifications are presented in table 2.9. Lab tests see the largest decline in annual utilization as a result of the expanded coverage of insulins, falling by 0.9 claims. Both imaging and office visits see statistically but not economically significant effects on utilization. Essential drug utilization increases by 0.3 claims as a result of the expansion. That essential drug utilization rises

while use of lab tests falls is consistent with a story of rationing of care in which insurers under-provide healthcare that is diagnostic and preventive in nature, but do not ration drugs which are necessary to avoid adverse health outcomes.

Table 2.10: Impact of expanded coverage of insulin on utilization of labs and office visits

	(1) Blood sugar labs	(2) Cholesterol labs	(3) Tryglicerides labs	(4) Creatinine labs	(5) A1C tests
Diabetic × policy	-0.26***	-0.08***	-0.05***	-0.09***	-0.06***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Diabetic	1.75***	1.01***	0.81***	0.87***	0.54***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.00)
Policy	0.06***	-0.05***	-0.02***	-0.03***	0.06***
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)
Model	Negative binomial	Negative binomial	Negative binomial	Negative binomial	Negative binomial
lnα	-0.52***	-0.28***	-0.21***	-0.60***	-0.24***
	(0.03)	(0.01)	(0.01)	(0.01)	(0.02)
Observations	2430423	2430423	2430423	2430423	2430423

Notes: Cells contain average marginal effects (standard errors). Equation (1) estimated on the sample of type-I diabetics and their exactly matched controls. All models control for sex, age, comorbidities, municipality, and insurance carrier. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Having presented evidence that reductions in non-drug spending stem primarily from reductions in discretionary services and in particular from lab tests, we now zoom in and examine which types lab tests are subject to rationing. We decompose lab tests into blood sugar, cholesterol, triglycerides, and creatinine lab tests, which

together comprise more than two-thirds of all lab tests. The difference-in-difference coefficients displayed in table 2.10 show significant reductions in all types of laboratory tests. The largest effect is observed for blood sugar lab tests, which fall by 0.26 claims for type I diabetics. The results presented in tables 2.9 and 2.10 also hold for the subsample of diabetics who exhibit no change in insulin consumption as seen in appendix tables 2.A5 and 2.A6.

The rationing of routine lab tests like those included in table 2.10 is consistent with weakened adherence to guidelines for diabetes management, which recommend lab testing at regular intervals. For example, the U.S. Department of Health and Human Services and the Centers for Disease Control both recommend annual cholesterol and kidney disease testing for diabetics, blood glucose testing every 3 months, and biannual A1C tests (CDC, 2019; National Institute of Diabetes and Digestive and Kidney Diseases, 2016). That these types of lab tests are a routine part of diabetes management is reflected in the fact that type I diabetics in 2011 received over twice as many lab tests as their exactly matched counterparts, a greater differential than any other type of outpatient care. In figure 2.2, we present effect of the expanded coverage of insulin on the share of type I diabetics that are up to date with lab testing. We see that expanded coverage has a small negative effect on the share of type I diabetics that are up to date with blood sugar, cholesterol, and A1C testing. Ellis and McGuire (2007) also find that lab tests are a relatively predictable

type of healthcare and that they are at higher risk of underprovision by insurers. Rationing of these lab tests reduces costs in the short run. In the long run, it may disincentivize the enrollment of diabetics and delay the diagnosis and treatment of comorbidities.

Figure 2.2: Predicted utilization of common laboratory tests by diabetics for years before and after formulary expansion

Notes: Plotted are the 95% confidence intervals of the marginal effects of the formulary expansion. Marginal effects are computed using estimates from specifications of equation (1) using binary indicators for being up to date with lab testing as outcome variables. Patients are up to date with cholesterol and kidney disease testing if they receive at least one lab test annually, blood glucose testing if they receive at least four tests annually, and A1C testing if they receive at least two tests annually.

2.11 Conclusion

In this paper, we measure the impact of expanded insulin coverage on type I diabetics in Colombia. We find that the expansion raises the relative cost of providing health insurance to type I diabetics by increasing the utilization of relatively more expensive types of insulin. Insurers respond to this decreased profitability of type I diabetics mostly by rationing discretionary outpatient care, including lab tests. This targeted rationing in part offsets the more than tripling of insulin costs. The ability to ration care to an identifiable subset of enrollees allows insurers in this market to respond to changes in selection incentives despite having no control over premiums, coinsurance rates, or co-pays. Our results are generalizable to other health systems where public health insurance is provided by private plans whose reimbursements are risk adjusted. This includes the ACA Marketplaces, Medicare Advantage, and Medicare Part D in the U.S. Despite the fact that the risk adjustment schemes in these systems take into account diagnostic and clinical information that makes them more robust than the scheme employed in the Colombian system, there is still evidence of adverse selection in these settings (Juhnke et al., 2016; Newhouse et al., 2013; Montz et al., 2016). Our results indicate that limiting insurers' ability to risk select through premiums and plan design does not keep them from responding to changing selection incentives. Carriers instead respond by rationing the amount of care provided to less profitable enrollees, reducing quality and encouraging disen-

rollment.

We focus on how the expansion of insulin coverage affects the profitability and quality of care received by type I diabetics. We do so because the exclusivity of insulin to diabetics facilitates identification of the causal effect of expanded insulin coverage on healthcare utilization and costs. Our findings, however, are not exclusive to Diabetics and insulin. If insurers can use prediction methods to identify patients who are likely to become less profitable conditional on risk adjustment, we show that they can ration care according to these predictions.

There are several limitations to our study. While our granular claims data allows us to identify the narrow types of discretionary care that insurers ration, our short time frame does not allow us to observe any effects of the policy that may take longer to manifest, such as changes in enrollment patterns and health status. While our focus on type I diabetics has some advantages, we cannot construct a control group for outcomes that are specific to this group. The results presented here suggest that insurers respond to changes in selection incentives with the tools available to them, no matter how unrefined they may be, and that careful consideration must be given to policy changes altering the profitability of identifiable groups of enrollees.

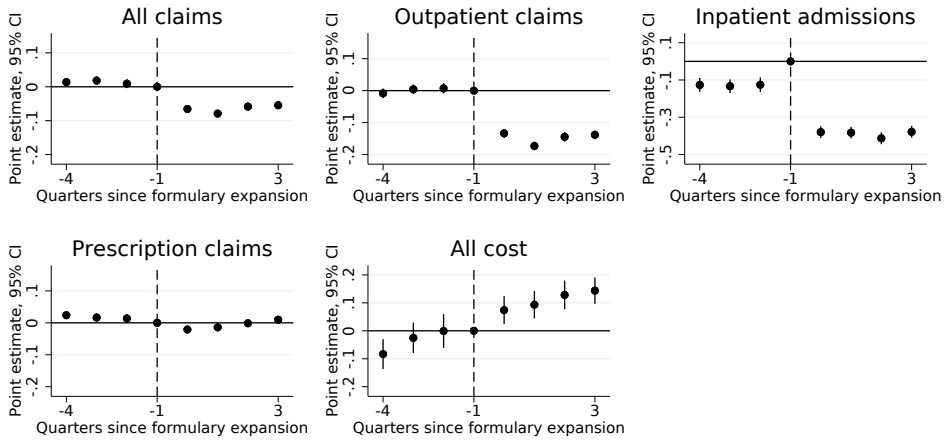
2.12 Appendix: Parallel trends

We test the parallel trends assumption of the differences-in-differences methodology presented in section 2.6 by conducting an event study for our primary outcomes: total claims, total cost, outpatient claims, prescription claims, and inpatient admissions. Let τ_{it} be the number of quarters since enrollee i was treated by the expanded coverage insulins, and let τ_{it} be normalized to -1 for control units. We estimate

$$G(\mathbb{E}[y_{it}]) = \alpha + \beta_{\tau_{it}} + \delta D_i + \eta_t + \mathbf{x}_i'\beta \tag{3}$$

where y_{it} is the outcome for patient i in year t; D_i is an indicator variable for patient i being a type I diabetic; η_t are month-year fixed effects; and $\mathbf{x}_i$ is a vector of demographic characteristics including sex, age group, comorbidity dummies, insurer dummies, and type of municipality dummies. As in the main analyses, we assume that y follows a negative binomial distribution for the total claims, outpatient claims, and prescription claims specifications; follows a Bernoulli distribution for the inpatient admission specification; and follows a gamma distribution for the total cost specification. We employ a log link throughout except in the case of inpatient admissions for which we employ a logit link. We plot the $\hat{\beta}_\tau$s in figure 2.A1 below.

Figure 2.A1: Plot of quarters-since-expansion fixed effects from event study regressions



Notes: Negative binomial regressions estimated for all claims, outpatient claims, and prescriptions claims specifications. Logistic regression estimated for inpatient admission specification. Gamma regression estimated for all costs specification. All specifications control for age group, sex, municipality, comorbidities, and insurance carrier. All and prescription claims include insulin utilization. Dashed vertical line at period before formulary expansion. Solid horizontal line at zero.

2.13 Appendix: Theoretical framework

Let a denote a measure of availability of prescription drugs, d the probability of rationing, TC^D drug-related costs, TC^M non-drug costs, R per patient reimbursement, and Q total demand. An insurer's profits are given by:

$$\pi(a,d) = (R - TC^D(a,d) - TC^M(a,d))Q(a,d) \tag{4}$$

Assume $\frac{\partial Q}{\partial a} > 0$, $\frac{\partial Q}{\partial d} < 0$, $\frac{\partial TC^D}{\partial a} > 0$, $\frac{\partial TC^D}{\partial d} < 0$, $\frac{\partial TC^M}{\partial d} < 0$, so that demand for an insurance carrier is increasing in the availability of drugs and decreasing in

the probability of rationing. Both types of costs are also decreasing in the probability of rationing. If there are spillovers from drug to non-drug spending then $\frac{\partial TC^M}{\partial a} < 0$, otherwise the partial derivative is non-negative. For simplicity assume $\frac{\partial^2 TC^D}{\partial a \partial d} = \frac{\partial^2 TC^M}{\partial a \partial d} = 0$. The availability of drugs is exogenous and determined by the government, while the probability of rationing is a choice variable to the insurer. The insurer's problem is to maximize profits choosing d, the FOC given by:

$$\partial \pi / \partial d = (R - TC^D(a,d) - TC^M(a,d))\partial Q/\partial d - (\partial TC^D/\partial d + \partial TC^M/\partial d)Q = 0 \quad (5)$$

We check whether the profit function is supermodular in (a,d) by taking the derivative of the FOC with respect to a as seen in the equation below:

$$\partial^2 \pi / \partial a \partial d = (R - TC^D(a,d) - TC^M(a,d))\partial^2 Q/\partial a \partial d - (\partial TC^D/\partial a + \partial TC^M/\partial a)\partial Q/\partial d$$
$$- (\partial TC^D/\partial d + \partial TC^M/\partial d)\partial Q/\partial a \quad (6)$$

If there are no spillovers from drug to non-drug spending, $\partial TC^M/\partial a \geqslant 0$ and π is supermodular in (a,d). In this case, exogenous increases in the availability of drugs, as the one generated by the formulary expansion, increases the probability of rationing. In the polar case where spillovers are present and large in magnitude, π is submodular in (a,d). Intuitively, if spillovers are large then insurers need not engage in rationing to achieve cost savings following an increase in the availability

of drugs. For moderate levels of spillovers from drug to non-drug spending, there is scope for a positive relation between a and d.

