Intertemporal Substitution in Response to Non-Linear Health Insurance Contracts

Katlyn Hettinger, Michigan State University

June 2023

Abstract

Health insurance contracts with high annual deductibles have become increasingly popular in the U.S. This feature of insurance contracts allows consumers to substitute healthcare in one period for healthcare in another period by, for example, increasing consumption in the year the annual deductible was met and decreasing future consumption. I obtain an estimate of the causal effect of meeting the deductible on healthcare consumption in the following year. I exploit variation in the timing of an injury that generates significant healthcare expenses and a regression discontinuity design to identify the effect of meeting the deductible. Data for the analysis are from the Marketscan database of medical claims on privately insured individuals at large firms. Estimates indicate that there is intertemporal substitution in healthcare consumption. 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 for every dollar of discretionary healthcare consumed in the year the coinsurance arm is reached, roughly $0.56 less is consumed in the following year.

1. Introduction

Insurance plans are increasingly using annual deductibles, which force consumers to take on the full price at the beginning of the plan year, to address moral hazard. However, when individuals meet their deductible, they face a much lower out-of-pocket price based on a coinsurance rate until the start of the next calendar year when the deductible resets. This non-linear plan design causes individuals to face sharp price changes that may create dynamic incentives. There is currently limited knowledge on to what degree consumers respond strategically to these dynamically changing out-of-pocket prices.

Given the dynamically changing out-of-pocket price, consumers have the ability to substitute consumption from the high-price periods to the low-price period, for example, by increasing discretionary consumption after meeting the deductible and before the end of the plan year. This response includes "ex post moral hazard" where individuals change their consumption in response to the current out-of-pocket prices they face (i.e., paying only the coinsurance rate), and intertemporal substitution where individuals pull forward consumption that they would have consumed at some point in the future in response to the change in relative prices across time periods. The ability to avoid the deductible, at least for consumption that is discretionary and for which the timing is not significant, suggests that high-deductible plans may not be as effective in reducing consumption as generally thought.

My work contributes to this research area by investigating whether individuals decrease their healthcare consumption in the year after unexpectedly meeting their deductible in the context of private health insurance. 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. My research design is able to isolate what portion of the increase in spending is intertemporal substitution versus moral hazard. It is important to differentiate these responses because they may have varying implications for optimal plan design. Further, this specific setting is significant as high deductibles have become an increasingly common part of private health insurance benefit designs in recent years. From 2010 to 2020, the share 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).

To obtain estimates of the effect of meeting the deductible, I use a fuzzy regression discontinuity design. I exploit differences in the timing of experiencing a major healthcare event —an injury—that causes differences in the timing of meeting a deductible among otherwise similar people. Those who are injured earlier in the year and hit the deductible earlier as a result have more opportunity to exploit the lower cost-sharing associated with meeting the deductible to substitute current healthcare for future healthcare.

I use the 2010-2012 IBM MarketScan Commercial Claims Database, which is well-suited to answer the research question because it follows privately insured individuals and their dependents throughout the healthcare system for three years. Further, the data includes individuals from a variety of large firms and private insurers, making it more representative of a broad population than previous research using a single employer or insurer.

Results of the analysis indicate that individuals substitute healthcare consumption across years in response to dynamically changing out-of-pocket price. Comparing those with similar injuries in late 2010 and early 2011, I find that those meeting their deductible in one year (i.e., 2011) consume $13,263 less of healthcare and spend $788 less out of pocket in the following year (i.e., 2012). For those induced to consume more healthcare by reaching the coinsurance arm of their plan, I find that for every dollar of discretionary healthcare consumed in the year the coinsurance arm is reached, roughly $0.56 less is consumed in the following year.

Broadly, this work contributes to the literature on consumer responses to health insurance contract design. 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 (Klein et al., 2022). These works focus on a single plan year either for simplicity or due to data limitations; however, below I show that certain estimates using only a single year could overstate savings from high-deductible health insurance plans.

My paper complements recent studies on intertemporal substitution in response other insurance features including dental insurance annual maximum benefits (Cabral, 2016), a Swedish policy eliminating primary care copayments for the elderly (Johansson, 2023), and the nonlinear contract design (Einav et al., 2015) and anticipation of program implementation (Alpert, 2016) of Medicare Part D. The most closely related work, 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. However, little is known about how individuals respond to non-linear health insurance contracts across years in the modern private US health insurance market.

My contributions to the literature are threefold. First, I add to the literature on response to dynamically changing prices by documenting consumers intertemporal substitution in response to varying out-of-pocket prices across years. While this response in the health insurance context is interesting in its own right as health care spending is a sizable portion of the US economy, the findings in this article may also shed light on how consumers are more generally able to strategically respond to non-linear pricing schemes and substitute intertemporally. Second, I use a new fuzzy regression discontinuity design leveraging accidental injuries around the plan year change to identify the effect of meeting a deductible. Third, I contribute to the literature on consumer responses to health insurance by quantifying the causal effect of meeting the deductible on healthcare consumption in the following year. This further demonstrates the importance of accounting for across-year intertemporal substitution in estimates of cost savings from non-linear health insurance plans and is especially important as recent works have estimated price elasticities using a single plan year which fails to account for across-year substitution that undermines cost savings.

2. Background on Non-Linear Health Insurance Contracts

2.1 Non-Linear Health Insurance Contract Design

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 a maximum out-of-pocket limit (or stoploss). The deductible is usually an annual limit and before reaching it the consumer is responsible for the entirety of their healthcare costs.

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

In practice, several services, for example, 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.[2] However, despite these exceptions the majority of medical procedures and diagnostic tests are subject to the deductible.

Once the deductible is met, the coinsurance arm is reached. In this period, the insured is only responsible for a coinsurance amount. 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. Once a certain amount has been paid out of pocket (through the deductible, coinsurance, and copays), the stoploss or out-of-pocket maximum applies, and the insured person has no cost-sharing during this period.

The complexity of the typical high-deductible plan increases in the context of a family unit. 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.

[2] Additionally, the Affordable Care Act (ACA) increased the number of services insurance plans must provide without consumer cost-sharing.

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.

2.2 Additional Related Literature

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 (Klein et al., 2022). 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.

3. Data Description

3.1 Dataset

Data for the analysis comes from the 2010-2012 IBM Truven Health MarketScan Commercial Database which is obtained from large companies and private health insurers across the United States. The sample is not representative of the U.S. population overall but may reflect reasonably well the population of people with large firm private insurance—a non-trivial group.

Further, because the dataset comes from a variety of employers and companies across the U.S., I am able to study a broader privately insured population than studies with data from a single employer or insurer.

The MarketScan database contains all insurance claims for individuals and their spouse/dependents including inpatient, outpatient, and pharmaceutical claims. Each claim provides information on the total amount paid by the insurer and out-of-pocket payment categorized as payment towards the deductible, coinsurance, or copayment. I aggregate this claim-level information to the individual level for all analyses. The dataset also includes limited demographic information on the age and sex of all individuals.

Claims data allows me to observe, in detail, the spending, diagnoses, and procedures of individuals. The major limitation of this claims data is that I do not observe individuals without any claims in a year. However, any small claim within the year such as an influenza vaccine or prescription refill would lead to inclusion.

To follow changes in healthcare consumption over time, I must limit my sample to those observed in my sample all three years. This means that the sample represents a group where the primary enrollee is linked to the same employer or private insurer for three consecutive years. I also exclude individuals under the age of 18 out of concern that guardians may be more altruistic towards their children and not be willing to delay their care, but I also show robustness to their inclusion in Appendix Table A1.

To select a sample of those with unexpected injuries, I use the injuries selected by Kowalski (2016), which were selected based on the fact that individuals that have the injury in their families do not spend more on their own medical care before the injuries occur. Those that have one of these injuries would make significant progress towards, if not meet, their deductible in the year that the injury occurs. Based on their selection, this class of injuries does not appear to be strategically timed in any manner.

In Table 1, I show the identifying injuries, their ICD-9 codes, and the counts overall and by year of injury for the 90-day bandwidth sample. The most common identifying injury is sprains and strains of joints and adjacent muscles occurring in 34 percent of the sample. Other common injuries occurring in roughly ten percent of the sample include fractures, dislocations, open wounds, contusions, and complications of trauma.

Summary statistics of observed covariates overall and by injury year are displayed in Table 2. The summary statistics across the two sides of the discontinuity appear quite similar. The sample is 58% women, and the most common age range is 45-54. The mean individual deductible in 2011 is $608. Table 3 presents summary statistics of the 2012 outcomes overall. The mean of total spending is $11,631, while the 99th percentile of total spending is $125,763.[3] The out-of-pocket mean is $1,468, while the 99th percentile, $7,121, is relatively smaller because out-of-pocket maximums exist on nearly all plans. The average number of care dates in 2012 is 16 with the majority being outpatient, while only 9% of individuals have an inpatient claim. In 2012, 31 and 60% have at least one elective or preventive claim observed, respectively.

We would expect those with the same injuries occurring a short time apart to be incredibly similar. However, when in the year the injury occurs determines how long the individual has to benefit from meeting their deductible and thus reaching the coinsurance arm of their plan. Therefore, an individual with an injury occurring in the end of the year would have little time to react from the increased likelihood of meeting their deductible, while an individual with the same injury in the beginning of the subsequent year could have twelve months to react

[3] See Section 3.3 for definitions of outcomes.

to the increased likelihood of reaching their deductible. In Figure 1, I show the probability of reaching the coinsurance arm (i.e., meeting the deductible) by the first date the injury is observed. The probabilities of reaching the coinsurance arm in 2011 for those with a 2010 or 2011 injury are 70.4 and 77.6, respectively (Table 4).

In my main specification, I focus on those with the first observed date of the injury occurring within 90 days of the 2010–2011-year change. This selection of a 90-day bandwidth is based on the average across outcomes for data-driven bandwidth selection methods following the procedures of Calonico, Cattaneo, and Titiunik (2014a, b, 2015b). I additionally show robustness to other bandwidths in Section 7.

