Intertemporal Substitution in Response to Non-Linear Health Insurance Contracts

Katlyn Hettinger, Michigan State University

Abstract

Non-linear health insurance contracts featuring high deductibles followed by coinsurance rates have become increasingly popular. This work studies whether individuals in the modern U.S. healthcare system optimize spending by consuming elective healthcare at a lower price in the calendar year when the coinsurance arm of the plan is unexpectedly met, and then in the next year reducing elective healthcare consumption. To obtain a causal effect of reaching the coinsurance arm on healthcare consumption in the following year, I use a fuzzy regression discontinuity design comparing those with specific injuries on either side of the new year. Using claim-level data from large firms following privately insured individuals from 2010-2012, I show evidence suggesting there is intertemporal substitution in healthcare consumption. Local average treatment effects indicate that reaching the coinsurance arm in one year leads to $13,263 less healthcare consumed, $788 less paid out of pocket, and 7.4 fewer care dates in the following year. For those induced to consume more healthcare by reaching the coinsurance arm of their plan, I find that that for every dollar of healthcare consumed in the year the coinsurance arm is reached, roughly $0.37 less is consumed in the following year. Ignoring this intertemporal substitution would cause previous estimates using a single year to overstate the cost-saving benefits of high-deductible plans.

1. Introduction

In recent years, deductibles have become an increasingly common part of private health insurance plan structures, and the size of deductibles continues to rise. From 2010 to 2020, the number of employer-sponsored health insurance plans with a deductible over $1,000 for singles rose from 27% to 57%. Among plans with a deductible, the average deductible rose from $917 to $1,644 (Kaiser Family Foundation, 2020). These non-linear plans cause individuals to face very different prices if they meet their deductible during the year and different prices again when their deductible resets at the beginning of each calendar year. When individuals meet their deductible, most face a much lower price based on a coinsurance rate until the start of the next calendar year when the deductible resets.

It is possible that consumers optimize healthcare spending by attempting to consume more elective healthcare at a lower price in the calendar year when the coinsurance arm of the plan is met, and then in the next year avoiding relatively elective healthcare spending. However, little is known about how individuals respond to non-linear health insurance contracts across years in modern private health insurance plans. Across-year intertemporal substitution has been shown in other contexts including dental insurance (Cabral, 2016), and Medicare Part D (Einav et al., 2015). Lin & Sacks (2019) use the RAND Health Insurance Experiment to conclude that failing to account for

intertemporal substitution could cause estimates to overstate savings from high deductible health insurance plans by 20% or more but suggest the importance of examining the topic in a modern setting where there are even more elective and preventative procedures available. This work fills this gap in the literature and suggests that for a group of consumers in modern private health insurance plans, estimates which fail to account for intertemporal substitution could lead researchers to overstate savings from deductible plans by 25% or more.

Within a single year, dynamic incentives, spot prices, and future prices all matter for healthcare consumption choices (Aron-Dine et al., 2015; Brot-Goldberg, 2017; Dalton et al., 2019; Guo & Zhang, 2019; Kowalski, 2016). If within year dynamic incentives are relevant for healthcare consumption decisions, it is likely that incentives across plan years are also important to consider. Guo & Zhang (2019) concludes that, relative to fully-forward looking behavior, the myopia of fathers in responding to nonlinear health insurance plans in the year of childbirth leads to a 21-24% decrease in annual medical spending. Brot-Goldberg (2017) exploits a firm switching from a free-healthcare to a high-deductible plan and estimates that the firm saved 11.8-13.8% on healthcare spending from switching. These works focus on a single plan year either for simplicity or due to data limitations; however, below I show that estimates using only a single year could overstate savings from high-deductible health insurance plans.

In this paper, I investigate whether individuals decrease their healthcare consumption in the year after unexpectedly meeting their deductible. A reduction in the following year suggests that individuals are not just increasing healthcare consumption in response to lower prices, but rather changing the timing or intertemporally substituting their healthcare consumption. For those induced to consume more healthcare by reaching the coinsurance arm of their plan, I find that for every dollar of healthcare consumed in the year the deductible is met, roughly $0.37 less is consumed in the following year.