2.14 Appendix tables

Table 2.A1: Balance table of insurer's network size across markets

Insurer	2011	2013	p-value
A	117 (145)	88.0 (119)	0.37
B	145 (144)	221 (165)	0.10
C	40 (38)	17 (15)	¡0.01
D	74 (74)	41 (60)	0.10
E	7 (14)	20 (36)	0.12
F	188 (191)	215 (187)	0.76
G	11 (13)	16 (39)	0.82
H	140 (42)	54 (16)	¡0.01
I	375 (268)	365 (237)	0.91
J	143 (222)	134 (231)	0.82
K	56 (106)	66 (109)	0.82
L	18 (30)	7 (11)	0.03
M	158 (108)	132 (89)	0.34

Notes: Cells contain the average and standard deviation in parenthesis of the number of unique reimbursed providers for each insurer in the pre- and post-policy periods. P-values for comparison of network sizes across periods are computed using Wilcoxon rank-sum test for continuous variables and Pearson's chi-squared test for categorical and binary variables.

Table 2.A2: Balance table of type I diabetics by whether they are matched in CEM

	Not matched by CEM	Matched by CEM	p-value
Demographics			
Male (%)	43.11	43.22	0.74
Age, mean (sd)	62.72 (13.10)	63.23 (12.59)	¡0.001
Diagnoses (%)			
Arthrosis	3.74	1.29	¡0.001
Cardiovascular disease	73.85	73.13	0.012
Long term pulmonary disease	6.50	3.33	¡0.001
Renal disease	19.26	12.34	¡0.001
Insurer (%)			¡0.001
A	2.13	1.54	
B	8.66	7.08	
C	4.24	0.07	
D	4.58	3.89	
E	5.63	5.59	
F	8.95	10.33	
G	0.94	2.31	
H	5.79	17.54	
I	14.99	11.11	
J	6.05	7.78	
K	0.03	0.04	
L	0.01	0.00	
M	0.35	0.29	
Type of municipality (%)			¡0.001
Metropolitan	67.20	77.25	
Normal	32.47	22.64	
Special	0.33	0.11	
N	40,673	56,537	

Notes: This table shows some descriptive summary statistics of type I diabetics who are and are not matched by conditional exact matching (CEM). Type 1 diabetics who are matched ahve no change in the level or composition of insulin composition across years. p-values for comparison of matched and unmatched type I diabetics are computed using Wilcoxon rank-sum test for continuous variables and Pearson's chi-squared test for categorical and binary variables.

Table 2.A3: Impact of expanded insulin coverage on healthcare costs for subsample of type I diabetics with no change in insulin consumption

	(1) All insulin	(2) All (net insulin)	(3) Outpatient	(4) Inpatient		(5) Prescriptions (net insulin)
Policy	-0.03***	-1.01***	-0.38***	-0.08***	-0.42**	-0.08***
	(0.00)	(0.08)	(0.02)	(0.00)	(0.14)	(0.02)
Model	Gamma	Gamma	Gamma	Logit	Gamma	Gamma
Observations	56537	56537	56537	56456		56537

Notes: Cells contain average marginal effects (standard errors). Equation (2) estimated on the sample of type I diabetics who have zero consumption of newly added insulins and similar consumption of continuously covered insulins before and after the formulary expansion. All models control for sex, age, comorbidities, municipality, and insurance carrier. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.A4: List of essential drugs

Essential drug	Included in measure?
Insulin	No
Anticoagulants	No
Angiotensin-converting enzyme inhibitors	Yes
Lipid-reducing medication	No
Antihypertensives	Yes
Furosemide	Yes
β-blockers	No
Antiarrhythmics	Yes
Asprin	Yes
Antiviral medication	Yes
Thyroid medication	Yes
Neuroleptics	Yes
Antidepressants	No
Anticonvulsants	No
Antiparkinsonian drugs	No
Prednisone	Yes
β-agonists	Yes
Inhaled steroids	Yes
Chloroquines	Yes
Primaquines	Yes
Cyclosporine	Yes

Notes: Essential drugs as defined in Tamblyn et al. (2001) and whether each is included in measure used in table 2.9 column 5. We exclude drugs on the basis of being part of a drug class that was expanded as part of the formulary expansion.

Table 2.A5: Impact of expanded insulin coverage on outpatient & essential drug utilization for subsample of type I diabetics with no change in insulin consumption

	(1) Imaging	(2) Labs	(3) Office vists/ consultations	(4) Essential drugs
Policy	-0.33***	-4.23***	-0.52***	-0.33***
	(0.03)	(0.23)	(0.07)	(0.08)
Model	Negative binomial	Negative binomial	Negative binomial	Negative binomial
$\ln\alpha$	0.28***	-0.41***	-1.35***	0.44***
	(0.01)	(0.01)	(0.01)	(0.01)
Observations	56537	56537	56537	56537

Notes: Cells contain average marginal effects (standard errors). Equation (2) estimated on the sample of type I diabetics who have zero consumption of newly added insulins and similar consumption of continuously covered insulins before and after the formulary expansion. All models control for sex, age, comorbidities, municipality, and insurance carrier. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 2.A6: Impact of expanded coverage of insulin on utilization of labs and office visits for subsample of type I diabetics with no change in insulin consumption

	(1) Blood sugar labs	(2) Cholesterol labs	(3) Tryglicerides labs	(4) Creatinine labs	(5) A1C tests
Policy	-0.85***	-0.41***	-0.36***	-0.34***	-0.07***
	(0.05)	(0.03)	(0.03)	(0.02)	(0.01)
Model	Negative binomial	Negative binomial	Negative binomial	Negative binomial	Negative binomial
$\ln\alpha$	-0.56***	-0.66 ***	-0.54***	-1.03***	-16.23***
	(0.02)	(0.02)	(0.02)	(0.02)	(0.17)
Observations	56537	56537	56537	56537	56537

Notes: Cells contain average marginal effects (standard errors). Equation (2) estimated on the sample of type I diabetics who have zero consumption of newly added insulins and similar consumption of continuously covered insulins before and after the formulary expansion. All models control for sex, age, comorbidities, municipality, and insurance carrier. Robust standard errors in parentheses. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

3 Healthcare Provider Referrals Under Mixed Contracts

3.1 Introduction

In markets where public health insurance is supplied by private insurers, payments from insurers to providers are typically made under either capitation or fee-for-service (FFS) contacts. These forms of reimbursement are antithetical with respect to the incentives they present to providers. Capitation payments are ex-ante lump sum transfers that cover all of a patient's health care claims. Because providers bear the full cost of treatment beyond the capitation payment, reimbursement by capitation may incentivize providers to under-provide services, substitute toward cheaper care, or provide care more efficiently (Frakt and Mayes, 2002; Frank et al., 1995; Hillman et al., 1989; Stearns et al., 1992; Brot-Goldberg and De Vaas, 2018). Papers studying provider behavior under capitation have found evidence that they respond to these incentives without affecting quality of care or patient outcomes (Ho and Pakes, 2014a).

Under a FFS reimbursement scheme, insurers and hospitals negotiate a service price that is paid every time the service is provided. Because provider profits are proportional to the number of services rendered under FFS, this reimbursement scheme generates incentives to over-provide or substitute to more expensive treat-

ments (Shafrin, 2010). Research has found that use of FFS contracts explains much of the rising healthcare costs in countries like the United States (RAND Corporation, 2019), Norway (Grytten and Skau, 2008), and China (Eggleston et al., 2004). The papers that examine provider responses to these polar contracts have focused on settings where one contract type is present between any insurer-provider pair.

In this paper, we study provider incentives and behavior in a setting where contracts are determined at the service level between each insurer-hospital pair. Contract variation at the service level has the potential to exacerbate provider responsiveness to financial incentives as it introduces asymmetric information between insurers and physicians in the hospitals. Providers could adjust their treatment and referral decisions based not only on their belief that a service is covered under a specific contract, but also on how profitable are the services that a patient claims. This paper examines the effect of profit variation across patients within an insurer introduced by service-level contract types, on the referral decisions of primary care providers and the treatment decisions of specialists.

Our setting is the Colombian health care system, characterized by one national health insurance plan with near universal coverage. Individuals choose a private insurer through which to access the national plan. Those above a monthly income threshold pay a monthly contribution to the government for access to the national plan, while those below that threshold have their health care subsidized. Enrollees

do not pay premiums and coinsurance rates, copays, and maximum out-of-pocket expenditures are set by the government.

Private insurers in Colombia form networks of health care providers and negotiate reimbursements for health care services. These negotiations entail determining both a contract type and a price for each service. Almost 40% of claims filed during 2010 were reimbursed under a capitation contract, while another 40% were reimbursed under a FFS contract. The government mandates that certain low complexity services with relatively high demand and low markups must be covered under capitation. While the government recommends that high complexity services be covered under FFS, the contract type is ultimately determined through negotiations between insurers and providers.

For many services there is variation in contract type across insurer-hospital pairs. This means that the services required to treat a single patient might be reimbursed under different schemes. Variation in contract types within patients, thus, raises the question of whether providers are still responsive to financial incentives in their referral and treatment decisions. Leveraging both the highly regulated nature of the setting we study as well as a detailed administrative claims-level data, which contains information on the contract type per claim, we are able to quantify the profitability of any given claim and assess its impact on provider behavior.

We quantify the extent to which hospitals respond to the financial incentives

present under mixed contracts in both their referral and treatment decisions. We do this by focusing on a subset of the population with a specific condition –pregnancy– that requires treatment by a specialist, namely an obstetrician (OB). We use the fact that most referrals in the Colombian health care system, must be made by a primary care provider (PCP). Although pregnant women in particular can go directly to the specialist, there is evidence that they use the PCP channel to access specialized care. In that context, we model the PCP's decision to refer the patient to a specialist within their own hospital (inside referral) or to a specialist at another hospital (outside referral). In the model, PCP's referral decisions are a function of expected profits to their own hospital from providing obstetric care to the woman. Expected profits vary across and within insurers. Profit variation within insurers is generated by variation in contract type across services and variation in the likelihood that a woman uses a particular service. We also examine the influence of procedure prices on the OB's decision as to whether to perform a c-section or vaginal delivery.

There is substantial interest in quantifying and characterizing healthcare provider moral hazard in the specific case of obstetric care. A large literature shows that c-section rates vary substantially across hospitals, even among women with similar health risks. Non-medical geographic factors including provider density and malpractice have been shown to explain part of the variation in c-section rates (Kozhimannil et al., 2013; Baiker et al., 2006). While there is work examining

the role of financial incentives and the treatment decision (Gruber and Owings, 1996), this work has exploited aggregate changes in the profitability of obstetric care rather than variation in the relative prices treatments across insurer-hospital pairs. Whether mixed contracts alleviate provider moral hazard is an important question generally, but is especially important in the specific case of delivery procedure choice. Procedure choice at the time of delivery has been shown to impact the rate of delivery injuries as well as rates of asthma, immune deficiencies, and breastfeeding among infants (Card et al., 2020). The determinants and consequences of delivery choice are also of keen interest to the many medical and international organizations that have called for regulation aimed at reducing the recent and uneven rise in global c-section rates (Teleki, 2020; World Health Organization, 2018; USA Today, 2020).