3.2 Determining Deductible Amount

The major limitation of this dataset is that it does not include specific plan details such as the deductible or out-of-pocket maximum. But I am able to back out deductible amounts for over 90% of individuals through a few reasonable assumptions and data features that are very similar to the procedures used in Guo and Zhang (2019). First, I assume that all family members will have the same deductible and that the family deductible is two times the individual deductible.

I also have a plan key that identifies all individuals on the same insurance plan for a portion of the sample and assume that all individuals with the same plan key have the same deductible. Because deductibles are most often chosen at common increments, I assume that individual deductibles must be multiples of 25 and family deductibles must be multiples of 50.

For each individual and each family, I sum the deductible spending throughout each calendar year. If the individual sum of deductible spending is a multiple of 25 or the family sum is a multiple of 50, I consider that to be the most likely deductible. Then for each plan key, I consider the most common deductible amount to be the deductible for all with the plan key with the exception of zero. If the most common deductible amount is zero, I consider the deductible to be zero if greater than 90% of those on the plan have a deductible of zero. If not, then the second most common deductible on the plan is chosen. This is done because some plans have deductibles but offer many deductible exemptions for things like standard office visits.

For those with no plan key, I assume the deductible is the sum of deductible spending observed for the individual (family) if it is a multiple of 25 (50). If I still have not found a deductible amount for a family in 2011, then I use the deductible estimated for 2012 or 2010, respectively, since it is reasonable to assume that those enrolled at the same employer would enroll in similar plans across years. In my main analysis, I exclude those with a deductible of $100 or less as the instrument shows little validity for this group (Table 7).

3.3.1 2012 Outcomes

The primary outcome of interest is total healthcare spending in year $t + 1$ (i.e., 2012). Total spending is defined as the total amount spent by the insurance company and individual out of pocket on all healthcare claims (outpatient, inpatient, and pharmaceutical) in year $t + 1$. To examine the spending most relevant to the healthcare consumer, I define the outcome total out-of-pocket spending on all healthcare claims (outpatient, inpatient, and pharmaceutical) in year $t + 1$. Out-of-pocket spending is the sum of all payments by the individual towards the deductible, coinsurance, or copayments. To address concerns about the skewed distributions of these spending outcomes, I also examine both outcomes with natural log, $\ln(1 + y_{it+1})$, and inverse hyperbolic sine, $arsinh(y_{it+1})$, transformations.

To examine a different margin and avoid concerns about the propensity to consume more medical care and the propensity to consume more expensive medical care being related, I also examine total care dates. A count of dates rather than a count of services is used because it is often difficult to distinguish separate services in claims or they may be charged as a bundle of services. I define care dates as a count of the number of service dates in year $t+1$ that an individual has outpatient or inpatient claims.

To better understand in what areas individuals respond, I examine outpatient and inpatient care dates separately, and classes of elective and preventive healthcare. Outpatient and inpatient care dates are counts of the number of service dates in year $t+1$ that an individual has outpatient or inpatient claims, respectively. The elective care dates outcome is a count of the number of dates in year $t+1$ with elective services being defined based on elective procedures defined using Berenson-Eggers Type of Service (BETOS) codes by Clemens and Gottlieb (2014) and also used by Guo and Zhang (2019). This class of elective services includes procedures such as cataract removal, joint replacement, colonoscopy, and minor skin procedures. The preventive care dates outcome is a count of the number of dates in year $t+1$ with preventive services being defined based on the Center of Medicare and Medicaid Services' list of preventive services. Examples of these preventive services include annual wellness visits, influenza vaccinations, and disease screenings. The full list of elective and preventive services is available in the Appendix. For outcomes that do not always occur and zeros are common (inpatient, elective, and preventive), I also examine the outcome as an indicator of any of that type of care consumed in the year.

Tables 2 and 3 present summary statistics on the year t+1 (i.e., 2012) outcomes. Mean total healthcare spending and out-of-pocket spending in year $t+1$ are \$11,631 and \$1,468, respectively. The average number of care dates is 16.2 with outpatient and inpatient averages of 15.8 and 0.5. Only nine percent of the sample has any inpatient care while 31 and 60 percent have any elective or preventive care, respectively.

3.3.2 2011 Outcomes

To give an estimate of the tradeoff between discretionary spending in the year the deductible is met and the following year, I also estimate the models for the outcomes in year t (i.e., 2011). First, I separate the outcomes in year t into those that are related to the injury or not. For the total healthcare spending outcomes, I measure spending related to the injury as the totals from any claim that has one of the injury ICD-9 codes. I then also measure the complement, spending unrelated to the injury, which is the total of any spending in year t which does not have one of the injury ICD-9 codes. However, it is possible that the injury codes are not always used on claims for follow-up care. This would lead to an underestimate of the tradeoff of discretionary spending because the unidentified follow up care would inflate the denominator.

To avoid issues of errors in measurement of discretionary care, I also measure all spending and care dates in year t, regardless of injury. These measures in year t contain the cost of treating the original injury for those with the first date the injury is observed in the beginning of year t. Follow-up care for these injuries could be contained in the measures for anyone in the sample.

In Table 4, I present averages of year t (2011) and year $t + 1$ (2012) outcomes by the year of injury to show that the expected patterns are observed in the raw data. Year t total spending is higher for those with an injury in year t comparing \$14,804 to \$11,761. Then in the following year average spending is lower for those with a year t injury relative to a year $t - 1$ injury (\$11,526 to \$11,726).

4. Empirical Strategy

I use a fuzzy regression discontinuity design to 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 on either side of the calendar year are similar, except that the year in which their injury occurs changes the probability of them meeting their deductible in 2011. My model relies on this variation to identify the effect of meeting a deductible (i.e., reaching the coinsurance arm) in one year (2011) on healthcare consumption in the following year (2012).

4.1 First Stage

I implement the fuzzy regression discontinuity using a local linear regression with a bandwidth of 90 days and a rectangular kernel. I use an instrumental variables estimation framework (Imbens and Lemieux, 2008) where the first stage of the model is:

$$w_{it} = \gamma_0 + \gamma_1 Injury_{it} + \gamma_2\, Injury_{it-1}\,'\, Date_{it-1} + \gamma_3\, Injury_{it}\,'\, Date_{it} + \boldsymbol{\gamma_4 X'_{it}} + v_{it} \quad (1)$$

where w_{it} is an indicator for individual i reaching the coinsurance arm of their plan in year t. An individual could reach the coinsurance arm of their plan by either meeting their individual deductible or their family deductible. The year is indexed with $t-1$ for 2010 measures, t for 2011 measures, and $t+1$ for 2012 measures. The excluded instrument, $Injury_{it}$, is an indicator equaling one if the injury was first observed in year t (early 2011) and zero if in year $t-1$ (late 2010). The running variable is represented by $Date_{it-1}$ and $Date_{it}$ which range from 1-90 in years $t-1$ and t, respectively. The interactions $Injury_{it-1}\,'\, Date_{it-1}$ and $Injury_{it}\,'\, Date_{it}$ between the running variable and what year the injury occurs allow the two sides of the discontinuity to have separate slopes.

The vector $\boldsymbol{X'_{it}}$ contains age, sex, individual deductible amount, and number of family members observed on the plan in ranges. Although there appear to be balance across the discontinuity in these observables (Figure 2), I include the deductible amount and number of family members because they are mechanical factors in predicting the likelihood of an individual reaching the coinsurance arm of the plan. I also show robustness to the exclusion of these covariates in Table 7.

Figure 1 shows a visual of the first-stage identification. Visually, the discontinuity is clear with little overlap across the year change in the probability of meeting the deductible. The first stage point estimate of γ_1 of 0.059 represents a roughly six percentage point increase at the discontinuity in the probability of an individual meeting their deductible in year t. The first stage meets all recent standards for power with a first stage F-statistic of 237.0.

To further support the validity of the research design, Figure 2 presents visualizations of the balance of covariates including the 2011 deductible, number of family members observed on the plan, birth year, and sex. They all show no major patterns or discontinuity at the cutoff supporting the validity of the design. Figure 3 shows the density of observations across the first observed service date for injuries. There are clear weekly patterns in the frequencies and some variations that can be attributed to holidays. Seasonal variation in these injuries is also plausible since many of the injuries could be connected to risky behaviors. However, the density plot still appears relatively smooth across the discontinuity.

4.2 Second Stage

The second stage is as follows:

$$y_{it+1} = \beta_0 + \beta_1 w_{it} + \beta_2 Injury_{it-1} \,\acute{}\, Date_{it-1} + \beta_3\, Injury_{it} \,\acute{}\, Date_{it} + \boldsymbol{\beta_4 X'_{it}} + \varepsilon_{it} \quad (2)$$

where y_{it+1}, is a healthcare consumption outcome in the year $t+1$. The second stage contains the same date-injury year interactions and covariates as the first stage. Here w_{it}, the indicator for reaching the coinsurance arm in year t is instrumented for by $Injury_{it}$, which is an indicator for having an injury in year t versus $t-1$. The estimates of β_1 are therefore identified by the discontinuities in the probability of reaching the coinsurance arm in year t at the year change.

5. Results and Discussion

5.1 2012 Outcomes

Table 5 presents estimates of β_1 based on Eqs. (1) and (2). Because injury date, $Date_{it}$, is discrete, all models report standard errors clustered by injury date following the inference procedure proposed by Lee and Card (2008). The first row has outcomes in the form of dollars spent or count of care dates, the second and third row presents the natural log and inverse hyperbolic sine of spending outcomes, and the fourth row presents outcomes as an indicator for the type of care occurring during the year.

In the first column, the outcome is total spending on healthcare in year $t+1$ and the point estimate of -13,263 implies that an individual meeting their deductible in the prior year is associated with a $13,263 decrease in total spending. This effect is large relative to the sample mean of $11,631 for total spending. In the second column, the point estimate for total out-of-pocket spending is -788. This implies that an individual meeting their deductible in the prior year is associated with a $788 decrease in out-of-pocket spending. The magnitude of this estimate is still large relative to the mean of $1,468, but not proportional to the total spending estimates because of the non-linear plan structure and out-of-pocket maximums.

