Using the age-based insurance eligibility criterion to estimate moral hazard in medical care consumption

YAN ZHENG, PhD*
TOMISLAV VUKINA, PhD*

Article**
JEL: C14, D82, I11
doi: 10.3326/fintp.39.3.3

* We are thankful to Xiaoyong Zheng, Barry Goodwin, Melinda Sandler Morrill and two anonymous referees for their valuable comments and suggestions.

** Received: June 1, 2015
Accepted: January 28, 2016

The article was submitted for the 2015 annual award of the Prof. Dr. Marijan Hanžeković Prize.

Yan ZHENG
Credit Analyst, JP Morgan Chase and Co. 201 N. Walnut Street, Wilmington, DE 19801, USA
e-mail: zoeyyzheng@gmail.com

Tomislav VUKINA
Professor, Department of Agricultural and Resource Economics, North Carolina State University, 3332 Nelson Hall, Raleigh, NC 27695-8109, USA
e-mail: tom_vukina@ncsu.edu

Abstract

This paper uses fuzzy regression discontinuity design to estimate the moral hazard effect in health care consumption in the population of young adults. We use invoice data for outpatient hospital services from a regional hospital in Croatia. The estimation is complicated by the fact that the data set consists only of users of medical services, which would tend to underestimate the moral hazard effect. To address this issue we use a modified version of the instrumental variables approach. We find a 92% reduction in the number of hospital visits for individuals who lost insurance coverage when crossing the 18th birthday threshold.

Keywords: fuzzy regression discontinuity design, instrumental variables, health insurance

1 INTRODUCTION

A large body of health economics literature documents a strong association between health insurance status and patterns of health care utilization. People with more generous insurance coverage tend to consume more health care, a phenomenon known as moral hazard. The literature suggests that when individuals lose health insurance, they change their consumption of health care services, i.e. they seek medical attention less frequently. But would the uninsured consume more health care if they had health insurance? Such a causal inference is difficult because uninsured individuals are likely to have different health conditions, attitudes towards risk, disposable income, wealth, and so on than insured individuals.

To overcome this problem, we exploit quasi-experimental variation in insurance status resulting from the rules insurance companies use to establish the coverage eligibility of dependents. One of the largest segments of the population that lacks health insurance is young adults (age 19 to 29). For example, in the United States, 29% of the uninsured are young adults (Schwartz and Schwartz, 2008). Because many health insurance contracts cover dependents until the age of 18 and only cover older dependents if they are full time students, a significant number of teenagers become uninsured after they reach the threshold birthday. Because young adults are relatively healthier than the older population, it is reasonable to assume that they would consume relatively low levels of health care regardless of whether they are insured or uninsured and that the social cost of extending coverage to uninsured young adults would be relatively modest.

The use of terms moral hazard has its origins in the insurance literature and has subsequently spread into contract theory and information economics. Contract theory refers to moral hazard as an asymmetric information problem arising when an agent's (insured's) behavior is not observable by the principal (insurance company). The magnitude of moral hazard is then measured by the welfare difference between the first best (symmetric information) and the second-best (asymmetric information) outcome. In the context of health insurance, however, moral hazard is often used in reference to the price elasticity of demand for health care, conditional on underlying health status (Paully, 1968; Cutler and Zeckhauser, 2000; Einav et al., 2013). The approach that we use in this paper is conceptually in line with this mainstream health insurance literature. In other words, our approach does not consider the potential impact of insurance on underlying health or its impact on risky behavior that may have detrimental effects on health, as for example in Yörük (2015). Instead, moral hazard is simply the consequence of the demand function for health care being less than perfectly inelastic (vertical) such that any reduction in the price of health care services resulting from owning an insurance policy increases the quantity demanded. The measure of moral hazard is then simply the difference in health care consumption between the full price of medical services and the lower price one pays as the owner of an insurance policy, conditional on health status.