We develop and estimate a Bayesian learning model of the PCP's referral decision. The PCP's prior beliefs about the profitability of the patient's obstetric care are a function of the share of obstetric services that the woman's insurer reimburses on a FFS basis. The PCP updates his beliefs regarding the woman's profitability based on the realization of claims and contract types for services that are rendered before the referral decision. The PCP's decision therefore depends both on an aggregate patient profitability measure at the insurer-provider level and on the profitability of services the woman in particular is relatively more likely to claim. Whether PCPs

are responsive to this second dimension of profitability is informative both about the role of the agency relationship between the hospital and physician in assessing a patient's profitability and the effect that mixed contracts has on the sorting of patients across providers.

We find that hospitals are responsive to the aggregate patient profitability at the insurer-provider level: a 1% decrease in the expected revenue of a woman's obstetric care results in a 13 percentage point (p.p). decline in the likelihood of an outside referral. We find no evidence that providers respond to variation in profitability across patients within an insurer, suggesting that mixed contracts do not meaningfully change the characterization of provider moral hazard relative to standard settings.

Our findings are consistent with other papers that study insurer and provider responses to financial incentives (McClellan, 2011; Hillman et al., 1989). Variation in FFS prices has been linked to the number and duration of PCP visits (Brekke et al., 2017), the likelihood of providing specialty care (Grant, 2009), and the likelihood of being prescribed a generic drug (Liu et al., 2009). The literature has also found evidence not only of under-provision of services under capitation (Stearns et al., 1992), but also of insurer responses to referral choices. In particular, Ho and Pakes (2014a) and Ho and Pakes (2014b) find that insurers with a higher share of capitated physicians are more price sensitive and more likely to refer their patients to cheaper,

farther away hospitals. Other papers have studied provider behavior when multiple contract types are present across patients and have found providers to be responsive (Hennig-Schmidt et al., 2011). We contribute to this literature by demonstrating that hospitals are most responsive to insurer-level measures of patient profitability, which is informative both the effects of moral hazard and the types of policies would be most effective at ameliorating it.

3.2 Background

The Colombian healthcare system is divided into two income-based regimes: the Contributory regime (CR) and the Subsidized regime (SR). The CR covers all individuals above a monthly income threshold, who pay a monthly contribution to the government for access to the national health insurance plan. Those below the monthly income threshold belong to the SR, and their health care is completely subsidized through tax revenue. Members of both regimes have access to the same national health insurance plan. The health care system has nearly universal coverage, with approximately 51% of the population enrolled to the CR, and the other 49% to the SR.

The design of the national health insurance plan is determined by the government. The plan covers a set of around 7 thousand procedures, services, and devices, as well as more than 700 prescription medications as of 2011, all of which are cho-

sen by the government. This plan is provided by private insurers through a network of providers. Insurers and providers engage in bilateral negotiations over contract types and prices at the service level. The government specifies a list of contract types that insurers and providers may use in their reimbursements. This list includes capitation, FFS, fee-for-package, and fee-for diagnosis. Capitation and FFS are by far the most commonly used contract types, together accounting for over 80% of all claims filed in the CR in 2010. The government mandates that certain common, low-complexity services, like visits to the PCP, must be covered under capitation. For all other services, contract type is decided by the insurer and hospital. Access to any specialist or high-complexity service can only be obtained through a referral from a PCP. This referral may be to another physician within the PCP's own hospital or to a physician at a different hospital; we will refer to the former type of referral as an inside referral and the latter as an outside referral.

Those who enroll in the national health insurance plan do not pay premiums to their insurer. Instead, the government reimburses carriers with risk-adjusted per-capita transfers. The risk adjustment formula controls only for sex, age category, and municipality of residence. The government also sets copays, coinsurance rates, and maximum out-of-pocket expenditures. Theses cost-sharing rules vary across individuals based on their monthly income but do not vary across services, providers, or insurers. For enrollees who made less than 2 times the monthly minimum wage

(MMW), the copay, coinsurance rate, and OOP limit in 2011 were 2,100 pesos, 11.5%, and 57.5%× MMW respectively. For those who made between 2 and 5 times the MMW, the copay, coinsurance rate, and OOP limit were 8,000 pesos, 17.3%, and 230%× MMW respectively. Finally, for those who made over 5 times the MMW, the copay, coinsurance rate, and OOP limit were 20,900 pesos, 23%, and 460% × MMW. These prices have remained unchanged since the establishment of Colombia's health care system in 1993.[1]

3.3 Data

Our data consists of all claims filed in 2011 by pregnant women belonging to the Contributory System in the nine months before their first childbirth. Women whose first childbirth takes place less than nine months into 2011 are left-censored, so we drop patients with childbirths that take place in the first eight months of 2011.[2] We observe the ICD-10 code for each claim and construct patient-level comorbidity indicators by grouping these codes into conditions according to Alfonso et al. (2013). The data also includes the woman's age, municipality of residence, and insurer.

We observe the hospital at which each claim was rendered as well as the specific service associated to the claim. We do not observe the specific physician that pro-

[1] The average exchange rate during 2011 was 1,847 COP/USD.

[2] Appendix figure 3.A1 shows how the total number of observed claims per woman varies according to the month of her childbirth and suggests that a meaningful number of claims are unobserved for censored women.

vides any particular service, only the hospital at which the physician operates. The level at which services are differentiated in the data is very granular. For instance, we are able to distinguish between first-time and follow-up visits, and between visits with primary care physicians and visits with specialists. Importantly, we observe the contract type under which each claim was reimbursed. This is unlike most administrative health care datasets that lack insurer-hospital contract information. For claims covered under FFS, we also observe the claim price, of which the woman pays a fraction equal to her coinsurance rate while her insurance carrier pays the remainder.

We assume that the referral decision is made at the end of the first trimester and define the woman's primary care hospital as the one at which the majority of her primary care visits during the first trimester of her pregnancy happened. All women in our sample have at least one primary care visit in the first trimester. A woman's delivery hospital is the one at which services for c-section or vaginal delivery are administered. We consider a woman to have been referred outside for her obstetric care and childbirth if her delivery and primary care hospitals differ. Of the over 68 thousand women who had their first childbirth in the last quarter of 2011, Two thirds of women visited a primary care hospital that did not perform any childbirths in 2011. Because we are interested in the decisions of PCPs whose hospitals regularly perform childbirths and therefore have the option to refer their patient inside,

we drop from our sample women whose primary care hospital does not perform childbirths. Despite accounting for two-thirds of all women, hospitals that provide primary care but not obstetric services comprise only one-third of all claims filed by primary care hospitals in 2011. Our final sample contains 5,381 unique women who visit 158 unique primary care hospitals and 304 delivery hospitals.

3.4 Model

Here we outline a model of the primary care hospital's referral decision that allows for learning about the expected profitability of individual patients. Pregnant patient i with insurance carrier j sees her PCP at hospital h. During the visit, the PCP decides whether to refer the patient to a specialist within their own hospital or a specialist at another hospital. We denote the former choice as an inside referral I and the latter choice as an outside referral O. Denote the referral decision as $R \in \{I, O\}$. The hospital's payoff from choosing an outside referral is

$$W_{ijh}^{O} = u_{ijh}^{O} + \gamma L_{jh} + \varepsilon_{ijh}^{O}, \tag{1}$$

while its payoff from an inside referral is

$$W_{ijh}^{I} = u_{ijh}^{I} + \gamma \left[L_{jh} + \sum_{s \in \mathscr{S}} \left(c_{sijh} \times Rev_{sijh} - mc_{sh} \right) \right] + \varepsilon_{ijh}^{I}. \tag{2}$$

In equations (1) and (2) above, L_{jh} are capitation payments from insurer j to hospital h which are rendered regardless of referral or treatment decisions. c_{sijh} is the expected share of services of type $s \in \mathscr{S}$ that patient i will require during her pregnancy that will reimbursed on a fee-for-service basis. We will use the word "service" for short to refer to these service types. Rev_{sijh} is the average FFS revenue generated by service s. mc_h the average marginal cost to hospital h from treating a woman's pregnancy, including her childbirth.

At the time of the referral choice, the PCP has observed a sample of contract type realizations for service s and patient i, $\mathbf{x}_{sijh}$, where a realization $x_{sijh} \sim$ Bernoulli(k_{sijh}, ρ_{sijh}). Here, k_{sijh} is the number of claims for service s that patient i has received from her PCP at hospital h and ρ_{sijh} is the likelihood that any one of those k_{sijh} claims is reimbursed on a FFS basis. Let $\mathbf{X}_{ijh} = \left[\mathbf{x}_{1ijh}, \mathbf{x}_{2ijh}, ..., \mathbf{x}_{Sijh}\right]$ be the random matrix of contract type realizations for the set of services $S = n(\mathscr{S})$ services types. Given $\mathbf{x}_{sijh}$, the PCP will refer the patient outside if $\mathbb{E}[W^O_{ijh} - W^I_{ijh} \mid \mathbf{X}_{ijh}] > 0$; that is, if

$$\mathbb{E}\left[\left[u^O_{ijh} - u^I_{ijh}\right] - \gamma\Big[\sum_{s\in\mathscr{S}}\left(c_{sijh} \times Rev_{sijh}\right) - mc_h\Big] + \left[\varepsilon^O_{ijh} - \varepsilon^I_{ijh}\right] \mid \mathbf{X}_{ijh}\right] > 0.$$

Here, we have summed over the service-type specific marginal costs m_{sh} incurred

by hospital h. If we assume that $\varepsilon^R_{ijh} \sim \text{T1EV}(0,1)$, then we can write down the probability of an outside referral as

$$\Pr(R_{ijh} = O|\mathbf{X}_{ijh}) = \frac{\exp\Big(u_{ijh} - \gamma\mathbb{E}\big[\sum_{s\in\mathscr{S}} c_{sijh} \times Rev_{sijh} \mid \mathbf{x}_{sijh}\big] + \gamma mc_h\Big)}{1 + \exp\Big(u_{ijh}\gamma\mathbb{E}\big[\sum_{s\in\mathscr{S}} c_{sijh} \times Rev_{sijh} \mid \mathbf{x}_{sijh}\big] + \gamma mc_h\Big)} \quad (3)$$

where $u_{ijh} = u^O_{ijh} - u^I_{ijh}$.

We now turn to modeling how physicians' beliefs over contract types are formed. We make the following assumptions.

> Assumption 1: Average service type FFS revenue Rev_{sijh} is independent of contract types and perfectly observable to physicians.

> Assumption 2: Contract types are independent across services types $s \in \mathscr{S}$.