To remove variations in costs as a factor, I examine total care dates in the third column. I estimate that meeting the deductible in the year prior leads to a decrease of 7.4 care dates relative to a mean of 16.2 care dates. The effect appears to be driven by decreases in both outpatient and inpatient care dates. When examining a class of elective care dates, the point estimates imply a 12.3 percentage point decrease in the probability of consuming any elective care in the year after the deductible is met, relative to a mean of 31 percent. In my main specification, I fail to detect any changes in preventive care dates. This result is suggestive that insurance companies and public policy efforts to exempt preventive care from consumer cost-sharing may successfully prevent decreases compared to other types of care.

To support the estimates of Table 5 visually, Figures 4 and 5 show the reduced-form relationship between the first date of service for the injury and outcomes. Despite being only the reduced form, the discontinuity in the expected direction is visually apparent.

5.2 2011 Outcomes

I use outcomes in the year the deductible is met to (1) verify that there is an observed increase in care in the year the deductible is met, and (2) estimate the tradeoff between spending across years. The first three rows of estimates in Table 6 contain the identifying injury for those with an injury in year t and potentially follow-up care for all individuals. The fourth and fifth rows decompose total spending based on whether an injury code is present on the claim. Overall, the point estimate for total spending is \$36,706 with the estimate for injury spending in year t being \$12,818 and for non-injury spending being \$23,887. Relative to the mean of total spending in year t of \$13,215, these estimates are large and economically significant increases in consumption.

Combining the year t overall estimates with those from year $t + 1$ implies that for those induced to consume more healthcare by meeting their deductible in one year, for every dollar of healthcare consumed in the year the deductible is met they consume \$0.37 less in the following year (\$13,263/\$36,706). Further, if we are interested in only the tradeoff between discretionary spending beyond that related to the injury, my estimates imply that for those induced to consume more healthcare by meeting their deductible in one year, for every dollar of elective care consumed in the year the deductible is met, they consume $0.56 less in the following year ($13,263/$23,887). From the consumer's perspective, the out-of-pocket estimates imply that for those induced to consume more healthcare by meeting their deductible in one year, for every dollar spent out of pocket in the year the deductible is met $0.24 less is spent in the following year ($788/$3,288).

While it is difficult to categorize what care is directly related to the injury as diagnosis and billing codes may vary, measurement error from failing to match all injury-related claims to the injury would cause an overestimate of discretionary (non-injury related) spending. Additionally, if those with injuries in year $t-1$ are participating in intertemporal substitution to any degree, the year t spending estimates will overestimate the true effect. However, the year t spending estimates serve as the denominator of the tradeoff, so potential overestimates lead to a conservative estimate and make my estimates represent a lower bound of the true tradeoff.

Examining elective and preventive care dates, which are unrelated to the injury, provides insights into where the increases in discretionary spending are occurring. The point estimate for elective care dates of 1.96 relative to a mean in year t of 0.65 suggests that these elective procedures are a major channel where this spending occurs. Interestingly, there is a 37.1 percentage point increase in the probability of consuming any preventive care relative to a mean of 59.4 percent. This contrasts with failing to detect any significant effect on preventive care in year $t+1$. A possible explanation is that increased interaction with the healthcare system increases preventive care usage, but cost-sharing exemptions are effective at minimizing decreases in preventive care due to non-linear plans.

5.3 Back-of-the-Envelope Calculation of Economic Impact

To understand the scope of this intertemporal substitution, I conduct a back-of-the-envelope calculation of the cost savings to the privately insured in the U.S. from intertemporally substituting when consuming the average out-of-pocket cost using 2020 insurance rates. I calculate

$$150{,}000{,}000 \ * \ 0.57 \ * \ 0.83 * \ 0.06 \ * \ 0.24 \ * \$1{,}468 \ = \ \$1{,}500{,}143{,}328 \qquad (3)$$

where there are roughly 150 million employed workers in the US, 57% of workers are covered by employer-provided insurance, 83% then have a deductible, roughly 6% of the population are compliers, the average out-of-pocket is $1,468, and the out-of-pocket spending tradeoff found is $0.24 less spent per dollar spent in the previous year (KFF, 2020).[4,5] I find that U.S. consumers are saving at least $1.5 billion per year through intertemporal substitution. While a strategic subpopulation is benefiting from these savings, all individuals with employer-provided health insurance likely bear the cost through slightly higher premiums (or lower wages).

Estimates of meeting the deductible from instrumental variables strategies using within year variation and a single year of data are unable to capture that consumers are strategically intertemporally substituting across years to consume healthcare at lower prices. Because intertemporal substitution is not accounted for in estimates of the cost-saving benefits of high-deductible plans extrapolating from their estimates will overstate savings for insurers and employers. Similarly, I estimate what savings health insurance companies are not receiving that certain single year estimates would suggest. I calculate

$$150{,}000{,}000 \ * \ 0.57 \ * \ 0.83 \ * \ 0.37 \ * \$11{,}631 \ * \ 0.06 = \$18{,}323{,}744{,}913 \qquad (4)$$

[4] While there should be standard errors on all the terms of the calculation, some are unpublished, so I am unable to provide a confidence interval.
[5] These back-of-the-envelope calculations require the assumption that those with an accidental injury in my sample are representative of a broader population of those with employer-provided health insurance with a deductible.

where there are roughly 150 million employed workers in the US, 57% of workers are covered by employer-provided insurance, 83% then have a deductible, roughly 6% of the population are compliers, , the average of total spending is $11,631, and the total spending tradeoff found is $0.37 less spent per dollar consumed in the previous year (KFF, 2020).[4, 5] Thus, I find that estimates using only a single year would overestimate the savings of high-deductible plans by $18.3 billion.

Further, if intertemporal substitution extends beyond the year after the deductible is met, all of these estimates will underestimate the true effect. It seems plausible that intertemporal substitution would extend beyond a single year as some types of care could be strategically timed across multiple years and some individuals may expect to meet their deductible only every few years Additionally, the size of deductibles has continued to grow over time which would likely lead to both increased rates and magnitudes of intertemporal substitution. Thus, these results are a conservative estimate of the economic impact of intertemporal substitution.

7. Sensitivity, Placebo, and Heterogeneity Analyses

7.1 Placebo Tests and Heterogeneity by Deductible Amount

In Figure 6, I show reduced form plots of 2011 and 2012 total spending for the main sample and then for those that are excluded from the sample because they have no deductible. Those with no deductible are a placebo test in the sense that we would expect to see a mechanical difference in spending 2011 because they experience the same injuries. However, we would expect to see little to no change in spending in 2012 because the group has no deductible and has no incentives to intertemporally substitute if they did not reach their out-of-pocket maximum in 2011, which would be unlikely for the average injury in the sample. The 2011 total spending plots look similar across groups, while the 2012 total spending plots show different patterns as expected. For those with a deductible greater than $100 there is a noticeable decrease in total spending at the discontinuity as expected, while for those without a deductible, there is an increase in 2012 total spending at the discontinuity of a much smaller magnitude. The fact that the placebo test of those without a deductible is of the opposite sign shows that the identification is coming solely from those with deductibles and is convincing evidence in support of the identification strategy's validity.

In Table 7, I conduct a heterogeneity analysis stratifying by deductible amount. But the first group of those with a deductible that is less than or equal to $100 serve as a placebo test because a deductible that small is much less likely to induce these behavioral responses across years. I find no significant results for this group and the instrument is weak with a first-stage F-statistic of 0.17.

The second through fifth columns of Table 7 presents the results from the main sample stratified by deductible amount. I would predict that individuals with mid-sized deductibles (i.e., $300-$1000) are the most likely to only meet their deductible in some years leading to strong incentives to substitute care across years. Those with smaller deductibles (i.e., <$300) are more likely to reach the coinsurance arm of their plan consistently. Similarly, those with much larger deductibles (i.e., ≥$1000) are less likely to meet their deductible often or from the injury. The first-stage coefficient is largest for mid-range deductibles, and smaller for both smaller and larger deductibles aligning with these predictions. Further, point estimates are largest for deductibles of $300-999 as these individuals are likely the most likely to only meet their deductible in some years leading to a substituting care across years. Sample size limitations make it difficult to comment on deductibles greater than $1000.

7.2 Sensitivity/Robustness

In my main specifications, I include covariates to address concerns about observables which may impact the likelihood of an individual meeting their deductible. I show robustness to this choice in Table 8 where the covariates in the vector $\boldsymbol{X'_{it}}$ (age, sex, individual deductible amount, and number of family members observed on the plan in ranges) are excluded. Overall, results appear quite similar with estimates having slightly larger magnitudes making the main specification a conservative estimate. Specifically, the estimates without covariates for total and out-of-pocket spending of -15,085 and -925 are very close to the corresponding estimates with covariates of -13,263 and -788.

Due to lack of plan information, I assume that the family deductible is twice the individual deductible meaning it is only relevant for families with more than two individuals enrolled. To show robustness to this assumption, I show results stratified by the number of observed family members enrolled in Table 9. Point estimates are similar and of a larger magnitude for those with one or two family members observed on the plan. Since the family deductible is not relevant for these individuals, assumptions about the family deductible are not driving results.

To address concerns about individuals strategically shifting treatment around the year change, I run two donut regression discontinuities with bandwidths of 7 and 14 days (Table 10). These specifications also address potential concerns about the uniqueness of injuries during the winter holiday season. These estimates are quite similar to the main specification with a slightly larger magnitude for the point estimates. Similarly, Figure 7 shows variations in bandwidth of 30-180 days. Point estimates are similar for bandwidths of roughly 60-180 days. For smaller bandwidths standard errors increase and the proportion of individuals with 2011 injuries who may have injury care in 2012 increases.