The empirical evidence on the importance of moral hazard in health insurance markets varies by country, type of health service provided and socio-economic status of insured. For example, using Australian data, Cameron et al. (1988) established that more generous coverage leads to higher utilization of a broad range of services. They found a significant price effect in health care consumption, which implies that moral hazard is an important determinant of overall health care utilization. Also for Australia, Savage and Wright (2003) found that after correcting for endogeneity, the extent of moral hazard can increase the expected length of a hospital stay up to three times.

Liu, Nestic and Vukina (2012), using matching estimators, found the presence of a significant moral hazard effect in the Croatian health care system. In the case of Spain, Vera-Hernandez (1999) found no evidence of moral hazard for heads-of-households and strong evidence for other household members. Holly et al. (1998), using data for Switzerland, found evidence of moral hazard in hospital stays. Coulson et al. (1995) found that supplemental insurance increases the number of prescriptions filled by the elderly in the United States. Manning et al. (1987) used a randomized experiment and found that a catastrophic insurance plan reduces expenditures 31% relative to zero out-of-pocket price, indicating a large moral hazard effect. Cardon and Hendel (2001) integrated health insurance and health care demand using 1987 National Medical Expenditure Survey data and found that the gap in expenditure between insured and uninsured can be attributed to observable demographic differences and to price sensitivity. They interpreted the elasticity of demand with respect to price (coinsurance rate) as evidence of moral hazard.

In this paper, we rely on the fuzzy regression discontinuity (RD) design to estimate the effect of losing insurance on health care consumption in young adults. We use invoice data for outpatient hospital services from a regional hospital in Croatia. Croatia has a state-run health insurance system dominated by a single public health insurance fund, the Croatian Institute for Health Insurance (HZZO). The HZZO offers compulsory and supplemental insurance. The former covers some medical care services fully and some subject to co-payments. Full coverage of medical care services is provided to children younger than 18 and to all patients suffering from specific serious illnesses. All other health services are subject to co-payments. In order to avoid these co-payments, supplemental insurance is required. The compulsory insurance coverage is universal whereas supplemental insurance can be either bought or is extended automatically free of charge to some categories of citizens such as, for example, full time students.

In this context, the 18th birthday represents a threshold for supplemental insurance coverage; young adults crossing the threshold will lose full coverage and face three options: stay in school and continue being fully covered for free, buy supplemental insurance and continue enjoying full coverage or refuse to buy the supplemental coverage and pay co-payments as required.

Since crossing the 18th birthday threshold is not a unique determinant of assignment into the treatment (losing the coverage), the problem fits into the fuzzy RD design. RD design can be used to determine the treatment effects in quasi-experimental settings where treatment is determined by a forcing variable exceeding the threshold. The probability of being treated at the threshold jumps from zero to one (see: Lee and Lemieux, 2010). In the fuzzy RD design, the forcing variable does not exclusively determine the treatment assignment, hence the discontinuity in the probability of being treated at the threshold is less than one.

Regression discontinuity design has been used extensively to determine the causal effect from an intervention. For example, Card, Dobkin, and Maestas (2009) compared health care consumption among people just before and just after the age of 65, the threshold for Medicare eligibility in the U.S., in a sharp RD design. They found that the Medicare eligibility causes a discontinuous increase in health care utilization. Van der Klaauw (2002) identified the causal effect of financial aid on college enrollment decisions using fuzzy RD. He found that enrollment rate increased by about 0.2 at the financial aid eligibility threshold. Chen and van der Klaauw (2008) evaluated the effect of disability insurance program on labor force participation among disability insurance beneficiaries using fuzzy RD design. They found that the labor force participation rate among those beneficiaries would have been 20% higher if none had received the benefits. Anderson, Dobkin, Gross (2012) used the age 19 as an instrument to identify the causal effect of loss of health insurance at age 19 on health care consumption. They found that not having insurance leads to a decreasing level of health care consumption, manifested in a 40 percent reduction in emergency department visits and a 61 percent reduction in inpatient hospital admissions.