With assumptions 1 and 2, we can say that $\mathbb{E}\big[\sum_{s\in\mathscr{S}} c_{sijh} \times Rev_{sijh} \mid \mathbf{x}_{sijh}\big] = \sum_{s\in\mathscr{S}} Rev_{sijh} \times \mathbb{E}\big[c_{sijh} \mid \mathbf{x}_{sijh}\big]$.

The PCP has prior beliefs regarding ρ_{sijh} such that $\rho_{sijh} \sim \text{Beta}(\alpha_{sjh}, \beta_{sjh})$. The physician therefore believes it more likely that the type of care provided by the

specialist is reimbursed under under a FFS contract if $\bar{\rho}_{sjh} = \frac{\alpha_{sjh}}{\alpha_{sjh}+\beta_{sjh}} > 0.5$.

Let k_{sijh}^{CAP} be the number of observed claims covered under capitation and k_{sijh}^{FFS} the number of observed claims covered under FFS so that $k_{sijh}^{CAP} + k_{sijh}^{FFS} = k_{sijh}$.

Proposition 1: Given α_{sjh}, β_{sjh}, and $\mathbf{X}_{ijh}$, a PCP at hospital h believes that the share of claims for service s that patient i with insurer j will require which will be reimbursed on a FFS basis is given by

$$\begin{aligned}\mathbb{E}[c_{sijh} \mid \mathbf{x}_{sijh}] &= \frac{\alpha_{sjh} + k_{sijh}^{FFS}}{\alpha_{sjh} + \beta_{sjh} + k_{sijh}^{FFS} + k_{sijh}^{CAP}} \\ &= \pi_{sijh}\bar{\rho}_{sjh} + (1 - \pi_{sijh})\hat{\rho}_{sijh} \\ &= \pi_{sijh}^{0}\bar{\rho}_{sjh} + \pi_{sijh}^{x}\hat{\rho}_{sijh},\end{aligned}$$

where

$$\hat{\rho}_{sijh} = \frac{k_{sijh}^{FFS}}{k_{sijh}^{CAP} + k_{sijh}^{FFS}},$$

and

$$\pi_{sijh} = \frac{\alpha_{sjh} + \beta_{sjh}}{\alpha_{sjh} + \beta_{sjh} + k_{sijh}^{FFS} + k_{sijh}^{CAP}}$$

We can now write the choice probability in (3) as

$$\Pr(R_{ijh} = O|\mathbf{X}_{ijh}) = \frac{\exp\left(u_{ijh} - \gamma\sum_{s\in S}\left(\pi^0_{sijh}\bar{\rho}_{sjh}Rev_{sijh} + \pi^x_{sijh}\hat{\rho}_{sijh}Rev_{sijh}\right) + \gamma mc_h\right)}{1 + \exp\left(u_{ijh} - \gamma\sum_{s\in S}\left(\pi^0_{sijh}\bar{\rho}_{sjh}Rev_{sijh} + \pi^x_{sijh}\hat{\rho}_{sijh}Rev_{sijh}\right) + \gamma mc_h\right)}.$$

3.5 Descriptives & identification

The model presented in section 3.4 distinguishes between the expected revenue that a woman is expected to generate based on the overall share of services that her insurer reimburses on a FFS basis and the expected revenue that she is expected to generate based on the contract type realizations of claims for service types that are observed by the primary care hospital and are a component of obstetric care. For this to be a meaningful distinction, it must be the case that in practice primary care providers and delivery providers overlap with respect to the service types they provide. Identification of the model in section 3.4 requires meaningful variation in the share of services that are reimbursed on a FFS basis between hospital h and insurer j. In this section, we provide descriptive evidence that our empirical setting satisfies these two criteria.

Table 3.1 provides some summary statistics of our model's objects of interest. Note that while for outside referrals, obstetric care, including procedures related to childbirth, are performed at a hospital other than the primary care hospital, we use

Table 3.1: Summary statistics stratified by referral status

	Inside referral		Outside referral	
Insurer-Hospital share of FFS services, $\bar{\rho}_{sjh}$				
All services	0.24	(0.23)	0.26	(0.25)
Shared services	0.20	(0.22)	0.22	(0.24)
Delivery services	0.32	(0.26)	0.34	(0.28)
Patient share of FFS services, $\hat{\rho}_{sijh}$				
All services	0.26	(0.37)	0.19	(0.34)
Office visits/consultations	0.22	(0.35)	0.16	(0.32)
Lab tests	0.22	(0.40)	0.26	(0.42)
Prophylactic procedures	0.65	(0.47)	0.31	(0.44)
Non-radiology imaging	0.28	(0.45)	0.18	(0.38)
Insurer-Hospital expected FFS revenue, Rev_{sijh}				
All services	1.03	(0.24)	0.97	(0.39)
Shared services	0.38	(0.16)	0.34	(0.26)
Delivery services	0.65	(0.08)	0.63	(0.13)
Patient characteristics				
Age	28.01	(5.47)	27.12	(4.92)
Asthma	0.01	(0.11)	0.00	(0.07)
Cancer	0.13	(0.34)	0.18	(0.39)
Cardiovascular disease	0.05	(0.22)	0.04	(0.21)
Long term Pulmonary Disease	0.01	(0.07)	0.00	(0.00)
Total healthcare utilization				
Prescription drug claims	16.20	(13.22)	23.22	(21.32)
Outpatient claims	56.99	(26.83)	56.53	(32.66)
Inpatient claims	1.17	(4.43)	3.69	(7.23)
Emergency room claims	12.11	(13.34)	11.51	(12.12)
Office visits	15.69	(6.54)	15.01	(7.45)
Observations	3,974		1,407	

Notes: An observation is a woman. Sample contains women whose primary care hospital performs at least one delivery during 2011 and whose childbirth takes place in the last quarter of 2011. Cells contain mean (standard deviation). Shared services include office visits/consultations, lab tests, prophylactic procedures, and non-radiology imaging. Delivery services include induction of vaginal delivery and abdominal delivery procedures.

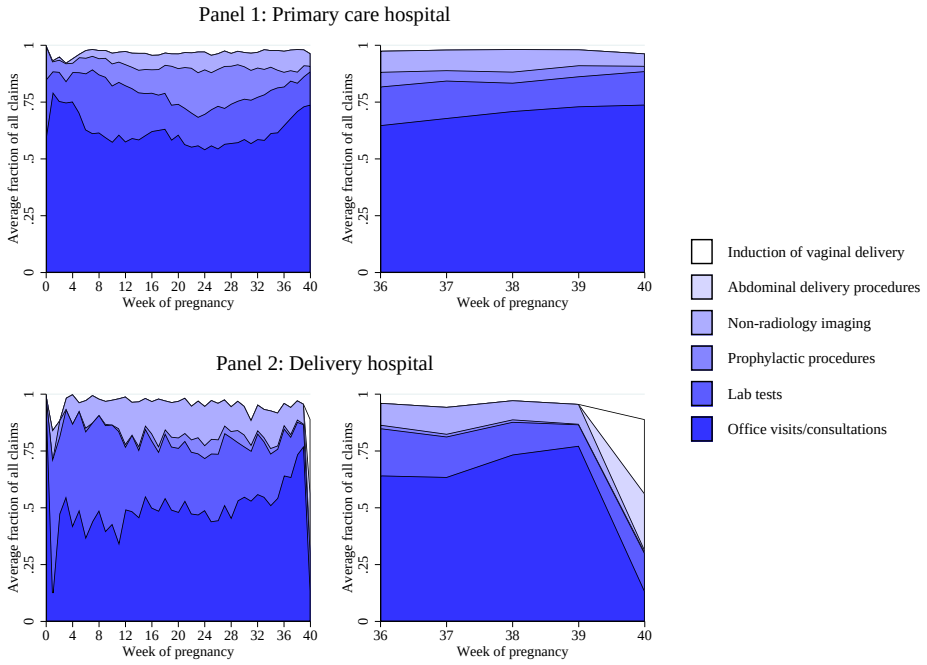
the primary care hospital's contract types and FFS prices for all women to compute the referring provider's expected revenue. We compute Rev_{sijh} as the predicted FFS revenue of service s conditional on patient characteristics including age, comorbidities, and annual healthcare utilization measures. Across all services, predicted FFS revenue is higher for inside referrals than for outside ones. The insurer-hospital level share of services that are reimbursed on a FFS basis, $\bar{\rho}_{sjh}$, is slightly higher for inside referrals than outside referrals. However, the realized share of services provided by the primary care hospital during the first trimester that are reimbursed on a FFS basis is higher for inside referrals. The difference in realized FFS rates is primarily driven by prophylactic procedures and non-radiology imaging.[3] Outside referrals also tend to have higher rates of cancer than inside referrals and to use more of most types of healthcare, including prescription drugs and inpatient services. It will be important to control for these measures of overall health in the empirical specification of the referral decision.

We first demonstrate that there is meaningful overlap between the types of health care that are provided by PCPs and specialists. We classify services into 103 health care types as defined by the Colombian Ministry of Health. Using the subsample of women who receive an outside referral, figure 3.1 shows the average share of total health care utilization provided at the primary care hospital and

[3]Prophylactic procedures are services aimed at disease prevention. They include educational services provided to women on postnatal care.

delivery hospital over the course of her pregnancy. The figure presents the share of six services: office visits/consultations, lab tests, prophylactic procedures, non-radiology imaging, abdominal delivery procedures, and induction of vaginal delivery. We restrict our attention to women who are referred outside since for those referred inside primary care and delivery hospitals are the same. The six services we focus on comprise 89% of all the care provided by primary care providers at any point during the pregnancy and no less than 75% of all care provided by specialists at any point during the pregnancy. While the particular services used by primary care and delivery providers within any given health care type may differ, the fact that almost all services rendered by both providers belong to these health care types suggests that these are the services most relevant for the physician's referral decision.

Figure 3.1: Service type average share of total healthcare utilization by week of pregnancy

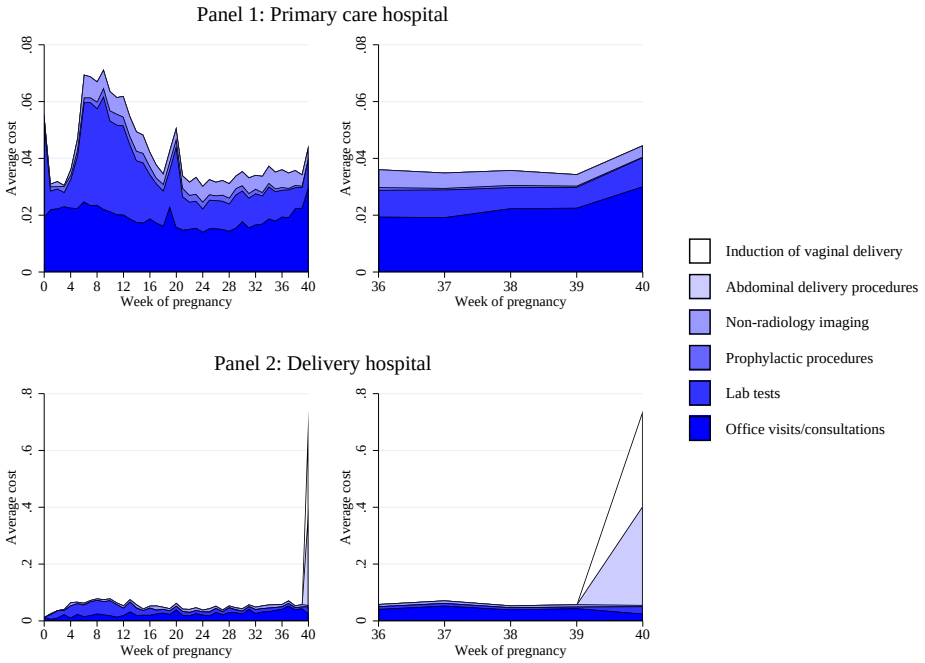


Notes: Figures created with subsample of women of women who are referred outside. Averages are taken across women within a week of pregnancy. A woman's primary care hospital is the one at which she receives the majority her office visits in the first trimester, while her delivery hospital is the one at which claims for childbirth are made.