8. Conclusion

In this work, I show that intertemporal substitution across years in response to non-linear health insurance plans exists in the modern U.S. healthcare context. This finding is especially important as high-deductible health insurance plans have become increasingly common, and many estimates of the cost savings of high deductibles come from sources that do not capture intertemporal substitution.

Using claims data following privately insured individuals over three years, I find that every dollar of discretionary healthcare consumed in the year the coinsurance arm is reached, roughly $0.56 less is consumed in the following year for those induced to consume more care by meeting their deductible. The local average treatment effects indicate that reaching the coinsurance arm in one year leads to $13,263 less healthcare consumed and $788 fewer paid out of pocket in the following year.

These results align with the conclusions of Lin and Sacks (2019), based on a simulation using the RAND Health Insurance Experiment, that failing to account for intertemporal substitution could cause estimates to overstate cost savings from high deductible plans by more than 20 percent. While my fuzzy regression discontinuity produces a local average treatment effect with strong internal validity, it does not immediately translate to total cost estimates for the entire population. To understand the scope of this intertemporal substitution, I conduct a back-of-the-envelope calculation of and find that U.S. consumers are saving at least $1.5 billion per year through intertemporal substitution. Further, single year estimates would overstate the national savings from high-deductible plans by at least $18.3 billion emphasizing the importance of the results that ignoring across-year intertemporal substitution would cause many previous estimates using a single year to overstate the cost-saving benefits of high-deductible plans.

References

Alpert, A. (2016). The anticipatory effects of Medicare Part D on drug utilization. Journal of health economics, 49, 28-45.

Aron-Dine A, Einav L, Finkelstein A, Cullen M. Moral Hazard In Health Insurance: Do Dynamic Incentives Matter? Rev Econ Stat. 2015 Oct;97(4):725-741

Brot-Goldberg Z, Chandra A, Handel B, Kolstad J. What does a Deductible Do? The Impact of Cost-Sharing on Health Care Prices, Quantities, and Spending Dynamics. The Quarterly Journal of Economics. 2017;132(3 1):1261-1318.

Cabral, M. (2016). Claim Timing and Ex Post Adverse Selection. *The Review of Economic Studies, 84*(1), 1-44.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014a). Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs. Econometrica, 82(6), 2295–2326.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2014b). Robust Data-Driven Inference in the Regression-Discontinuity Design. The Stata Journal, 14(4), 909–946.

Calonico, S., Cattaneo, M. D., & Titiunik, R. (2015). Optimal Data-Driven Regression Discontinuity Plots. Journal of the American Statistical Association, 110(512), 1753–1769.

U.S. Centers for Medicare and Medicaid Services. Preventive & Screening Services, https://www.medicare.gov/coverage/preventive-screening-services.

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

Dalton, C. M., Gowrisankaran, G., & Town, R. J. (2019). Salience, Myopia, and Complex Dynamic Incentives: Evidence from Medicare Part D. *The Review of Economic Studies*.

Einav L, Finkelstein A, Schrimpf P. The Response Of Drug Expenditure To Non-Linear Contract Design: Evidence From Medicare Part D. Q J Econ. 2015 May;130(2):841-899.

Einav L, Finkelstein A. Moral Hazard in Health Insurance: What We Know and How We Know It. J Eur Econ Assoc. 2018 Aug;16(4):957-982.

Guo A, Zhang J. What to expect when you are expecting: Are health care consumers forward-looking? J Health Econ. 2019 Sep; 67:102216.

Imbens, G. W., & Lemieux, T. (2008). Regression discontinuity designs: a guide to practice. Journal of Econometrics, 142(2), 615.

Johansson, N., de New, S. C., Kunz, J. S., Petrie, D., & Svensson, M. (2023). Reductions in out-of-pocket prices and forward-looking moral hazard in health care demand. Journal of health economics, 87, 102710.

Kaiser Family Foundation, Health Research and Educational Trust, 2020. Employer Health Benefits, 2020 Annual Survey.

Klein, T. J., Salm, M., & Upadhyay, S. (2022). The response to dynamic incentives in insurance contracts with a deductible: Evidence from a differences-in-regression-discontinuities design. Journal of Public Economics, 210.

Kowalski A. Censored Quantile Instrumental Variable Estimates of the Price Elasticity of Expenditure on Medical Care. J Bus Econ Stat. 2016 Jan 2;34(1):107-117.

Lee, D. S., & Card, D. (2008). Regression discontinuity inference with specification error. Journal of Econometrics, 142(2), 655–674.

Lin, H., & Sacks, D. W. (2019). Intertemporal substitution in health care demand: Evidence from the RAND Health Insurance Experiment. *Journal of Public Economics, 175*, 29-43.