A distinct feature of our data set which considerably complicates the estimation is that it consists of users only. People that did not use the medical services of the hospital during the time period covered by the data do not show up in the data. The problem of estimating the moral hazard effect with users-only data is caused by the fact that some sick people do not seek medical attention at all or seek it less often precisely because they do not have the insurance. Using users-only (patients) data clearly underestimates the moral hazard effect. To deal with this attenuation bias we use the instrumental variables (IV) approach from Anderson, Dobkin and Gross (2012), which relies on the assumption that the net change in the observed hospital visits after the age threshold of 18 is driven only by individuals who have lost their insurance coverage, an assumption implied by the standard IV exclusion restriction. We found a statistically significant reduction in the number of hospital visits by 92% for young adults who had lost their supplemental insurance after turning 18, confirming the moral hazard hypothesis.

2 INSTITUTIONAL FRAMEWORK AND DATA DESCRIPTION

The healthcare system in Croatia is still largely dominated by the institutional setup inherited from socialism. Despite the fact that the generous benefits and exemptions inherited from the old system have been politically difficult to roll back, Croatia has embarked on a number of reform initiatives that resulted in a relative decline in total spending on health care. Reforms have included enlarging the participation scheme (co-payments), reducing the number of individuals exempt from participation, introduction of administrative fees, and some cost savings in prescription drugs expenditures (Liu, Nestic and Vukina, 2012). Croatia still spends 7.9% of its GDP on health, among the highest for new EU members. In a fiscally constrained environment, the Croatian health system faces a mismatch between declining available public resources, growing expenditures and the increasing needs of an ageing population (Word Bank, 2015).

The main characteristics of the system provided by the HZZO can be summarized as follows. The compulsory insurance, which is funded by a 15% payroll tax, covers two kinds of medical care services: one with full coverage and the other with a system of co-payments. Full coverage medical care services are provided to children up to 18 years of age, pregnant women and for everybody else for life-threatening types of conditions such as infectious diseases, psychiatric care, surgeries, cancers and mandatory vaccinations, all other health services (including but not limited to primary care, hospitals stays and prescription drugs) are subject to a system of co-payments. The patients are required to pay 20% of the full price of medical care, with the largest out-of-pocket cost share amount set at 3,000.00 HRK per invoice.[1] Supplemental insurance is voluntary and can be acquired by a person 18 years or older who has compulsory insurance, by signing a contract with the HZZO. Certain categories of citizens are entitled to the supplemental insurance free of charge, i.e. their premiums are covered from the state budget. The list of people entitled to free supplemental insurance includes full time secondary school and college students. For those not entitled to free supplemental coverage, the premiums range from 50 to 130 HRK per month depending on income and whether the insured is active or retired. A person having the supplemental insur-

[1] The figures are reflective of the year 2009 which is the year covered by the dataset. The exchange rate for the local currency, Croatian kuna (HRK), as of June 20, 2009 was 1USD=5.19 HRK. A more detailed description of the health insurance system in Croatia can be found in Liu, Nestic and Vukina (2012).

FINANCIAL THEORY AND PRACTICE
40 (3) 337-356 (2016)

ance is entitled to full waiver of all the medical expense co-payments mentioned before.