I use the 2010-2012 IBM MarketScan Commercial Claims Database, which includes a detailed breakdown of individual claim payments and diagnosis and procedure codes. The data is extremely well-suited for this topic, as it follows privately insured individuals and their dependents through the healthcare system for three years. Further, the data includes individuals from a variety of large firms and private insurers, making it representative of a broader population than previous research using a single employer or insurer.

Using a fuzzy regression discontinuity design, I identify the effect of meeting a deductible in one year on healthcare consumption in the following year. I exploit the fact that most health insurance deductibles reset at the beginning of the calendar year by comparing those with a class of unexpected injuries, which would be unlikely to be strategically delayed, in late 2010 and early 2011. We would expect that those who suffer

injuries near the end of a calendar year are similar to those who suffer injuries near the beginning of the subsequent year, except that the year in which their injury occurs changes the probability of them meeting their deductible in 2011. In turn, this variation allows for the identification of the effect of meeting a deductible in one year (2011) on healthcare consumption in the following year (2012).

Using this strategy, I present evidence that there are individuals who substitute healthcare consumption across years to lower prices for elective care. Comparing those with similar injuries in late 2010 and early 2011, I find that those meeting their deductible in one year (2011) consume $13,263 less of healthcare and spend $788 less out of pocket in the following year (2012). To avoid concerns about the propensity to consume more medical care and more expensive medical care being related, I also examine care dates and find a decrease of 7.4 care dates in the year after meeting their deductible. To better understand in what areas individuals respond, I examine classes of elective and preventative healthcare. I find a marginally significant decrease of 12.3 percentage points in the probability of consuming any elective care and fail to detect any significant differences in the usage of any preventive care.

The remainder of the paper proceeds as follows. Section 2 provides relevant background on non-linear health insurance plans. Section 3 develops a theoretical model of intertemporal substitution in response to non-linear health insurance plans. Section 4

discusses the data. Section 5 describes the identification strategy, empirical model, and specification details. In Sections 6 and 7, respectively, I discuss the main results, and a variety of sensitivity analyses. I conclude in Section 8 with a discussion of my findings and their potential relevance to both the literature and insurance markets.

2. Background on Non-Linear Health Insurance Plans

First, I discuss the simple case of an individual with a non-linear (or high deductible) insurance plan[1], and then explain the relevant variations for family plans. The most common form a non-linear plan takes includes a deductible arm, coinsurance arm, and stoploss (or maximum out-of-pocket). The deductible arm is when at the beginning of the plan the consumer is responsible for the entirety of their healthcare costs. This can also be framed as a 100% coinsurance rate.

In reality, preventative care and prescriptions like an annual physical or contraception are often not included in the deductible and instead are subject to a copay or zero out-of-pocket cost. These exemptions make it less likely that effects will be observed among preventive care outcomes. As a simplification, I will not model copays, but will capture them in spending measures, which is common in the literature. Additionally, the Affordable Care Act (ACA) increased the number of services insurance

[1] I use the term high-deductible plan broadly and do not tie it to the legal definition for Health Savings Account eligibility.

plans must provide without consumer cost-sharing. However, despite these exceptions the majority of medical procedures and diagnostic tests are subject to the deductible.

Once the deductible is met by the insured paying for the entirety of their medical costs up to the deductible at the beginning of the plan year, the coinsurance arm is reached. In this section of the plan, the insured is only responsible for paying a certain percentage of their healthcare costs. If a consumer is within the coinsurance arm of their plan and has a 20% coinsurance rate, the out-of-pocket cost for a $500 scan would be $100. The range of the coinsurance arm in terms of total costs is usually much larger than the deductible arm.

Once a certain amount has been paid out of pocket (through the deductible, coinsurance, and copays), the stoploss or out-of-pocket maximum is met. After this stoploss is hit, the insured no longer contributes to the cost of healthcare and is fully insured.

While a single enrollee has a complex non-linear pricing scheme when facing the common deductible, coinsurance arm, and stoploss insurance plan structure, it can become even more complex for a family plan. Many plans include both individual and family deductibles, and out-of-pocket maximums. Most commonly the family deductible and stoploss are two to three times the individual deductible and stoploss.