Tables and Figures

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

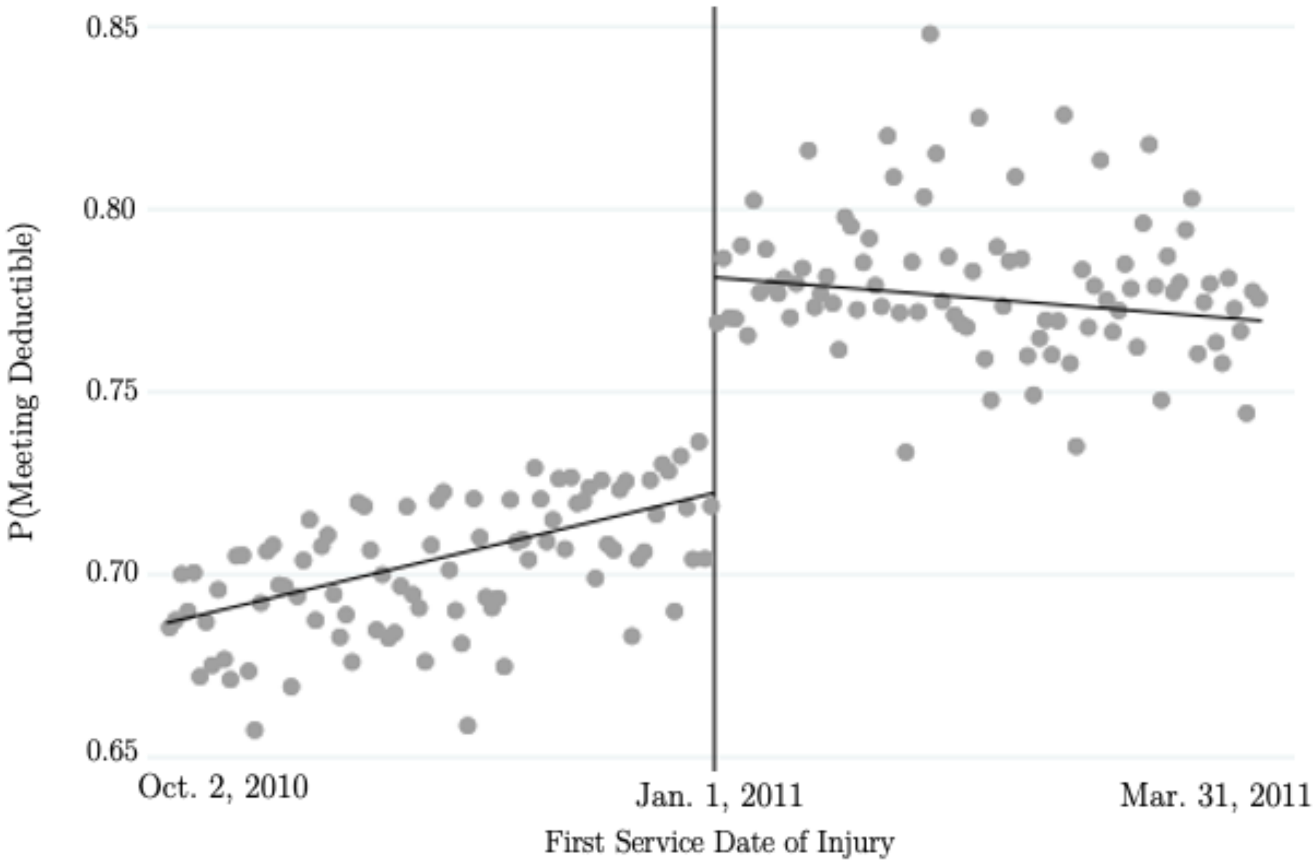


Figure 2. Balance Tests

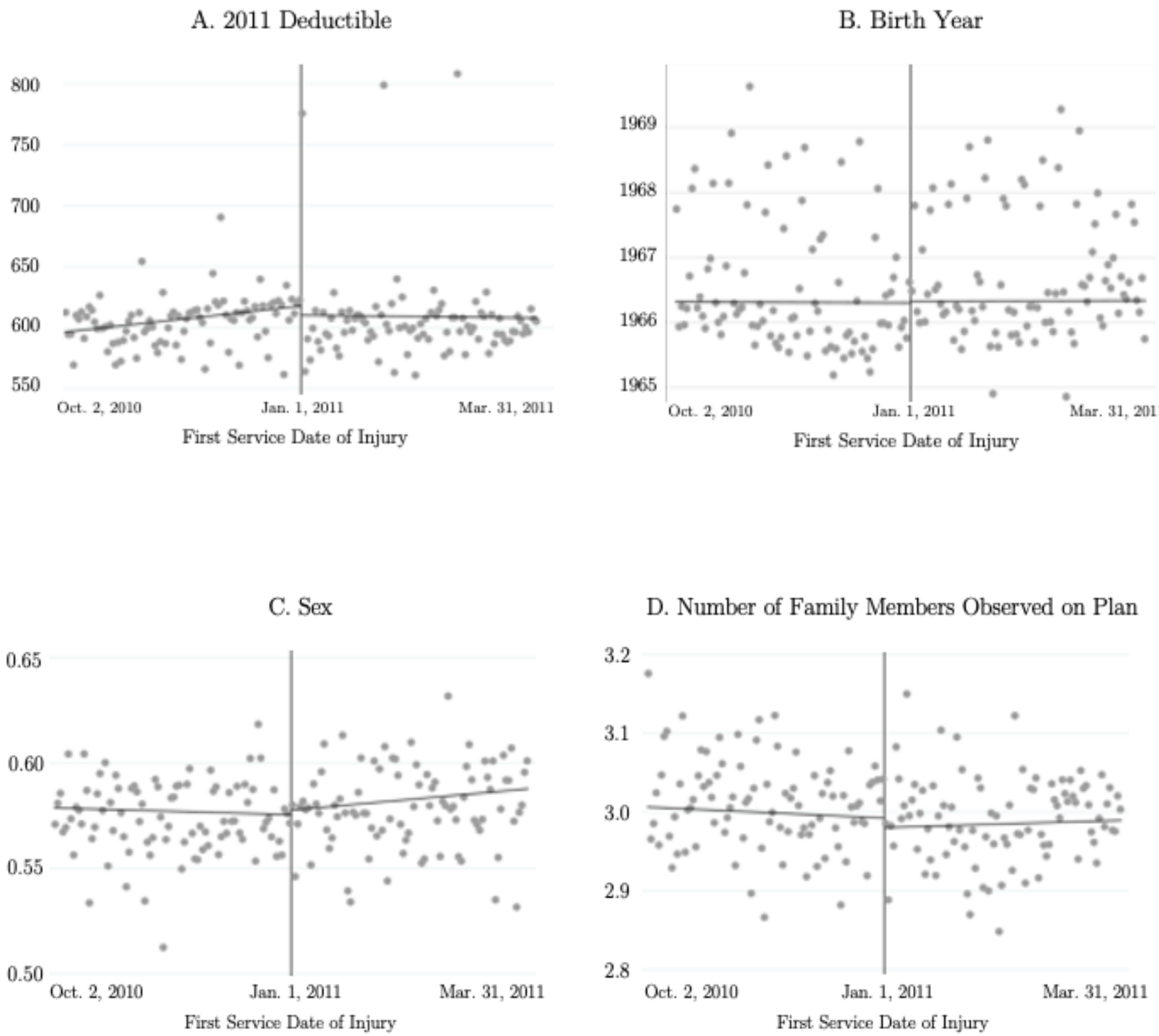


Figure 3. Number of Individuals Observed by First Service Date of Injury

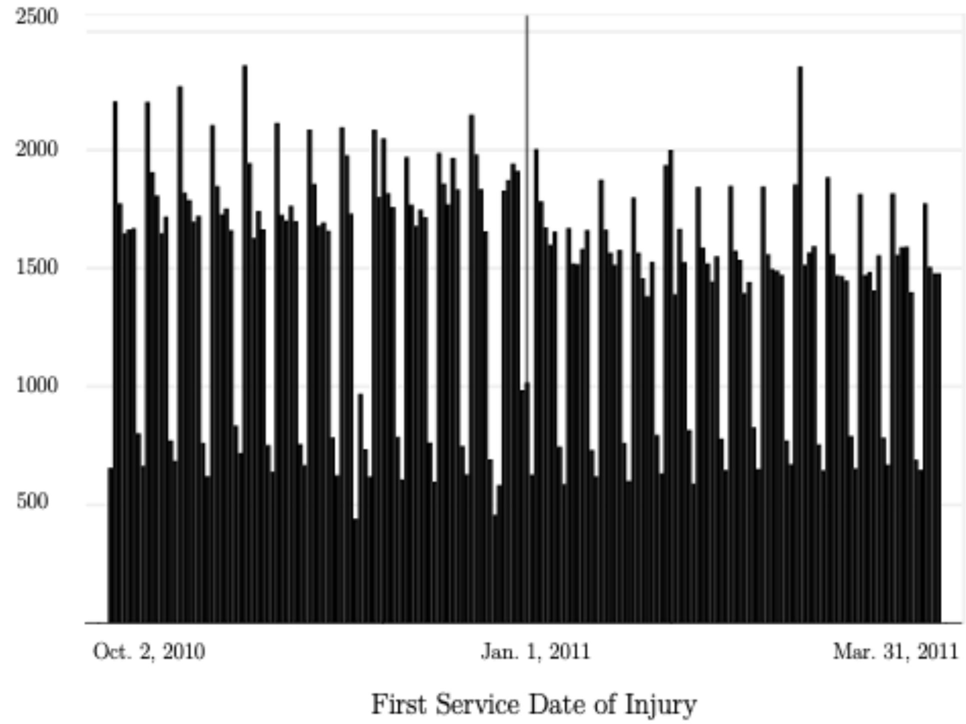
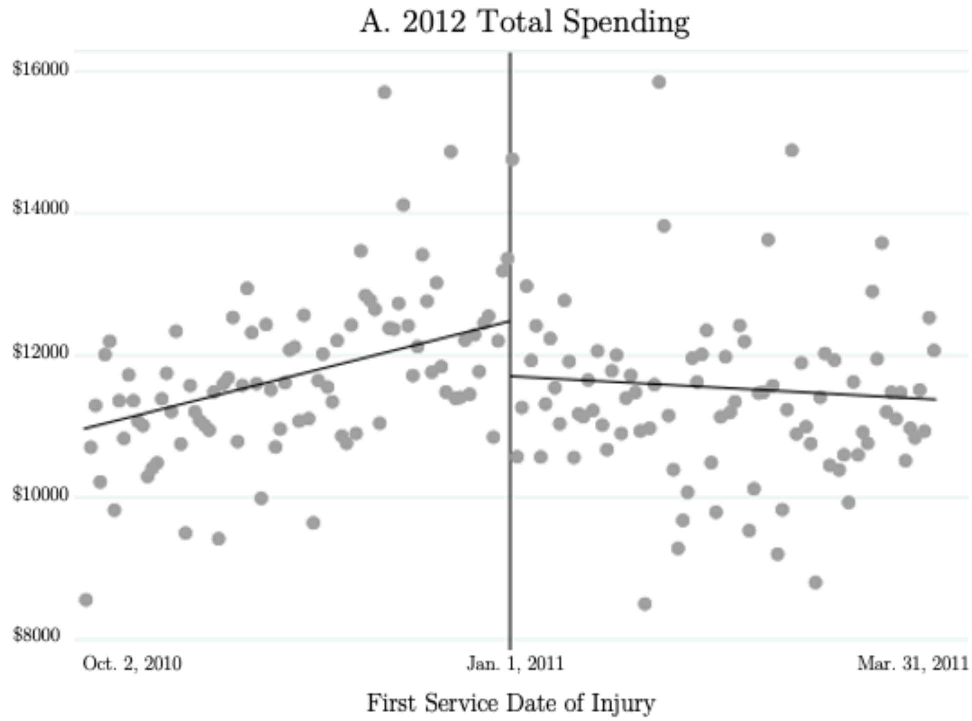