The original data set consists of all invoices for all outpatient services from a small hospital in Croatia during the period from March 1 to June 30, 2009.[2] The data set consists of 105,646 observations. Each observation reflects one hospital visit (invoice). The data contains the following set of variables: a numeric code for the type of hospital service provided, compulsory health insurance number, supplemental insurance number (if the patient has one), period covered by the supplemental insurance, numeric code for categories entitled to supplemental insurance free of charge, eligibility category for compulsory insurance, cost of hospital service, part of the cost covered by compulsory insurance, part of cost covered by supplemental insurance, part of cost covered by participation (co-payment), date of birth and sex of the patient. The invoices do not record the exact date when the patient visited the hospital, but are chronologically ordered. To determine the date of the visit, we first divided all invoices into 122 days (March 1 to June 30) and designated the first 866 invoices as March 1st invoices, the next 866 invoices as March 2nd invoices, etc. This is a fairly innocuous simplification because the actual number of daily appointments is determined by the hospital's outpatient capacity and it is, therefore, reasonable to assume that the average number of patients treated each day is approximately the same. Since we know the date of birth of each patient and knowing the day of the visit, we can calculate the number of days each patient was away from the 18th birthday when they visited the hospital. For patients younger than 18, the days away from the 18th birthday are recorded as negative numbers and for patients older than 18 as positive numbers.

3 REGRESSION DISCONTINUITY DESIGN

In the basic setting for the regression discontinuity design researchers are interested in the causal effect of a binary intervention or treatment. Individuals are either exposed or not exposed to a treatment. Let W_iÎ{0, 1} denote a treatment with $W_i = 1$ if unit i was exposed to the treatment and $W_i = 0$ otherwise. Let $Y_i(W_i)$ denote the outcome where $Y_i(0)$ is the outcome without the exposure to treatment and $Y_i(1)$ is the outcome with exposure to treatment. We are interested in the difference $Y_i(1) - Y_i(0)$. Since one can never observe $Y_i(0)$ and $Y_i(1)$ at the same time, one needs to use the average effects of the treatment over the population or a sample to estimate the average treatment effect.

It is important to distinguish two RD settings: sharp and fuzzy regression discontinuity designs. Based on Imbens and Lemieux (2008), in the sharp regression discontinuity (SRD) design, the assignment W_i is a deterministic function of one

[2] The hospital in question is one of the general medicine regional hospitals in Croatia which provides all typical services such as internal medicine, surgery, anesthesiology, pediatrics, psychiatry, neurology, obstetrics and gynecology, radiology, pathology and emergency care at a fairly standard (basic) level. All complicated cases are typically referred to one of the larger medical centers in the capital Zagreb. The name and the location of the hospital are suppressed for confidentiality reasons.

of the covariates, i.e. the forcing (or treatment-determining) variable X such that $W_i = 1$ if X_i ³ c and $W_i = 0$ if $X_i <$ c, where c is some constant. Therefore, all units with the value of the forcing variable $X_i = x$ at least c are assigned to the treatment group and participation is mandatory whereas all units with the value of forcing variable $X_i = x$ less than c are assigned to the control group. In the SRD design, the interest is to use the discontinuity in the conditional expectation of the outcome given the forcing variable in order to estimate the average treatment effect:

$$\lim_{x \downarrow c} E[Y_i | X_i = x] - \lim_{x \uparrow c} E[Y_i | X_i = x], \tag{1}$$

which is interpreted as the average causal effect of the treatment:

$$\tau_{SRD} = E[Y_i(1) - Y_i(0) | X_i = c] \tag{2}$$

at the discontinuity point.

In the fuzzy regression discontinuity (FRD) design, however, the probability of being treated does not need to jump from zero to one at the threshold. The design allows other factors to influence the assignment to treatment, besides the forcing variable. Therefore, only a jump in the probability of assignment to the treatment at the threshold is required:

$$\lim_{x \downarrow c} \Pr(W_i = 1 | X_i = x) \neq \lim_{x \uparrow c} \Pr(W_i = 1 | X_i = x). \tag{3}$$

The average causal effect of the treatment:

$$\tau_{FRD} = \frac{\lim_{x \downarrow c} E[Y | X = x] - \lim_{x \uparrow c} E[Y | X = x]}{\lim_{x \downarrow c} E[W | X = x] - \lim_{x \uparrow c} E[W | X = x]} \tag{4}$$

is identified as the ratio of the jump in the outcome at the threshold over the jump in the treatment indicator at the threshold. In our particular case expression (4) measures the average causal effect on health care utilization (as measured by the number of hospital visits) of losing insurance as the result of turning 18.