While both primary care and delivery hospitals provide office visits/consultations, lab tests, prophylactic procedures, and non-radiology imaging, only delivery hospitals render services relating to abdominal delivery procedures, and induction of vaginal delivery. There is therefore no scope for primary care providers to learn about the profitability of these services prior to making their referral decisions. However, primary care providers' priors with respect to the profitability of these

services are likely to be relevant to the referral decision, since these procedures are by far the most expensive of those performed during the pregnancy as can be seen in 3.2. Figure 3.A2 shows that delivery performed in the last week of pregnancy constitutes approximately one-third of the total cost of services provided during the first eight months of pregnancy. We will distinguish between the services provided both by the primary and delivery hospitals and those that are provided exclusively by the delivery hospitals. We refer to the former as shared services and the latter as delivery services. These groups of services differ in that the PCP has scope to learn about the profitability of the former but not the latter, and in that delivery services are much more expensive than shared services and with certainty will be required by the pregnant patient.

Figure 3.2: Service type average cost by week of pregnancy



Notes: Costs are measured in millions of Colombian pesos. Figures created with subsample of women of women who are referred outside. Averages are taken across women within a week of pregnancy. A woman's primary care hospital is the one at which she receives the majority her office visits in the fist trimester, while her delivery hospital is the one at which claims for childbirth are made.

Our model distinguishes between the insurer-hospital level average profitability of service s, $\bar{\rho}_{sjh}$, and the average profitability of realized claims of type s for patient i, $\hat{\rho}_{sijh}$. Identifying the effect of $\bar{\rho}_{sjh}$ on outside referral rates separately from hospital-level marginal costs requires within-hospital variation in average revenue across insurers. Panel 1 of table 3.2 summarizes variation in the share of services reimbursed on a FFS basis across insurer-hospital pairs. Summary statistics are

computed across all services as well as separately for shared services and delivery services. In constructing these measures, insurer-hospital pairs are weighted by the number of unique women they serve. Panel 1 shows that there is meaningful variation across insurer-hospital pairs in the share of services that are reimbursed on a FFS basis, the interquartile range is 0.25 percentage points. This variation, however, does not allow us to separately identify the effect of insurer-level patient profitability measures from hospital-level marginal costs on referral decisions if hospitals do not contract with multiple insurers, or if hospitals have the same contracts with all insurers for all services. The second panel of table 3.2 provides the distribution of the within-hospital coefficient of variation for $\bar{\rho}_{jh}$. More than a third of hospitals have variation in the share of services covered under FFS across the insurers that they contract with. It is this variation that allows us to separately identify marginal costs from financial incentives to providers that vary across insurers.

Table 3.2: Summary statistics for insurer-provider level contract type variation

Panel 1: Share of claims reimbursed under FFS ($\bar{\rho}_{jh}$)

	Mean	S.d.	p25	p50	p75
$\bar{\rho}_{jh}$	0.23	0.23	0.06	0.20	0.29
$\bar{\rho}_{jh}^{SHR}$	0.19	0.22	0.04	0.15	0.22
$\bar{\rho}_{jh}^{DEL}$	0.31	0.26	0.10	0.31	0.43
Observations	182				

Panel 2: Coefficient of variation of $\bar{\rho}_{jh}$ within provider

	Mean	S.d.	p25	p50	p75
CV $\bar{\rho}_{jh}$	0.40	0.64	0.00	0.00	0.85
CV $\bar{\rho}_{jh}^{SHR}$	0.41	0.65	0.00	0.00	0.92
CV $\bar{\rho}_{jh}^{DEL}$	0.39	0.63	0.00	0.00	0.76
Observations	158				

Notes: CV short for coefficient of variation. CV of $\bar{\rho}_{jh}$ computed using mean and standard deviation of $\bar{\rho}_{jh}$ within a provider. Distributions of $\bar{\rho}_{jh}$ and CVs are weighted by number of unique women serviced by each insurer-provider pair or provider.

In order to identify the effect of variation in contract type across services within an insurer-hospital pair, we require that patients within an insurer-hospital pair vary

with respect to the types of services that they use. In table 3.3, we summarize the share of services that the primary care hospital provides during the first trimester of the pregnancy before the referral decision is made, $\hat{\rho}_{ijh}$, across insurer-hospital pairs. We also provide the coefficient of variation of $\hat{\rho}_{ijh}$ across insurer-hospital pairs. In the second panel of table 3.3 we see that there is substantial variation in the share of claims provided to pregnant patients in the first trimester of pregnancy that are covered under FFS, especially for lab tests and prophylactic procedures.

Table 3.3: Summary statistics for insurer-provider level contract type variation

Panel 1: Within insurer-provider mean share of claims rendered in first trimester reimbursed on FFS basis ($\hat{\rho}_{ijh}$)

	Mean	S.d.	p25	p50	p75
$\hat{\rho}_{ijh}$	0.24	0.27	0.01	0.12	0.46
$\hat{\rho}_{sijh}$, s = Office visits/consultations	0.20	0.24	0.01	0.11	0.42
$\hat{\rho}_{sijh}$, s = Lab tests	0.39	0.40	0.00	0.44	1.00
$\hat{\rho}_{sijh}$, s = Prophylactic procedures	0.54	0.48	0.00	1.00	1.00
$\hat{\rho}_{sijh}$, s = Non-radiology imaging	0.48	0.46	0.01	0.35	0.88
Observations	182				

Panel 2: Coefficient of variation of $\hat{\rho}_{jh}$ within insurer-provider

	Mean	S.d.	p25	p50	p75
CV $\hat{\rho}_{ijh}$	3.33	3.76	0.78	1.87	6.91
CV $\hat{\rho}_{sijh}$, s = Office visits	3.57	3.82	0.89	1.88	7.85
CV $\hat{\rho}_{sijh}$, s = Lab tests	4.93	7.76	0.00	1.07	2.15
CV $\hat{\rho}_{sijh}$, s = Prophylactic procedures	3.26	4.29	0.00	0.36	3.38
CV $\hat{\rho}_{sijh}$, s = Non-radiology imaging	0.76	1.22	0.00	0.00	2.28
Observations	158				

Notes: CV short for coefficient of variation. Mean of $\hat{\rho}_{ijh}$ computed across women within an insurer-provider pair. Distributions of the mean of $\hat{\rho}_{ijh}$ and CVs are weighted by number of unique women serviced by each insurer-provider pair or provider.

3.6 Empirical specification

Let O_{ijh} be an indicator that patient i enrolled to insurer j treated at the primary care hospital h is referred outside. We empirically specify O_{ijh} as

$$O_{ijh} = \mathbf{w}_i'\beta + \gamma^0 \bar{\rho}_{jh} Rev_{ijh} + \gamma^x \hat{\rho}_{ijh} Rev_{ijh} + \eta_h + \eta_j + \varepsilon_{ijh} \tag{4}$$

where $\varepsilon_{ijh} \sim T1EV(0,1)$. $\mathbf{w}_i$ is a vector of patient characteristics that capture variation in the mean utility of an outside referral. This includes indicators for individual comorbidities, age, age squared, and measures of various types of healthcare utilization including the total number of prescription drug claims, outpatient claims, inpatient claims, and ER claims. $\bar{\rho}_{jh} Rev_{ijh} = \sum_{s \in \mathscr{S}} \bar{\rho}_{sjh} Rev_{sijh}$ is the PCP's expectation of the revenue to be generated from a patient with insurer j in the second and third trimesters of the pregnancy. $\hat{\rho}_{ijh} Rev_{ijh} = \sum_{s \in \mathscr{S}} \bar{\rho}_{sjh} Rev_{sijh}$ is the PCP's expectation of the revenue to be generated from patient i in the second and third trimesters of the pregnancy based on the realized contract types of services provided in the first trimester of pregnancy. Note that because the primary care hospital does not observe claims for delivery services in the first trimester, then $\hat{\rho}_{ijh} = 0$ for induction of vaginal delivery and abdominal delivery procedures. η_h and η_j are hospital and insurer fixed effects, respectively. Hospital fixed effects capture the average marginal cost of providing care to woman i. Insurer fixed effects help us

capture any source of unobserved insurer heterogeneity that may be correlated with contract types and service prices. We also estimate an empirical specification of the form

$$O_{ijh} = \mathbf{w}_i'\beta + \gamma^{0,Del}\bar{\rho}_{jh}^{Del}Rev_{jh}^{Del} + \gamma^{0,Shr}\bar{\rho}_{jh}^{Shr}Rev_{jh}^{Shr} + \gamma^{x}\hat{\rho}_{ijh}Rev_{ijh} + \eta_h + \eta_j + \varepsilon_{ijh} \tag{5}$$

where we separate the services associated to delivery *Del*, from those that are provided by both the primary care hospital and the delivery hospital, *Shr*. In particular, $\bar{\rho}_{jh}^{Del}Rev_{jh}^{Del} = \sum_{s\in\mathscr{S}^{Del}}\bar{\rho}_{sjh}Rev_{sijh}$, $\bar{\rho}_{jh}^{Shr}Rev_{jh}^{Shr} = \sum_{s\in\mathscr{S}^{Shr}}\bar{\rho}_{sjh}Rev_{sijh}$, and $\mathscr{S}^{Shr} = \{$Office visits, Lab tests, Prophylactic procedures, Non-radiology imaging$\}$. This specification allows us to estimate the effect of expected revenue for shared and delivery services on outside referral rates separately.

3.7 Results

Table 3.4 summarizes the results of estimating equations (4) and (5). Results are presented as the percentage point change in the likelihood of an outside referral resulting from a 1% increase expected revenue. The estimates for specification (4) presented in column 1 indicate that PCPs are much more likely to refer outside women enrolled with insurers that reimburse relatively less for obstetric care, but

that they are not responsive to the observed profitability of care that they actually provide. A 1% increase in the expected revenue of obstetric care at the insurer-hospital level, decreases the likelihood of an outside referral by a physician at hospital h by 13, a nearly 50% reduction relative to the baseline rate of 26%. To put this marginal effect into context, 1% of the expected revenue of obstetric care is equal to approximately 2,000 Colombian pesos. There is no statistically significant effect of the realized profitability of woman i on the likelihood of an outside referral. In column 2 of table 3.4, we separate $\bar{\rho}_{jh}Rev_{ijh}$ into the insurer-hospital level expected revenue from the obstetric service-types provided by both primary care hospitals, $\bar{\rho}_{jh}^{Shr}Rev_{jh}^{Shr}$, and those provided exclusively by obstetricians, $\bar{\rho}_{jh}^{Del}Rev_{jh}^{Del}$. We see that the responsiveness of primary care hospitals to aggregate insurer-hospital profits comes entirely from their responsiveness to the expected revenue of delivery services. Since primary care physicians do not themselves provide these services, this suggests that the decision-making agent driving the referral of less profitable patients to outside hospitals is likely the hospital rather than the primary care physician himself.