Figure 4. Reduced Form Plots for Spending Outcomes



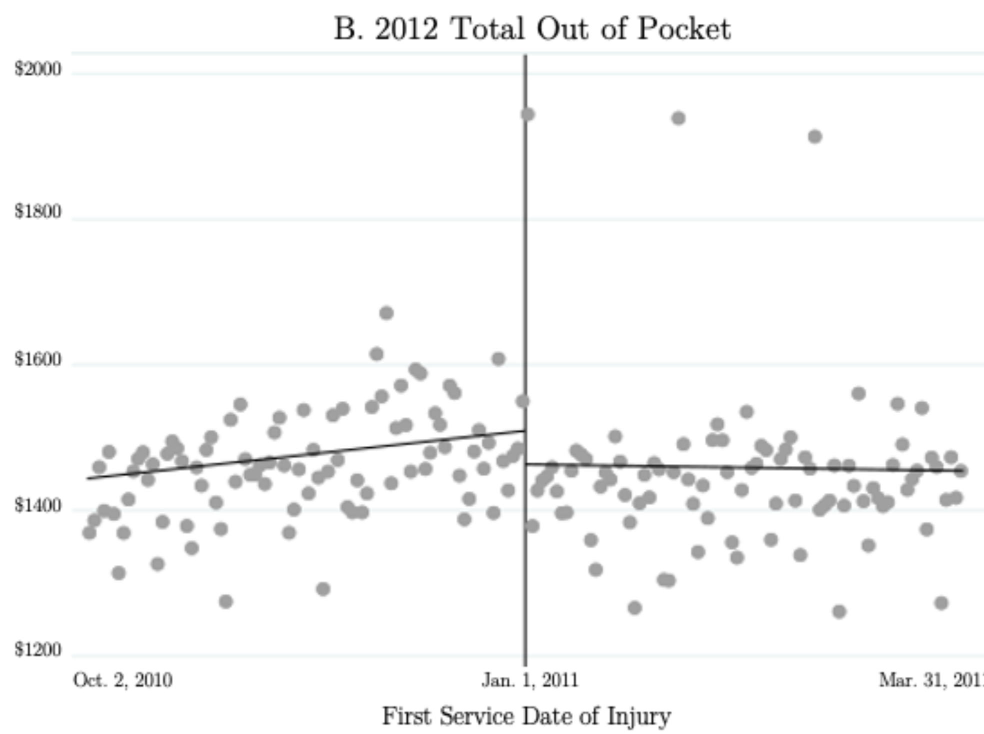


Figure 5. Reduced Form Plots for Care Date Outcomes

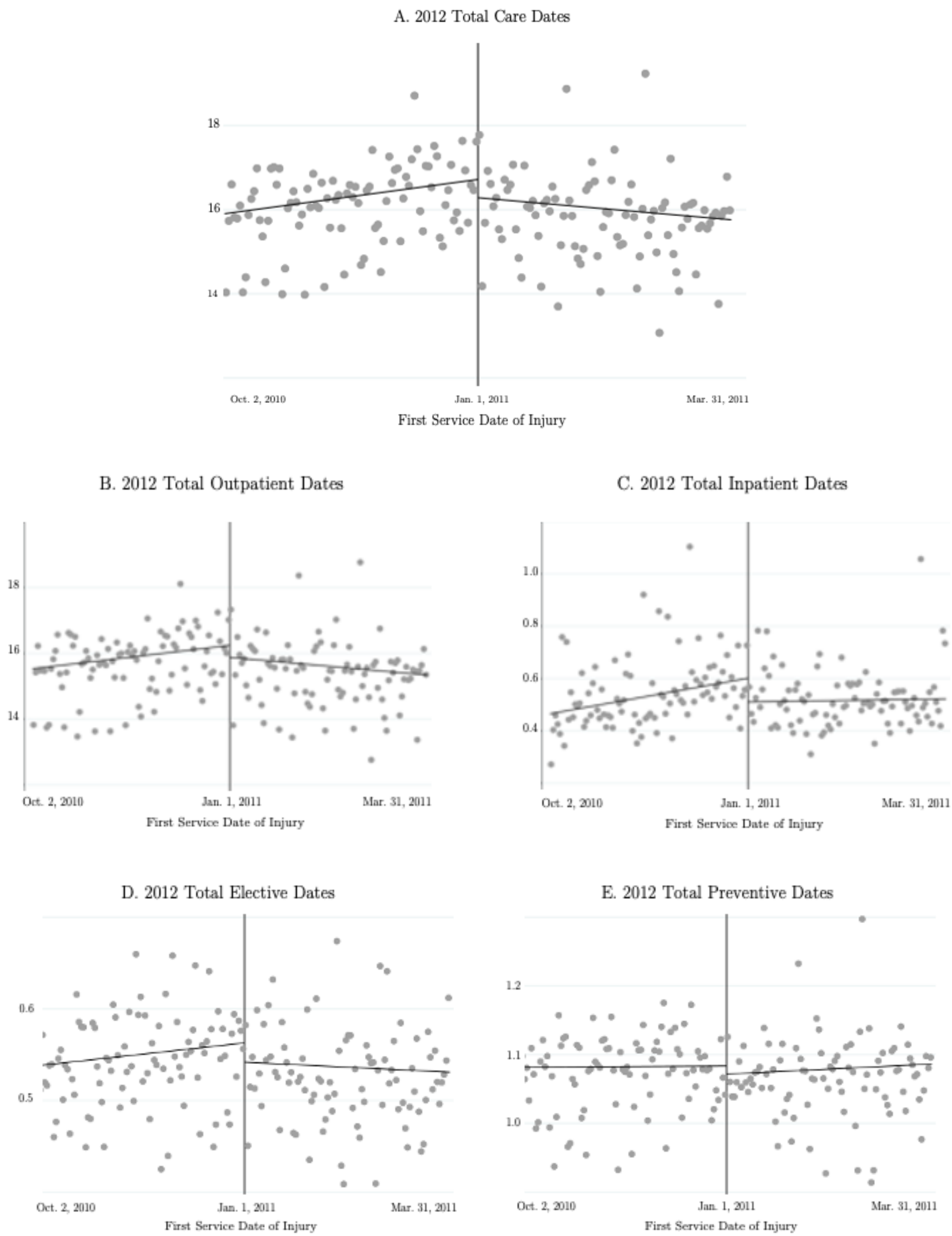


Figure 6. Reduced-Form Plot Placebo Test

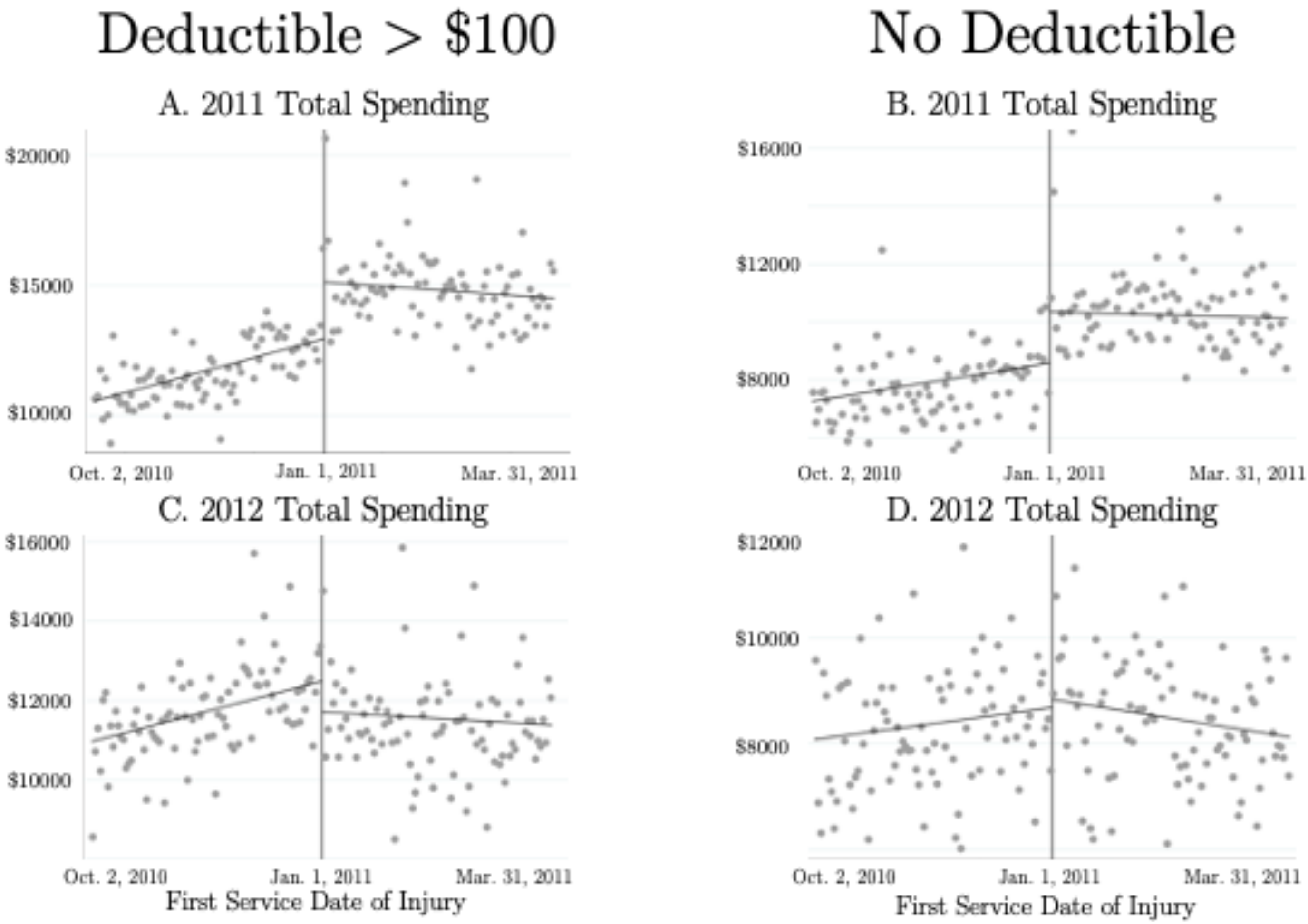


Figure 7. Estimates for 30-180 Day Bandwidths

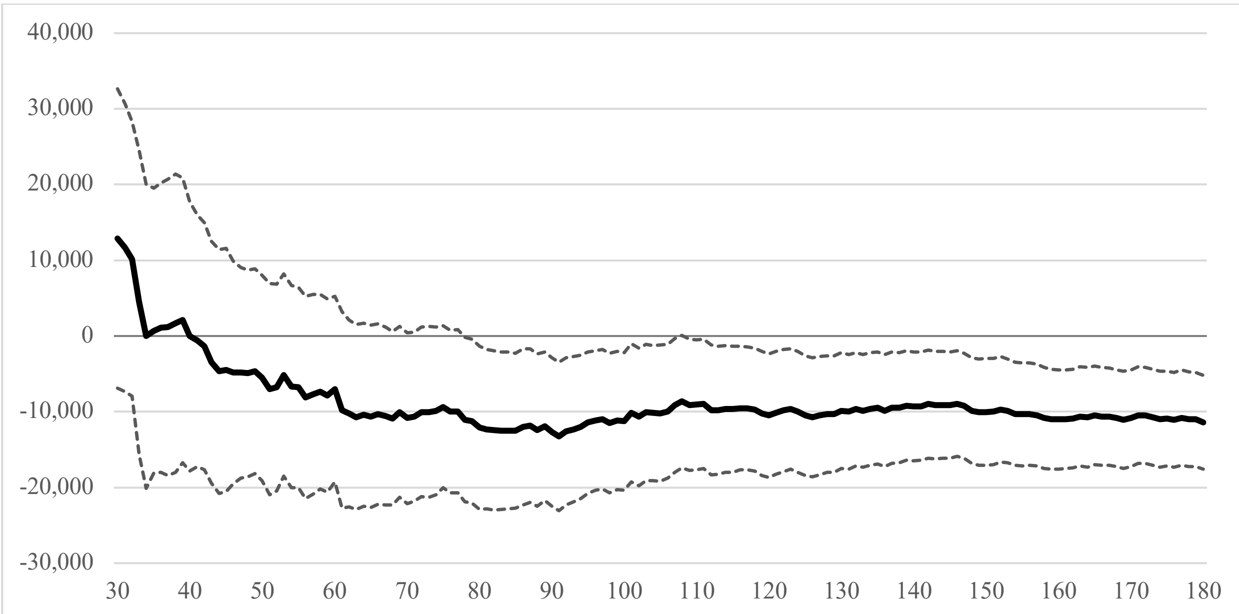


Table 1. Identifying Injuries in Sample, Overall and By Year of Injury

Injuries from Kowalski (2016)	ICD-9	Total Injuries (Percent)	Late 2010 Injuries (Percent)	Early 2011 Injuries (Percent)
Entire Sample		254,902 (100)	133,069 (100)	121,833 (100)
Fractures	800-829	23,847 (9.4)	12,187 (9.2)	11,660 (9.6)
Dislocation	830-839	24,742 (9.7)	13,509 (10.2)	11,233 (9.2)
Sprains and Strains of Joints and Adjacent Muscles	840-849	87,431 (34.3)	44,608 (33.5)	42,823 (35.1)
Intracranial Injuries, Excluding Skull Fractures	850-859	4,334 (1.7)	2,099 (1.6)	2,235 (1.8)
Internal Injury of Thorax, Abdomen, and Pelvis	860-869	0 (0.0)	0 (0.0)	0 (0.0)
Open Wounds	870-899	25,453 (10.0)	13,749 (10.3)	11,704 (9.6)
Injury to Blood Vessels	900-904	267 (0.1)	131 (0.1)	136 (0.1)
Late Effects of Injuries, Poisonings, Toxic Effects, and Other External	905-909	1,018 (0.4)	609 (0.5)	409 (0.3)
Superficial Injuries	910-919	13,535 (5.3)	7,748 (5.8)	5,787 (4.7)
Contusion with Intact Skin Surface	920-924	23,908 (9.4)	12,256 (9.2)	11,652 (9.6)
Crushing Injuries	925-929	675 (0.3)	383 (0.3)	292 (0.2)
Foreign Body Injuries	930-939	5,031 (2.0)	2,685 (2.0)	2,346 (1.9)
Burns	940-949	2,346 (0.9)	1,268 (1.0)	1,078 (0.9)
Injuries to Nerves and Spinal Cord	950-957	1,165 (0.5)	651 (0.5)	514 (0.4)
Complications of Trauma	958-959	32,188 (12.6)	15,882 (11.9)	16,306 (13.4)
Poisoning by Drugs, Medicinal and Biological Substances	960-979	2,254 (0.9)	1,117 (0.8)	1,137 (0.9)
Toxic Effects of Substances Chiefly Nonmedicinal and Other External	980-995	17,729 (7.0)	9,177 (6.9)	8,552 (7.0)
Complications of Surgical and Medical Care, Not Elsewhere Classified	996-999	16,789 (6.6)	8,469 (6.4)	8,320 (6.8)

Table 2. Covariate Summary Statistics, Overall and By Year of Injury

	Overall	2010 Injury	2011 Injury		Overall	2010 Injury	2011 Injury
2011 Deductible (in USD)				Sex			
Mean	607.9	606.8	609.1	Male	42.0	42.3	41.7
101-199	6.9	6.7	7.0	Female	58.0	57.7	58.3
200-299	18.6	18.8	18.3	Age (in 2010)			
300-399	20.0	20.2	19.9	18-34	24.7	24.4	25.0
400-499	6.6	6.5	6.7	35-44	24.6	24.5	24.8
500-749	22.0	22.1	21.8	45-54	33.0	33.3	32.6
750-999	5.3	5.2	5.4	55-64	17.7	17.9	17.6
1000-1249	9.7	9.5	9.9	Number of Family Members Enrolled			
1250-1499	3.3	3.1	3.5	1	20.0	19.5	20.6
1500-1749	4.3	4.3	4.2	2	23.6	24.0	23.3
1750-1999	0.4	0.5	0.4	3	18.1	18.2	18.0
2000-2499	1.2	1.2	1.1	4	22.7	22.8	22.7
2500-2999	0.6	0.6	0.6	5-6	13.4	13.4	13.4
3000-4999	1.1	1.2	1.1	7-8	1.7	1.7	1.7
5000-10000	0.0	0.0	0.0	≥ 9	0.4	0.4	0.4

Table 3. Summary Statistics of Year $t + 1$ (2012) Outcomes

2012 Outcomes	Mean	SD	P1	P25	P50	P75	P99
Total Spending	11631	32495	52	1368	3867	10403	125763
Out of Pocket	1468	1900	0	383	958	2038	7121
Care Dates	16.17	19.43	0	5	10	21	88
Outpatient	15.75	18.39	0	5	10	20	83
Inpatient	0.53	3.64	0	0	0	0	11
Elective	0.54	1.47	0	0	0	1	5
Preventive	1.08	1.31	0	0	1	2	5
1(Inpatient>0)	0.09	0.28	0	0	0	0	1
1(Elective>0)	0.31	0.46	0	0	0	1	1
1(Preventive>0)	0.60	0.49	0	0	1	1	1

Table 4. Means of Outcomes Overall and by Year of Injury

2011 Outcomes (y_{it})	Overall	2010 Injury	2011 Injury	2012 Outcomes (y_{it+1})	Overall	2010 Injury	2011 Injury
Probability of Meeting Deductible	73.83	70.42	77.55				
Total Spending	13,215	11,761	14,804	Total Spending	11,631	11,726	11,526
Out of Pocket	1,583	1,461	1,717	Out of Pocket	1,468	1,474	1,461
Care Dates	19.08	18.02	20.23	Care Dates	16.17	16.32	16.01
Outpatient	18.52	17.57	19.57	Outpatient	15.75	15.89	15.60
Inpatient	0.69	0.56	0.83	Inpatient	0.52	0.53	0.52
Elective	0.65	0.59	0.72	Elective	0.54	0.55	0.53