3.1 INSTRUMENTAL VARIABLE APPROACH

An important reason to choose FRD instead of some other causal inference methods is that it naturally solves the endogeneity problem by using the discontinuity as an instrument (Lee and Lemieux, 2010). To estimate the ratio in equation (4) we follow the approach of Anderson, Dobkin and Gross (2012). To explain the procedure, consider the following reduced form model of the effect of health insurance coverage on health care utilization:

$$Y_i = \gamma_0 + \gamma_1 D_i + \varepsilon_i, \tag{5}$$

where Y_i indicates health care utilization of individual i, D_i is an indicator variable equal to 1 if individual i has health insurance and 0 otherwise and γ_0 and γ_1 are coefficients to be estimated. All other determinants of health care utilization are summarized in the error term ε_i. Therefore, the coefficient γ_1 denotes the causal effect of health insurance on health care utilization, also known as the moral hazard effect.

However, this moral hazard effect is contaminated by other factors because the insurance coverage variable D_i is correlated with other determinants of health care utilization in ε_i, for example, the unobserved health status of that individual. An individual chooses to acquire health insurance based on idiosyncracies that simultaneously affect the decision to be insured and the consumption of health care. Therefore the estimate for γ_1 is inconsistent due to endogeneity. To solve the problem one typically relies on the IV approach. The objective is to estimate the causal effect of losing health insurance coverage on the number of hospital visits. Using crossing the 18th birthday threshold as an instrument, we estimate the first stage equation – the share of young adults who lose the insurance coverage while crossing the age 18 threshold and the reduced form equation – the change in the number of hospital visits associated with attaining the age of eighteen. We identify the effect of the moral hazard (coefficient γ_1) by dividing the reduced form estimate (the effect of turning 18 on the number of visits) by the first stage estimate (the effect of turning 18 on health insurance coverage).[3]

Notice however, that having hospital invoices data introduces sample selection bias in the first stage estimation because we only observe the insurance status for individuals who do show up in the hospital and do consume some services for which they or their insurer is charged. Regression estimate of the change in the proportion of uninsured after crossing the age 18 threshold understates the true size of this change. Because loss of insurance reduces the likelihood of a hospital visit and therefore affects the probability of appearing in the sample, this selection mechanism leads to an attenuation bias when estimating the change in the insurance coverage because newly uninsured individuals are more likely to leave the sample.

To correct for the bias in the first-stage estimates we assume that the net change in the observed hospital visits after crossing the age 18 threshold is driven only by individuals who have lost insurance coverage. This assumption is implied by the standard IV exclusion restriction and is quite reasonable. Because all our patients in the *Y-18* group are fully insured, those who did not go to the hospital before 18 (when they had insurance) and hence did not show up in the data as users, are unlikely to visit hospital immediately after turning 18 even if they have insurance and even less likely if they don't have insurance.

Let D_i denote the insurance coverage indicator and A_i = a denote the age of an individual. Then the effect of crossing the age 18 threshold on insurance coverage at the population level can be expressed as:

[3] This strategy is analogous to using the age 18 discontinuity as an instrument to identity the causal effect of health insurance; see Hahn, Todd and van der Klaauw (2001).

$$\lim_{a \downarrow 18} E[D_i | A_i = a] - \lim_{a \uparrow 18} E[D_i | A_i = a], \tag{6}$$