Consider a family plan where each family member faces an individual deductible of $1,000 and a family deductible of $2,000. For a family of two this is equivalent to individual deductibles of $1,000, but it is not equivalent for larger families. For a larger family, the family deductible means that instead of an individual being guaranteed to pay a $1,000 deductible, they could pay a maximum of $1,000 before reaching the coinsurance arm. The coinsurance arm could be met by no family member meeting their individual deductible and instead multiple family members contributing a sum larger than the family deductible.

A single family member could still hit their individual deductible of $1,000 and then individually move to the coinsurance arm of the plan. Then the threshold for the other family members to reach the coinsurance arm is either $1,000 individually or summed among the other members. I study the effect of reaching the deductible by identifying individuals that reach the coinsurance arm of their plan by meeting either their individual or family deductible.

Figure 1. Probability of Reaching Coinsurance Arm (Meeting Deductible) in 2011

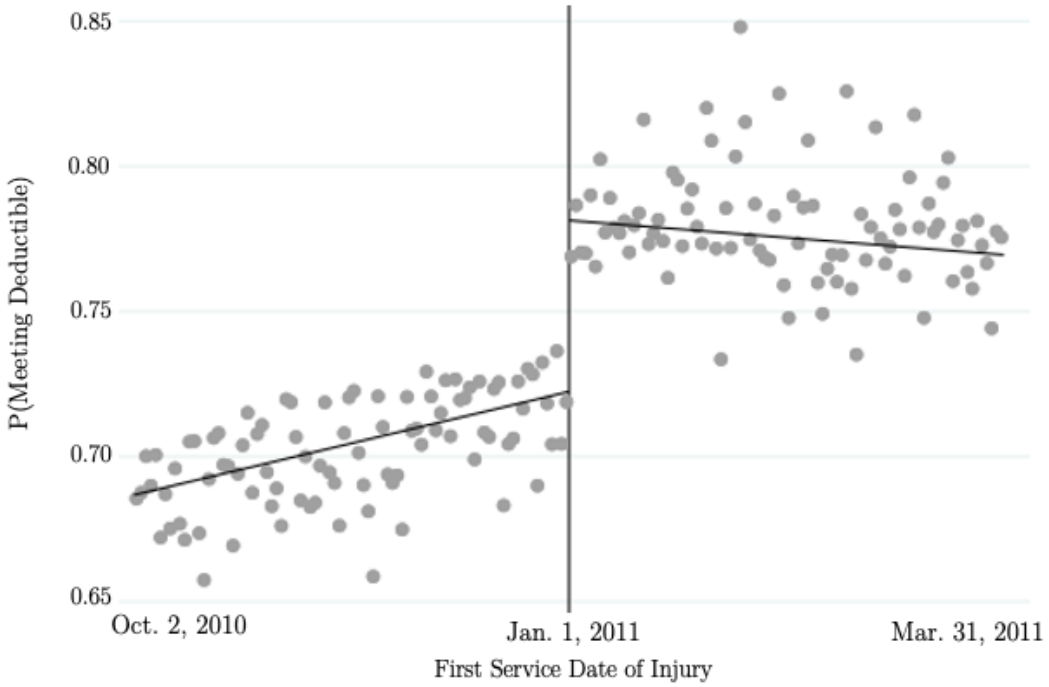

Table 5. Main Specification: Effect of Meeting Deductible in Year t on Year $t+1$ Outcomes

Outcome Form	Total Spending	Total Out of Pocket	Total Care Dates	Outpatient Care Dates	Inpatient Care Dates	Elective Care Dates	Preventive Care Dates
y_{it+1}	-13,263***	-788.4**	-7.39**	-6.19*	-1.56***	-0.36*	-0.20
	(5,005)	(367.5)	(3.50)	(3.29)	(0.59)	(0.19)	(0.18)
$\ln(1+y_{it+1})$	-0.503*	-0.559**					
	(0.276)	(0.235)					
$\mathbf{1}(y_{it+1}>0)$					-0.0863**	-0.123*	-0.0667
					(0.042)	(0.064)	(0.059)
Mean of y_{it+1}	11,631	1,468	16.17	15.8	0.52	0.54	1.08

$SD\ of\ y_{it+1}$	32,495	1,900	19.42	18.4	3.64	1.47	1.31