Preventive	1.08	1.07	1.08	Preventive	1.08	1.08	1.08
1(Inpatient>0)	0.110	0.096	0.126	**1**(Inpatient>0)	0.087	0.087	0.087
1(Elective>0)	0.375	0.337	0.416	**1**(Elective>0)	0.309	0.313	0.305
1(Preventive>0)	0.594	0.594	0.595	**1**(Preventive>0)	0.603	0.603	0.604

Table 5. 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)					
$arsinh(y_{it+1})$	-0.504*	-0.577**	-	-	-	-	-
	(0.277)	(0.277)					
$\mathbf{1}(y_{it+1} > 0)$	-	-	-	-	-0.0863**	-0.123*	-0.0667
					(0.042)	(0.064)	(0.059)
N	254,902	254,902	254,902	254,902	254,902	254,902	254,902
$Mean\ of\ y_{it+1}$	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

Note: Each coefficient is estimated from a single regression and represents β_1 in Eq. (2). All standard errors are clustered at the first service date of the injury. *p<0.10, **p<0.05, ***p<0.10.

Table 6. Effect of Meeting Deductible in Year t on Year t 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}	36,706***	3,288***	23.09***	21.46***	2.159***	1.962***	0.122
	(5,938)	(263.5)	(3.838)	(3.654)	(0.686)	(0.277)	(0.216)
$\ln(1 + y_{it})$	4.896***	4.423***	-	-	-	-	-
	(0.344)	(0.212)					
$arsinh(y_{it})$	4.905***	4.502***	-	-	-	-	-
	(0.345)	(0.216)					
$\mathbf{1}(y_{it} > 0)$	-	-	-	-	0.486***	1.544***	0.371***
					(0.0372)	(0.104)	(0.0427)
y_{it} on $Injury$	12,818**	-	-	-	-	-	-
	(5,297)						
y_{it} Not on $Injury$	23,887***	-	-	-	-	-	-
	(1,918)						
N	254,902	254,902	254,902	254,902	254,902	254,902	254,902
$Mean\ of\ y_{it}$	13,215	1,583	19.08	18.52	0.691	0.651	1.075
$SD\ of\ y_{it}$	32,861	1900	20.4	19.12	4.187	1.450	1.343

Note: Each coefficient is estimated from a single regression and represents β_1 in Eq. (2) with year $\boldsymbol{t}$ (2011) outcomes. All standard errors are clustered at the first service date of the injury. *p<0.10, **p<0.05, ***p<0.10.

Table 7. Effect of Meeting Deductible in Year t on Year $t+1$ Outcomes

	Deductible (d)				
2012 $Outcome\ (y_{it+1})$	$0 < d \leq 100$	$100 < d < 300$	$300 \leq d < 500$	$500 \leq d < 1000$	$1000 \leq d < 10000$
Total Spending	-342,356	-7,004	-18,791**	-15,106**	-9,222
	(878,255)	(9,629)	(8,970)	(7,498)	(15,000)
Total Out of Pocket	-18,383	-644.3	-294.9	-1,788***	148.8
	(47,906)	(465.1)	(357.1)	(596.0)	(1,195)
Total Care Dates	-76.56	-1.561	-6.719	-14.19***	-2.410
	(234.5)	(6.255)	(4.870)	(5.240)	(8.845)
Outpatient Care Dates	-26.37	-0.605	-5.138	-12.75***	-1.919
	(130.7)	(5.971)	(4.575)	(4.854)	(8.252)
Inpatient Care Dates	-61.27	-1.248	-1.890**	-1.887**	-0.839
	(150.8)	(1.268)	(0.889)	(0.896)	(1.488)
Elective Care Dates	-0.0106	0.0699	-0.190	-0.579**	-0.665
	(11.12)	(0.485)	(0.383)	(0.263)	(0.509)
Preventive Care Dates	-8.509	0.164	-0.249	-0.389	-0.132
	(22.02)	(0.329)	(0.364)	(0.260)	(0.0987)
Sample Size	27,538	64,822	67,919	69,492	52,669
First-stage coefficient	0.004	0.048	0.058	0.073	0.054
First-stage F-statistics	0.17	43.33	93.48	91.48	50.69

Note: Each coefficient is estimated from a single regression and represents β_1 in Eq. (2) stratified by deductible amount. All standard errors are clustered at the first service date of the injury. *p<0.10, **p<0.05, ***p<0.10.

Table 8. **Effect of Meeting Deductible in Year t on Year $t+1$ Outcomes without Covariates**

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}	-15,085***	-925.2**	-8.954**	-3.822	-1.594***	-0.486**	-0.356*
	(5,133)	(463.9)	(3.822)	(3.640)	(0.578)	(0.203)	(0.209)
$\ln(1+y_{it+1})$	-0.712**	-0.709**					
	(0.322)	(0.309)					
$1(y_{it+1}>0)$					-0.0880**	-0.185**	-0.130*
					(0.0418)	(0.0718)	(0.0728)
$Mean\ of\ y_{it+1}$	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

Note: Each coefficient is estimated from a single regression and represents β_1 in Eq. (2) without covariates (deductible amount, number of family members observed on plan, age, sex). All standard errors are clustered at the first service date of the injury. *p<0.10, **p<0.05, ***p<0.10.