Table 3.4: Referral choice logistic regression results

	(1)		(2)	
	$\mathbb{1}\{\text{Outside referral}\}$		$\mathbb{1}\{\text{Outside referral}\}$	
$\bar{\rho}_{jh}Rev_{jh}$, γ^0	-12.64***	(2.97)		
$\hat{\rho}_{ijh}Rev_{jh}$, γ^x	-0.00	(0.50)	-0.04	(0.50)
$\bar{\rho}^{Shr}_{jh}Rev^{Shr}_{jh}$, $\gamma^{0,Shr}$			1.95	(1.87)
$\bar{\rho}^{Del}_{jh}Rev^{Del}_{jh}$, $\gamma^{0,Del}$			-32.97***	(7.98)
Observations	5,244		5,244	
Pesudo R^2	0.46		0.46	

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

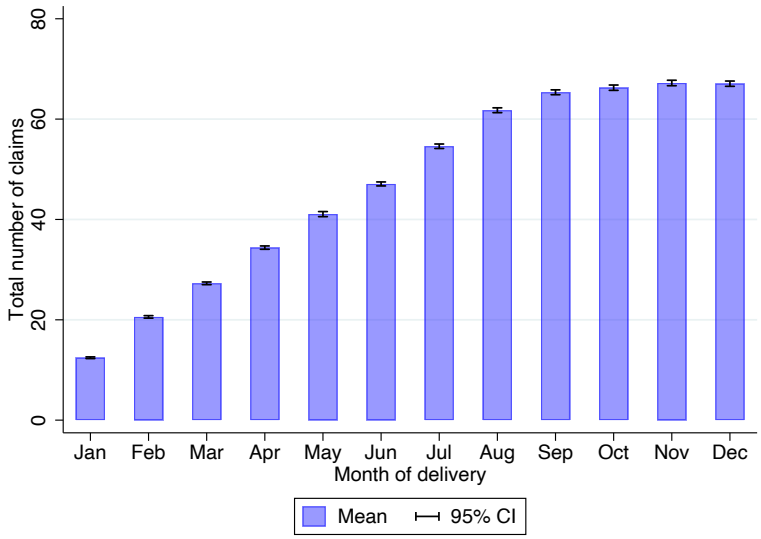
3.8 Conclusion

This paper characterizes physician moral hazard in the presence of mixed contracts. Our context is the Colombian healthcare system, where private insurers provide coverage of the national health insurance plan by engaging in bilateral negotiations over contract types and prices with hospitals. Unlike health care systems like the United States, where agreement over a capitation contract involves having all services subject to capitation, in Colombia insurers and hospitals have discretion over which services to reimburse under capitation and which services under fee-for-service. We exploit this variation in contracts across services to study referral and treatment decisions for pregnant women. For these patients health care is relatively standard: the primary care physician who first sees the woman decides whether to

refer her to an obstetrician in the same hospital or at a different hospital, and the obstetrician decides whether to perform a vaginal delivery or a c-section. Our findings show that there is moral hazard in this setting, but that it is not exacerbated by mixed contracts. Providers are responsive to the aggregate insurer-hospital level of patient profitability, but their referral and treatment decisions do not depend on the woman's particular use of services and their profitability, despite meaningful variation in profitability along this dimension. Allowing insurers and hospitals to more flexibly negotiate contracts, such as at the service level, may allow them to establish more optimal contracts that contain provider responsiveness to financial incentives.

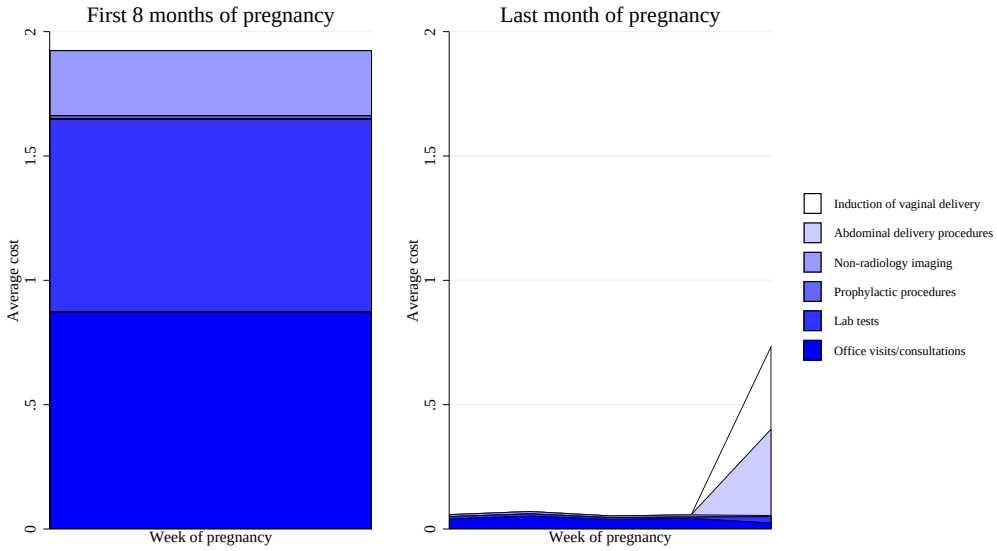
3.9 Appendix

Figure 3.A1: Total number of observed claims by month of delivery



Notes: Mean and standard error of total number of observed claims computed across women stratified by month of delivery.

Figure 3.A2: Service type average cost by period of pregnancy



Notes: Figures created with subsample of women of women who are referred outside. Cumulative cost for first eight months of pregnancy computed by summing over weekly average costs for each service type. A woman's primary care hospital is the one at which she receives the majority her office visits in the fist trimester, while her delivery hospital is the one at which claims for childbirth are made.

References

Abadie, A., Chingos, M., and West, M. (2018). Endogenous Stratification in Randomized Experiments. *The Review of Economics and Statistics*, 100(4):567–580.

Abaluck, J., Agha, L., Chan, D., Singer, D., and Zhu, D. (2020). Fixing Misallocation with Guidelines: Awareness vs. Adherence, Working Paper.

Agency for Healthcare Research and Quality (2019). United States Preventive Services Task Force. https://www.ahrq.gov/cpi/about/otherwebsites/uspstf/index.html.

Agha, L. (2014). The effects of health information technology on the costs and quality of medical care. *Journal of Health Economics*, 34:219–30.

Akerlof, G. (1970). The Market for "Lemons": Quality Uncertainty and the Market Mechanism. *The Quarterly Journal of Economics*, 84(3):488–500.

Alfonso, E., Riascos, A., and Romero, M. (2013). The performance of risk adjustment models in Colombia competitive health insurance market. *Documentos CEDE 012062, Universidad de los Andes - CEDE.*

Allcott, H. and Mullainathan, S. (2010). Behavior and Energy Policy. *Science*, 327(5970):1204–1205.

Américo, P. and Rocha, R. (2020). Subsidizing access to prescription drugs and health outcomes: The case of diabetes. *Journal of Health Economics*, 72:1–19.

Andersen, M. (2017). Constraints on Formulary Design under the Affordable Care Act. *Health Economics*, 26(12):160–178.

Aouad, M. (2019). Are physicians sensitive to changes in patients' demands?: Evidence from a health insurance reform.

Aron-Dine, A., Einav, L., Finkelstein, A., and Cullen, M. (2015). Moral Hazard in Health Insurance: Do Dynamic Incentives Matter? *The Review of Economics and Statistics*, 97(4):725–741.

Bai, L., Handel, B., Miguel, E., and Rao, G. (2021). Self-Control and Demand for Preventive Health: Evidence from Hypertension in India. *The Review of Economics and Statistics.* Retrieved from https://direct.mit.edu/rest/article/doi/10.1162/rest_a_00938/97676/Self-Control-and-Demand-for-Preventive-Health.

Baiker, K., Buckles, K., and Chandra, A. (2006). Geographic variation in the appropriate use of cesarean delivery. *Health Affairs*, 25(5):w355–w367.

Baker, L., Bundorf, K., and Kessler, D. (2014). Vertical Integration: Hospital Ownership Of Physician Practices Is Associated With Higher Prices And Spending. *Health Affairs*, 5(33):756–763.

Basu, A., Manning, W., and Mullahy, J. (2004). Comparing alternative models: log vs Cox proportional hazard? *Health Economics*, 13(8):749–765.

Batty, M. and Ippolito, B. (2017). Financial Incentives, Hospital Care, and Health Outcomes: Evidence from Fair Pricing Laws. *American Economic Journal: Economic Policy*, 9(2):28–56.

Berrington de González, A., Pyo Kim, K., Knudsen, A., Lansdorp-Vogellar, I., Rutter, C., Smith-Bindman, R., Yee, J., Kuntz, K., van Ballegooijen, M., Zauber, A., and Berg, C. (2011). Radiation-Related Cancer Risks from CT Colonography Screening: A Risk-Benefit Analysis. *American Journal of Roentgenology*, 196(4):816–23.

Brekke, K., Holmas, T., Monstad, K., and Straume, O. (2017). Do treatment decisions depend on physicians' financial incentives? *Journal of Public Economics*, 155(1):74–92.

Brot-Goldberg, Z. and De Vaas, M. (2018). Intermediation and vertical integration in the market for surgeons.

Card, D., Fenizia, A., and Silver, D. (2020). The health impacts of hospital delivery practices.

CDC (2019). Your diabetes care schedule. https://www.cdc.gov/diabetes/managing/care-schedule.html.

Centers for Disease Control and Prevention (2001). Utilization of Ambulatory Medical Care by Women: United States, 1997–98 . Technical report, U.S. Department of Health and Human Services, Hyattsville, Maryland.

Centers for Disease Control and Prevention (2021). Quick Facts: Colorectal Cancer (CRC) Screening in Wisconsin. Technical report, U.S. Department of Health and Human Services, Hyattsville, Maryland.

Centers for Medicare and Medicaid Services (2019). NHE Fact Sheet. https://www.cms.gov. Last checked on Dic 23, 2019.

Chen, B., Gertler, P., and Yang, C. (2013). Moral hazard and economies of scope in physician ownership of complementary medical services.

Chen, S., Moreno, C., Duszak, R., and Hornỳ, M. (2021). U.S. Preventive Services Task Force Update and Computed Tomography for Colorectal Cancer Screening Among Privately Insured Population. *American Journal of Preventive Medicine*, 61(1):128–132.

Clemens, J. and Gottlieb, J. (2014). Do physicians' financial incentives affect medical treatment and patient health? *American Economic Review*, 104(4):1320–1349.