Next, let $D_i(1) = 1$ indicate an individual older than 18 with insurance coverage and $D_i(1) = 0$ an individual older than 18 without insurance. $D_i(0) = 1$ and $D_i(0) = 0$ are defined similarly for individuals younger than 18. Also, let $Y_i(1) = 1$ denote an individual older than 18 who did visit the hospital and $Y_i(1) = 0$ an individual older than 18 who did not visit the hospital. $Y_i(0) = 1$ and $Y_i(0) = 0$ are defined similarly for individuals younger than 18. Since we could only observe individuals who visited the hospital (users), the effect of turning 18 on insurance coverage among users is estimated as:

$$E[D_i(1)|Y_i(1)=1] - E[D_i(0)|Y_i(0)=1]. \tag{7}$$

However, we would like to estimate the effect of turning 18 on insurance coverage at the population level, given that they visited the hospital when they were younger than 18:

$$E[D_i(1)|Y_i(0)=1] - E[D_i(0)|Y_i(0)=1]. \tag{8}$$

The desired effect is estimated as follows. First we denote the number of visits made before 18 as y_0, the number of insured visits made before 18 as d_0, the number of visits made after 18 as y_1 and the number of insured visits made after 18 as d_1. The ratios $\frac{d_0}{y_0}$ and $\frac{d_1}{y_1}$ represent the fractions of insured visits before and after the age of 18. The corresponding fractions of uninsured visits before and after the age of 18 are denoted as $(1-\frac{d_0}{y_0})$ and $(1-\frac{d_1}{y_1})$. It can be shown that the bias-corrected estimate for the effect of crossing 18 on the insurance coverage is obtained as:

$$\frac{d_1 - d_0}{y_0} \xrightarrow{p} E[D_i(1) - D_i(0)|Y_i(0)=1], \tag{9}$$

where $\xrightarrow{p}$ denotes the convergence in probability. Expression (9) then provides the population level estimate defined by equation (8).[4]

Under a simplifying assumption that number of visits per patient is constant[5], equation (9) can be estimated as:

$$\frac{\frac{d_1}{y_1} * y_1}{y_0} - \frac{d_0}{y_0} = \frac{\lim_{a \downarrow 18} E[D_i | A_i = a] * \lim_{a \downarrow 18} E[Y_i | A_i = a]}{\lim_{a \uparrow 18} E[Y_i | A_i = a]} - \lim_{a \uparrow 18} E[D_i | A_i = a]. \tag{10}$$

Since we are interested in estimating how an increase in the proportion of the uninsured affects the change in hospital visits as people turn 18, we need to estimate the following equation:

[4] For proof see web support to Anderson, Dobkin and Gross (2012).

[5] As explained later in the estimation procedure section, this approach does not exploit the panel structure of the data set, i.e. we are only using the number of visits (insured and uninsured) in each week regardless of the number of patients that generated those visits.

$$\frac{d_0-d_1}{y_0}=\frac{\lim_{a\downarrow 18}E[U_i|A_i=a]*\lim_{a\downarrow 18}E[Y_i|A_i=a]}{\lim_{a\uparrow 18}E[Y_i|A_i=a]}+(1-\frac{\lim_{a\downarrow 18}E[Y_i|A_i=a]}{\lim_{a\uparrow 18}E[Y_i|A_i=a]})$$
$$-\lim_{a\uparrow 18}E[U_i|A_i=a] \tag{11}$$

where U_i is an indicator equal to one if individual i is uninsured, i.e. $U_i=1-D_i$.

In the second step, we estimate the reduced form equation, i.e. the percentage decline in visits due to crossing the age 18 threshold, which can be written as:

$$\frac{y_1-y_0}{y_0}, \tag{12}$$

and obtained by the following estimator:

$$\frac{\lim_{a\downarrow 18}E[Y_i|A_i=a]-\lim_{a\uparrow 18}E[Y_i|A_i=a]}{\lim_{a\uparrow 18}E[Y_i|A_i=a]}. \tag{13}$$

It can be shown that (12) converges in probability to:

$$\frac{y_1-y_0}{y_0}\xrightarrow{p}E[Y_i(1)-Y_i(0)|D_i(1)-D_i(0)$$
$$=-1,Y_i(0)=1]*E[D_i(1)-D_i(0)|Y_i(0)=1] \tag{14}$$

where the first factor on the right hand side denotes the measure of moral hazard for individuals who visited the hospital before turning 18 and lost the insurance coverage after turning 18 and the second factor denotes the first stage estimate from equation (9), i.e. the change in insurance coverage due to reaching one's 18th birthday.[6] Therefore, the measure of moral hazard is computed as the ratio of the reduced form estimate and the first stage estimate.