Table 9. Effect of Meeting Deductible in Year t on Year $t+1$ Outcomes by Number of Family Members Observed on Plan

	Number of Family Members Observed on Plan		
2012 $Outcome\ (y_{it+1})$	1-2	3-4	≥ 5
Total Spending	-18,762***	-5,424	-12,603
	(6,221)	(8,241)	(10,279)
Total Out of Pocket	-844.9*	-754.6	-721.6
	(434.5)	(571.1)	(564.7)
Total Care Dates	-11.23**	-4.094	-1.523
	(4.395)	(5.164)	(6.888)
Outpatient Care Dates	-9.909**	-3.713	1.100
	(4.212)	(4.827)	(6.370)
Inpatient Care Dates	-1.782**	-0.494	-3.209**
	(0.765)	(0.885)	(1.297)
Elective Care Dates	-0.676***	0.0273	-0.163
	(0.245)	(0.371)	(0.475)
Preventive Care Dates	-0.520**	0.165	0.0568
	(0.213)	(0.364)	(0.351)
Sample Size	111,247	104,121	39,534
First-stage coefficient	0.070	0.049	0.056
First-stage F-statistics	125.68	71.33	47.00

Note: Each coefficient is estimated from a single regression and represents β_1 in Eq. (2) stratified by number of family members observed on the plan. All standard errors are clustered at the first service date of the injury. *p<0.10, **p<0.05, ***p<0.10.

Table 10. Donut Hole Specification of Effect of Meeting Deductible in Year t on Year $t+1$ Outcomes

	Donut Hole Length in Each Year	
2012 $Outcome\ (y_{it+1})$	7 days	14 days
Total Spending	-17,300***	-22,092***
	(5,877)	(7,060)
Total Out of Pocket	-1,176***	-1,524***
	(404.2)	(471.8)
Total Care Dates	-8.973**	-14.38***
	(4.115)	(4.765)
Outpatient Care Dates	-7.386*	-12.44***
	(3.948)	(4.535)
Inpatient Care Dates	-2.051***	-2.400***
	(0.637)	(0.719)
Elective Care Dates	-0.351	-0.699***
	(0.238)	(0.268)
Preventive Care Dates	-0.297	-0.312
	(0.213)	(0.279)
Sample Size	234,965	215,995
First-stage coefficient	0.058	0.059
First-stage F-statistics	191.46	118.51

Note: Each coefficient is estimated from a single regression and represents β_1 in Eq. (2) with donut holes of 7 and 14 days. All standard errors are clustered at the first service date of the injury. *p<0.10, **p<0.05, ***p<0.10.

Appendix

- Count of Elective Care Dates in 2012 is a count of the number of elective service dates in 2012 with elective services being defined based on elective procedures defined using Berenson-Eggers Type of Service (BETOS) codes by Clemens and Gottlieb (2014) and also used by Guo and Zhang (2019).
    - List of BETOS codes used: P2A: Major procedure, cardiovascular – CABG, P2C: Major procedure, cardiovascular – thrombo-endarterectomy, P2D: Major procedure, cardiovascular – coronary angioplasty (PTCA) P3B: Major procedure, orthopedic – hip replacement, P3C: Major procedure, orthopedic – knee replacement, P4B: Eye procedure – cataract removal/lens insertion, P5A: Ambulatory procedures – skin • P5B: Ambulatory procedures – musculoskeletal, P6A: Minor procedures – skin, P6B: Minor procedures – musculoskeletal, P8A: Endoscopy – arthroscopy, P8B: Endoscopy – upper gastrointestinal, P8C: Endoscopy – sigmoidoscopy, P8D: Endoscopy – colonoscopy, P8E: Endoscopy – cystoscopy, P8F: Endoscopy – bronchoscopy, P8G: Endoscopy – laparoscopic cholecystectomy, P8H: Endoscopy – laryngoscopy, I4A: Imaging/procedure – heart including cardiac catheter
- Count of Preventive Care Dates in 2012 is a count of the number of preventative service dates in 2012 with preventative services being defined based on the Center of Medicare and Medicaid Services' list of preventive services.
    - Preventive Service Categories include: Alcohol Misuse Screening & Counseling, Annual Wellness Visit, Bone Mass Measurements, Cardiovascular Disease Screening Tests, Cervical Cancer Screening, Colorectal Cancer Screening, Counseling to Prevent Tobacco Use, Depression Screening, Diabetes Screening, Diabetes Self-Management Training, Flu Shot & Administration, Glaucoma Screening, Hepatitis B Screening, Hepatitis B Shot & Administration, Hepatitis C Screening, HIV screening, IBT for Cardiovascular Disease, IBT for Obesity, Initial Preventive Physical Exam, Ling Cancer Screening, STI Screening & HIBC to Prevent STIs, Screening Pelvic Exams, Ultrasound AAA Screening

Appendix Table A1. Effect of Meeting Deductible in Year t on Year $t+1$ Outcomes Sample Exclusions

	Group Included along with Main Sample	
2012 $Outcome\ (y_{it+1})$	Under 18	Deductible < 100
Total Spending	-9,888***	-15,270***
	(3,748)	(5,426)
Total Out of Pocket	-476.0*	-892.6**
	(261.5)	(399.3)
Total Care Dates	-4.74**	-7.79**
	(2.42)	(3.65)
Outpatient Care Dates	-3.79*	-6.28*
	(2.30)	(3.42)
Inpatient Care Dates	-1.21***	-1.94***
	(0.40)	(0.63)
Elective Care Dates	-0.27*	-0.36*
	(0.15)	(0.21)
Preventive Care Dates	-0.13	-0.25
	(0.12)	(0.19)
Sample Size	375,739	282,440
First-stage coefficient	0.063	0.054
First-stage F-statistics	370.35	211.81

Note: Each coefficient is estimated from a single regression and represents β_1 in Eq. (2) with varying observations included. All standard errors are clustered at the first service date of the injury. *p<0.10, **p<0.05, ***p<0.10.

Appendix Table A2. Main Estimate Stratified by Identifying Injury and Main Sample Excluding Each Injury

Outcome Form	Fractures	Dislocations	Sprains	Intracranial	Open Wounds	Vessels	Late Effects of Poison	Superficial	Contusion
Only									
y_{it}	41,812***	36,972***	14,054**	96,520***	21,729*	463,632	6,367	21,594	9,086
	(10,174)	(11,569)	(7,044)	(30,952)	(11,890)	(545,816)	(63,755)	(17,127)	(7,941)
y_{it+1}	-2,115	-14,409	-2,978	-1,395	-18,813	-97,881	-34,581	4,786	-5,328
	(8,948)	(10,760)	(7,897)	(14,468)	(12,384)	(248,501)	(52,880)	(16,476)	(9,413)
N	23,847	24,742	87,431	4,334	25,453	267	1,018	13,535	23,908
First-stage coefficient	0.105	0.059	0.041	0.141	0.080	0.073	-0.067	0.057	0.076
First-stage SE	(0.011)	(0.009)	(0.006)	(0.027)	(0.011)	(0.079)	(0.056)	(0.014)	(0.011)
First-stage F-statistic	88.79	39.13	44.94	27.50	56.03	0.85	1.45	16.70	44.75
Less									
y_{it}	35,703***	36,522***	46,137***	33,549***	38,755***	35,850***	36,542***	37,055***	40,597***
	(6,700)	(6,365)	(6,950)	(5,982)	(7,201)	(5,904)	(5,887)	(6,123)	(6,747)
y_{it+1}	-15,539***	-13,501**	-14,932**	-13,910***	-12,707**	-13,185***	-13,374***	-14,339***	-14,413***
	(5,910)	(5,340)	(5,940)	(5,178)	(5,823)	(5,037)	(4,953)	(5,202)	(5,394)
N	231,055	230,160	167,471	250,568	229,449	254,635	253,884	241,367	230,994
First-stage coefficient	0.054	0.059	0.069	0.057	0.056	0.059	0.059	0.058	0.057
First-stage SE	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
F-stat	194.95	209.62	250.94	229.50	199.65	236.88	237.29	213.75	201.74

Outcome Form	Crushing	Foreign	Burns	Nerve	Trauma	Poison	Toxic	Complications
Only								
y_{it}	11,789	69,844	173,041	273,968	16,365*	192,216	2,987	148,992***
	(14,675)	(109,289)	(168,140)	(335,290)	(8,760)	(144,225)	(31,920)	(30,440)
y_{it+1}	2,155	-70,976	64,271	-130,958	-13,939**	-58,502	-59,856	-32,788
	(12,367)	(102,413)	(86,344)	(208,248)	(6,693)	(130,720)	(36,410)	(22,414)
N	675	5,031	2,346	1,165	32,188	2,254	17,729	16,789
First-stage coefficient	0.218	0.022	0.034	0.038	0.101	0.051	0.045	0.079
First-stage SE	(0.073)	(0.024)	(0.034)	(0.043)	(0.009)	(0.030)	(0.015)	(0.011)
F-stat	8.78	0.81	1.02	0.78	123.71	2.94	9.27	54.60
Less								
y_{it}	36,948***	36,496***	36,022***	36,139***	42,769***	35,320***	38,653***	22,336***
	(5,974)	(5,970)	(5,944)	(5,940)	(7,066)	(6,109)	(5,775)	(4,533)
y_{it+1}	-13,335***	-12,900**	-13,736***	-12,819***	-13,146**	-13,199***	-10,877**	-12,279***
	(5,045)	(5,042)	(5,051)	(4,946)	(6,460)	(4,979)	(5,069)	(4,210)
N	254,227	249,871	252,556	253,737	222,714	252,648	237,173	238,113
First-stage coefficient	0.059	0.059	0.059	0.059	0.052	0.059	0.060	0.058
First-stage SE	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)	(0.004)
F-stat	233.05	241.53	234.24	234.85	176.78	233.44	274.96	197.96