Coffman, Lucas C., F. C. R. and Kessler, J. B. (2017). Can Social Information Affect What Job You Choose and Keep? *American Economic Journal: Applied Economics*, 9(1):1–22.

Coscelli, A. and Shum, M. (2004). An empirical model of learning and patient spillovers in new drug entry. *Journal of Econometrics*, 122(2):213–246.

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

Cullis, J. and Propper, P. (2000). *Handbook of Health Economics*, chapter Waiting lists and medical treatment, pages 1201–1249. North-Holland, Amsterdam.

Currie, J. and MacLeod, B. (2017). Diagnosing expertise: Human capital, decision making and performance among physicians. *Journal of Labor Economics*, 35:1–43.

Curto, L., Einav, L., Finkelstein, A., Levin, J., and Bhattacharya, J. (2019). Health care spending and utilization in public and private medicare. *American Economic Journal: Applied Economics*, 11(2):302–332.

Cutler, D., Skinner, J., Stern, A., and Wennberg, D. (2019). Physician payment reform and hospital referrals. *American Economic Journal: Economic Policy*, 11(1):192–221.

de Moor, J. S., Cohen, R. A., Shapiro, J. A., Nadel, M. R., Sabatino, S. A., Robin Yabroff, K., Fedewa, S., Lee, R., Paul Doria-Rose, V., Altice, C., and Klabunde, C. N. (2018). Colorectal cancer screening in the United States: Trends from 2008 to 2015 and variation by health insurance coverage. *Preventive Medicine*, 112:199–206.

Dougherty, M., Brenner, A., Crockett, S., Shivani, G., Wheeler, S., Cocker-Schwimmer, M., Cubilos, L., Malo, T., and Reuland, D. (2018). Evaluation of Interventions Intended to Increase Colorectal Cancer Screening Rates in the United States: A Systematic Review and Meta-analysis. *JAMA Internal Medicine*, 178(12):1645–58.

Douven, R., Halbersma, R., Katona, K., and Shestalove, V. (2014). Vertical Integration and Exclusive Behavior of Insurers and Hospitals. *Journal of Economics and Management Strategies*, 23(2):344–368.

Dranove, D. and Satterthwaite, M. (2000). *The Industrial Organization of Healthcare Markets*.

Duggan, M. (2000). Hospital Ownership and Public Medical Spending. *The Quaterly Journal of Economics*, 115(4):1343–1373.

Duggan, M. and Scott Morton, F. (2010). The Effect of Medicare Part D on Pharmaceutical Prices and Utilization. *American Economic Review*, 100(1):590–607.

Eggleston, K., Ling, L., Qingyue, M., Lindelow, M., and Wagstaff, A. (2008). Health service delivery in china: A literature review. *Health Economics*, 17(2):149–165.

Eggleston, K., Norman, G., and Pepall, L. (2004). Pricing Coordination Failures and Health Care Provider Integration. *Journal of Economic Analysis and Policy*, 3(1):1–31.

Einav, L. and Finkelstein, A. (2018). Moral Hazard in Health Insurance: What we Know and How we Know It. *Journal of the European Economic Association*, 16(4):957–982.

Einav, L., Finkelstein, A., and Mahoney, N. (2018). Provider Incentives and Healthcare Costs: Evidence from Long-term Care Hospitals. *Econometrica*, 86(6):2161â€“2219.

Einav, L., Finkelstein, A., Ryan, S., Shrimpf, P., and Cullen, M. (2013). Selection on Moral Hazard. *American Economic Review*, 103(1):178–219.

Ellis, R. and McGuire, T. (2007). Predictability and predictiveness in health care spending. *Journal of Health Economics*, 26(1):25–48.

Fabbri, D. and Monfardini, C. (2009). Rationing the public provision of healthcare in the presence of private supplements: Evidence from the italian nhs. *Journal of Health Economics*, 28(2):290–304.

FDA (2020). Insulin. https://www.fda.gov/consumers/free-publications-women/insulin.

Fedewa, S., Corley, D., Jensen, C., Zhao, W., Goodman, M., Jemal, A., Ward, K., Levin, T., and Doubeni, C. (2017). Colorectal Cancer Screening Initiation After Age 50 Years in an Organized Program. *American Journal of Preventive Medicine*, 53(3):335–344.

Frakt, A. and Mayes, R. (2002). Beyond capitation: How new payment experiments seek to find the "sweet spot" in amount of risk providers and payers bear. *Health Affairs*, 9:1951–1958.

Frank, R., Glazer, J., and McGuire, T. (2000). Measuring adverse selection in managed health care. *Journal of Health Economics*, 19(1):829–854.

Galenianos, M., Liccardo Pacula, R., and Persico, N. (2012). A search-theoretic model of the retail market for drugs. *Review of Economic Studies*, 79:1239–1269.

Gaynor, M. and Gertler, P. (1995). Moral Hazard and Risk Spreading in Partnerships. *RAND Journal of Economics*, 26(4):591–613.

Gaynor, M., Rebitzer, J., and Taylor, L. (2004a). Physician Incentives in Health Maintenance Organizations. *Journal of Political Economy*, 112(4):915–931.

Gaynor, M., Rebitzer, J. B., and Taylor, L. J. (2004b). Physician Incentives in Health Maintenance Organizations. *Journal of Political Economy*, 112(4).

Geruso, M., Layton, T., and Prinz, D. (2019). Screening in Contract Design: Evidence from the ACA Health Insurance Exchanges. *American Economic Journal: Economic Policy*, 11(2):64–107.

Giedion, U. and Uribe, M. V. (2009). Colombia's Universal Health Insurance System. *Health Affairs*, 28(3):853–863.

Gowrisankaran, G., Nevo, A., and Town, R. (2015). Mergers When Prices Are Negotiated: Evidence from the Hospital Industry. *American Economic Review*, 105(1):172–203.

Gowrisankaran, G. and Town, R. (2003). Competition, Payers, and Hospital Quality. *Health Services Research*, 38(6):1403–1426.

Grant, D. (2009). Physician financial incentives and cesarean delivery: New conclusions from the healthcare cost and utilization project. *Journal of Health Economics*, 28(1):244–250.

Green, B. B., BlueSpruce, J., Tuzzio, L., Vernon, S. W., Aubree Shay, L., and Catz, S. L. (2017). Reasons for Never and Intermittent Completion of Colorectal Cancer Screening after Receiving Multiple Rounds of Mailed Fecal Tests. *BMC Public Health*, 17(1):531.

Grigolon, L., Lasio, L., and Evans, W. (2021). Stigma as a Barrier to Treatment and Adoption of Innovation, Working Paper.

Gruber, J. and Owings, M. (1996). Physician financial decisions and cesarean delivery. *RAND Journal of Economics*, 27(1):99–123.

Grytten, J. Carlsten, F. and Skau, I. (2008). Primary physicians' response to changes in fees. *European Journal of Health Economics*, 9:117–125.

Hackmann, M., Kolstad, J., and Kowalski, A. (2015). Adverse Selection and an Individual Mandate: When Theory Meets Practice. *American Economic Review*, 105(3):1030–1066.

Handel, B. (2013). Adverse Selection and Inertia in Health Insurance Markets: When Nudging Hurts. *American Economic Review*, 103(7):2643–2682.

Handel, B., Hendel, I., and Whinston, M. (2015). Equilibria in Health Exchanges: Adverse Selection versus Reclassification Risk. *Econometrica*, 83(4):1261–1313.

Hennig-Schmidt, H., Selten, R., and Wiesen, D. (2011). How payment systems affect physicians' provision behaviour-an experimental investigation. *Journal of Health Economics*, 30(4):637–646.

Hillman, A., Pauly, M., and Kerstein, J. (1989). How do financial incentives affect physicians' clinical decisions and the financial performance of health maintenance organizations? *New England Journal of Medicine*, 321(2):86–92.

Ho, K. and Pakes, A. (2014a). Hospital choice, hospital prices, and financial incentives to physicians. *American Economic Review*, 104(12):3841–3884.

Ho, K. and Pakes, A. (2014b). Hospital Choices, Hospital Prices, and Financial Incentives to Physicians. *American Economic Review*, 104(12):3841–3884.

Ho, K. and Pakes, A. (2014c). Physician payment reform and hospital referrals. *American Economic Review: Papers and Proceedings*, 104(5):200–205.

Iacus, S., King, G., and Porro, G. (2011a). Causal inference without balance checking: Coarsened exact matching. *Political Analysis*, 20(1):1–24.

Iacus, S., King, G., and Porro, G. (2011b). Multivariate Matching Methods That are Monotonic Imbalance Bounding. *Journal of the American Statistical Association*, 106(493):345–361.

Jelovac, I. (2001). Physicians' payment contracts, treatment decisions and diagnosis accuracy. *Health Economics*, 10(1):9–25.

Johns Hopkins University Health Services Research & Development Center (2008). The Johns Hopkins ACG System Reference Manual. Version 8.2. Technical report, Johns Hopkins University Bloomberg School of Public Health, Baltimore, MD.

Johnson, E. and Rehavi, M. (2016). Physicians treating physicians: Information and incentives in childbirth. *American Economic Journal: Economic Policy*, 8(1):115–141.

Jones, A., Lomas, J., and Rice, N. (2014). Applying beta-type size distributions to healthcare cost regressions. *Journal of Applied Econometrics*, 29(4):649–670.

Jones, R., Devers, K., Kuzel, A., and Woolf, S. (2010). Patient-Reported Barrieres to Colorectal Cancer Screening: A Mixed-Methods Analysis. *American Journal of Preentive Medicine*, 38(5):508–16.

Juhnke, C., Bethge, S., and Mhlbacher, A. C. (2016). A Review of Methods of Risk Adjustment and their Use in Integrated Healthcare Systems. *International Journal of Integrated Care*, 16(4):1–18.

Knight, R., Kanotra, S., Siameh, S., Jones, J., Thompson, B., and Thomas-Cox, S. (2015). Understanding Barriers to Colorectal cancer Screening in Kentucky. *Preventing Chronic Disease*, 12:140586.

Kozhimannil, K., Law, M., and Virnig, B. (2013). Caesarean delivery rates vary tenfold among Us hospitals; reducing variation may address quality and cost issues. *Health Affairs*, 32(3):527–535.

Lamprea, E. and Garcia, J. (2016). Closing the gap between formal and material health care coverage in colombia. *Health and Human Rights*, 18(2):49–65.

Lavetti, K. and Simon, K. (2018). Strategic Formulary Design in Medicare Part D Plans. *American Economic Journal: Economic Policy*, 10(3):154–192.

Levine, S., Malone, E., Lekiachvili, A., and Briss, P. (2019). Health Care Industry Insights: Why the Use of Preventive Services Is Still Low. *Preventive Chorinic Disease*, 16:180625.

Levy, B., Xu, Y., Daley, J., and Ely, J. (2013). A Randomized Controlled Trial to Improve Colon Cancer Screening in Rural Family Medicine: an Iowa Research Network (IRENE) study. *Journal of the American Board of Family Medicine*, 26(5):486–97.