3.2 ESTIMATION PROCEDURE

The key decision in implementing RD design relates to the choice of bandwidth. In most current empirical studies researchers often choose a bandwidth by either cross-validation or by *ad hoc* methods. We rely instead on the optimal, data-dependent, bandwidth choice rule introduced by Imbens and Kalyanaraman (2012). Estimation of the optimal bandwidth was performed with the invoice level data using uniform kernel on [-1,+1] interval. Both a one-year and a two-year data window surrounding the 18th birthday yield very similar estimates of the optimal bandwidth of 1.24 and 1.23 years, respectively (64 weeks surrounding the threshold).[7]

Once optimal bandwidth has been determined, all invoices for the patients outside the 64 weeks surrounding the 18th birthday threshold were dropped, leaving the

[6] The proof of this result is also contained in the web support to Anderson, Dobkin and Gross (2012).
[7] A detailed technical description of the estimation algorithm is available from authors upon request.

sample of 1,883 invoices in total. The summary statistics for this 64 weeks window are presented in table 1. In the left panel we look at the number of visits (invoices). If the invoice shows that the patient visited the hospital with both compulsory and supplemental insurance, we call it an insured visit, otherwise it is an uninsured visit. In other words, an uninsured visit in this paper means a visit to the hospital when the patient did not have the supplemental insurance. Therefore, the younger than 18 group (*Y-18*) has only insured visits whereas the older than 18 group (*O-18*) has both insured and uninsured visits. The number of uninsured visits is only 3% of the total visits.

Table 1

Summary statistics: 64-week window around the age 18

	Visits					Patients			
Sample	Invoices	Insured	Uninsured	Male	Female	Number	Male	Female	Visits/patient
Y-18	1,063	1,063	0	480	583	393	181	212	2.71 (3.16)
O-18	820	763	57	372	448	360	152	208	1.98 (1.85)
Total	1,883	1,826	57	852	1,031	753	333	420	2.36 (2.64)

Note: The numbers in parentheses are standard deviations.

We also compared the number of visits by gender. In both *Y-18* and *O-18* groups, women visited the hospital more frequently than men. In the right panel we look at the summary statistics at the patient level. We see that, overall, the sample contains more women than men, which probably explains why we observe more female than male visits. Also, the average number of visits per patient has dropped when the age 18 threshold is crossed from 2.71 to 1.98 visits per patient. The *O-18* group has fewer people visiting the hospital and a lower number of visits per patient but the number of female patients dropped less than the number of male patients as they passed their 18th birthday.

We start the estimation by collapsing the individual invoices data into weekly data and count the number of total and uninsured visits in each week around the patients' 18th birthday. The numbers of days away from the 18th birthday are converted into the numbers of weeks away from the 18th birthday.

The scatter plot and fitted values of the percentages of uninsured visits in each of the 64 weeks away from age 18 are depicted in figure 1. The fitted lines are estimated separately on each side of the cutoff point. There is a clear jump in the percentage of uninsured visits at the cutoff point at age 18. The proportion of uninsured visits is zero for people younger than 18 because all are fully covered by both compulsory and supplemental insurance.

The scatter plot and fitted values of the log of the total visits in each of the 64 weeks surrounding the 18th birthday are displayed in figure 2. There is a sharp drop in hospital visits at the cutoff point at age 18. The graphs provide powerful visual evidence supporting our choice of the regression discontinuity design.