Lin, J., Perdue, L., Henrikson, N., Bean, S., and Blasi, P. (2021a). Screening for Colorectal Cancer: An Evidence Update for the U.S. Preventive Services Task Force. Evidence Synthesis No. 202. AHRQ Publication No. 20-05271-EF-1. Technical report, Agency for Healthcare Research and Quality, Rockville, MD.

Lin, J., Perdue, L. A., Henrikson, N., Bean, S. I., and Blasi, P. R. (2021b). Screening for Colorectal Cancer: An Evidence Update for the U.S. Preventive Services Task Force. Technical report, Agency for Healthcare Research and Quality.

Lin, J. S., Perdue, L. A., Henrickson, N. B., Bean, S. I., and Blasi, P. R. (2017). Screening for Colorectal Cancer: An Evidence Update for the U.S. Preventive Services Task Force. Technical report, Kaiser Permanente Research Affiliates Evidence-based Practice Center.

Liu, Y., Kao, Y., and Hsieh, C. (2009). Financial incentives and physicians' prescription decisions on the choice between brand-name and generic drugs: Evidence from taiwan. *Journal of Health Economics*, 28(2):341–349.

Lundin, D. (2000). Moral hazard in physician prescription behavior. *Journal of Health Economics*, 19(5):639–662.

Marone, V. (2019). Should There be Vertical Choice in Health Insurance Markets?

Martínez-Roa, G. and Serna, N. (2019). Hospital Incentives Under Capitation and Fee-For-Service Contracts.

McClellan, M. (2011). Reforming payments to healthcare providers: The key to slowing healthcare cost growth while improving quality? *The Journal of Economic Perspectives*, 25(2):69–92.

Melnick, G., Zwanziger, J., Bamezai, A., and Pattison, R. (1992). The effects of market structure and bargaining position on hospital prices. *Journal of Health Economics*, 11(3):217–233.

Milcent, C. (2005). Hospital ownership, reimbursement systems and mortality rates. *Health Economics*, 14(11):1151–1168.

Montz, E., Layton, T., Ellis, R. Rose, S., and McGuire, T. (2016). Risk-Adjustment Simulation: Plans May Have Incentives to Distort Mental Health and Substance Use Coverage. *Health Affairs*, 35(6):1022–1028.

Murphy, C. C., Vernon, S. W., Haddock, N. M., Anderson, M. L., Chubak, J., and Green, B. B. (2014). Longitudinal Predictors of Colorectal Cancer Screening Among Participants in a Randomized Controlled Trial. *Preventive Medicine*, 66(1):123–130.

National Institute of Diabetes and Digestive and Kidney Diseases (2016). 4 steps to manage your diabetes for life. https://www.niddk.nih.gov/health-information/diabetes/overview/managing-diabetes/4-steps.

NCI Surveillance, Epidemiology, and End Results Program (2019). Cancer Stat Facts: Colorectal Cancer. https://seer.cancer.gov/statfacts/html/colorect.html.

Newhouse, J., McWilliams, J. M., Price, M., Huang, J., Fireman, B., and Hsu, J. (2013). Do Medicare Advantage Plans Select Enrollees in Higher Margin Clinical Categories? *Journal of Health Economics*, 32(6).

NIH (2020). Preventing diabetes problems. https://www.niddk.nih.gov/health-information/diabetes/overview/preventing-problems.

OECD (2015). *OECD Reviews of Health Systems: Colombia 2016*. OECD Publishing, Paris.

OECD (Accessed September 26 2020). Health resources - pharmaceutical spending. https://data.oecd.org/healthres/pharmaceutical-spending.htm.

Office of the National Coordinator for Health Information Technology (2019). Office-based Physician Electronic Health Record Adoption, Health IT Quick-Stat 50. dashboard.healthit.gov/quickstats/pages/physician-ehr-adoption-trends.php.

Pauly, M. (2004). Medicare drug coverage and moral hazard. *Health Affairs*, 23(1):113–122.

Pickhardt, P. J., Hanson, M. E., Vanness, D. J., Lo, J. Y., Kim, D. H., Taylor, A. J., Winter, T. C., and Hinshaw, J. L. (2008). Unsuspected Extracolonic Findings at Screening CT Colonography: Clinical and Economic Impact. *Radiology*, 249:151–159.

RAND Corporation (2019). Private Health Plans Pay Hospitals 241% of what Medicare Would Pay. https://www.rand.org/news/press/2019/05/09.html. Last checked on Dic 23, 2019.

Regan, T. (2008). Generic entry, price competition, and market segmentation in the prescription drug market. *International Journal of Industrial Organization*, 26(4):930–948.

Reiffen, D. and Ward, M. (2005). Generic drug industry dynamics. *The Review of Economics and Statistics*, 87(1):37–49.

Riascos, A. and Camelo, S. (2014). Una estimación del costo y cambios en el bienestar de los colombianos con el nuevo plan de beneficios en salud. *Revista de Economía del Rosario*, 17(2):299–314.

Riascos, A., Serna, N., and Guerrero, R. (2017). Capital requirements of health insurers under different risk-adjusted capitation payments. *Documentos CEDE*, 6:1–29.

Rodrigues Llorian, E. and Mason, G. (2021). Electronic medical records and primary care quality: Evidence from Manitoba. *Health Economics*, 30:1124–1138.

Ruiz, F., Amaya, L., Garavito, L., and Ramírez, J. (2008). *Precios y Contratos en Salud: Estudio Indicativo de Precios y Análisis Cualitativo de Contratos*. Ministerio de la Protección Social.

Scott Morton, F. (2005). Entry Decisions in the Generic Pharmaceutical Industry. *NBER Working Paper 6190*.

Sequist, T., Zaslavsky, A., Marshall, R., Fletcher, R., and Ayanian, J. (2009). Patient and physician reminders to promote colorectal cancer screening: a randomized control trial. *Archives of Internal Medicine*, 169(4):364–71.

Serna, N. (2019). Spot Prices in Healthcare and the Demand for Health Services.

Serna, N., Carranza, J., and Riascos, A. (2018). The effect of reimbursement contracts on health outcomes: The case of patients with chronic diseases in the colombian health care system.

Shafrin, J. (2010). Operating on comission: analyzing how physician financial incentives affect surgery rates. *Health Economics*, 19(5):562–580.

Siegel, R. L., Miller, K. D., and Jemal, A. (2019). Cancer statistics, 2019. *CA: A Cancer Journal for Clinicians*, 69(1):7–34.

Siu, A. L., Bibbins-Domingo, K., and Grossman, D. (2015). Evidence-Based Clinical Prevention in the Era of the Patient Protection and Affordable Care Act the Role of the U.S. Preventive Services Task Force. *JAMA*, 314(19):2021–2022.

Stearns, S., Wolfe, B., and Kindig, D. (1992). Physician responses to fee-for-service and capitation payment. *Inquiry*, 29(4):416–425.

Superintendencia Nacional de Salud (2010). Indicadores de Calidad de las IPS. https://www.supersalud.gov.co/es-co/Paginas/Delegada%20Supervisi%C3%B3n%20Institucional/Estad%C3%ADsticas-en-Salud-IPS.aspx.

Tai-Seale, M. and McGuire, T. G. (2012). Time is Up: Increasing Shadow Price of Time in Primary-Care Office Visits. *Health Economics*, 21:457–476.

Tai-Seale, M., McGuire, T. G., and Zhang, W. (2007). Time Allocation in Primary Care Office Visits. *Health Services Research*, 42(5):1871–1894.

Tamblyn, R., Laprise, R., Hanley, J., Abrahamowicz, M., Scott, S., Mayo, N., Hurley, J., Grad, R., Latimer, E., Perreault, R., McLeod, P., Huang, A., Larochelle, P., and Mallet, L. (2001). Adverse Events Associated with Prescription Drug Cost-Sharing Among Poor and Elderly Persons. *JAMA*, 285(4).

Teleki, S. (2020). Birthing A Movement To Reduce Unnecessary C-Sections: An Update From California. Health Affairs Blog.

Tenn, S. and Wendling, B. (2014). Entry Threats and Pricing in the Generic Drug Industry. *The Review of Economics and Statistics*, 96(2):214–228.

Thaler, R. and Sustein, C. (2008). *Nudge: Improving Decisions on Health, Wealth, and Happiness*. Yale University Press, New Haven, CT.

US Preventive Services Task Force (2008). Screening for Colorectal Cancer: U.S. Preventive Services Task Force Recommendation Statement. *Annals of Internal Medicine*, 149(9):627–638.

US Preventive Services Task Force (2016). Screening for colorectal cancer: US Preventive Services Task Force Recommendation Statement. *JAMA*, 315(23):2564–2575.

USA Today (2020). Unnecessary C-sections are a problem in the Us. Will publicizing hospital rates change that? https://www.usatoday.com/story/news/health/2020/12/21/the-joint-commission-report-cesarean-section-birth-rates/3943700001/.

Wernli, K., Hubbard, R., Johnson, E., Chubak, J., Kamineni, A., Green, B., and Rutter, C. (2014). Patterns of Colorectal Cancer Screening Uptake in Newly-Eligible Men and Women. *Cancer, Epidemiology, Biomarkers, and Prevention*, 23(7):335–344.

Whitlock, E., Lin, J., Liles, E., Beil, T., and Fu, R. (2008). Screening for Colorectal Cancer: A Targeted, Updated Systematic Review for the U.S. Preventitive Services Task Force. *Annals of Internal Medicine*, 149(149):638–58.

Wisconsin Collaborative for Healthcare Quality (2011). WCHQ Ambulatory Measure Specification: Colorectal Cancer Screening Performance Measure.

Wisconsin Collaborative for Healthcare Quality (2021a). About Wisconsin Collaborative for Healthcare Quality. https://www.wchq.org/about-wchq.

Wisconsin Collaborative for Healthcare Quality (2021b). UW Health Colorectal Cancer Screening Historical Trend. https://reports.wchq.org/trend/71/30/0.

World Health Organization (2018). Who recommendations non-clinical interventions to reduce unnecessary caesarean. https://apps.who.int/iris/bitstream/handle/10665/275377/9789241550338-eng.pdf.

Xiang, J. (2020). Physicians as persuaders: Evidence from hospitals in china.

Yarnall, K., Pollak, K., Ø stbye, T., Krause, K., and Michener, L. (2003). Primary Care: Is There Enough Time for Prevention? *American Journal of Public Health*, 93(4):635–641.

Zapka, J., Klabunde, C. N., Taplin, S., Yuan, G., Ransohoff, D., and Kobrin, S. (2012). Screening Colonoscopy in the US: Attitudes and Practices of Primary Care Physicians. *Journal of General Internal Medicine*, 27(9):1150–1158.

Zauber, A., Knudsen, A., Rutter, C., Naber, S., Doria-Rose, P., Pabiniak, C., Fischer, S., Lansdorp-Vogelaar, I., and Kuntz, K. (2016). Evaluating the Benefits and Harms of Colorectal Cancer Screening Strategies: A Collaborative Modeling Approach to Inform the US Preventive Services Task Force. *Gastroenterology*, 150(4):30279–7.

Zwanziger, J. and Melnick, G. (1996). Can Managed Care Plans Control Health Care Costs? *Health Affairs*, 15(2):1–15.