FIGURE 1
Uninsured visits surrounding patients' 18th birthday (64-week window)

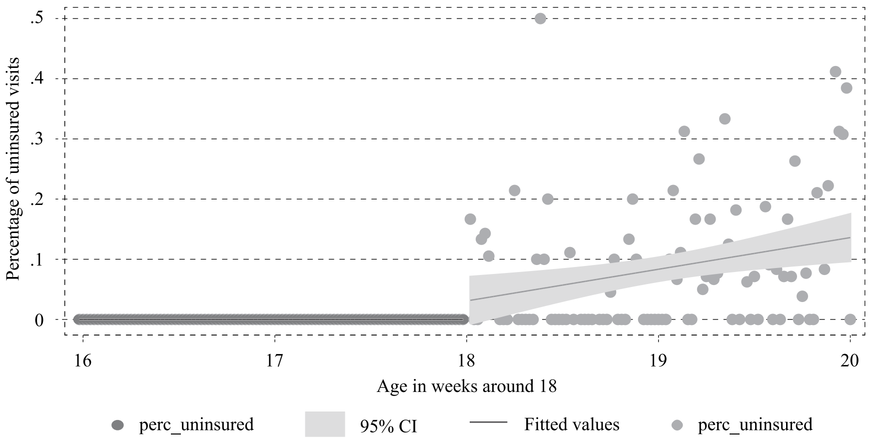
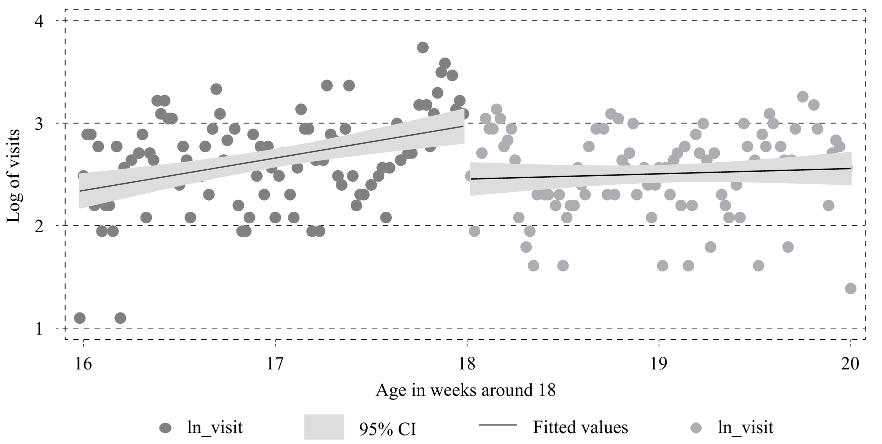


FIGURE 2
Hospital visits around patients' 18th birthdays (64-week window)



The primary concern in the RD design is that factors other than insurance coverage such as high school graduation, starting college, obtaining driving license, entering legal drinking age and starting employment could change discontinuously at the age of 18 and dramatically alter the need to access health care. Because we measure age at the weekly level, only factors that change sharply within a few weeks around the age 18 threshold could bias our estimates. As it turns out, most of those obvious confounders should not bias our estimates. First, high school graduations occur in June and universities start classes in September or October but 18th birthdays are distributed throughout the year. Second, Croatia has a law on the books prohibiting people younger than 18 to purchase or consume alcohol. Yet, there is ample anecdotal evidence that the enforcement is fairly lax, especially in establishments outside the larger cities and along the Adriatic coast during the tourist season. Finally, a potentially interesting factor is the legal driving age. At age 18 young adults in Croatia become eligible to drive. However, since the drivers' education process is quite lengthy and expensive, most young adults will not obtain their licenses exactly on their 18^{th} birthday but rather during that year or even later.

Based on the framework developed before, in the first step we estimate the limit expectations in equation (11) by seemingly unrelated regression (SUR) using the following two equations:

$$1-\frac{d_j}{y_j}=\alpha_0+\alpha_1 I_{over18j}+\alpha_2 aweek_j+\alpha_3 aweek_j * I_{over18j}+\varepsilon_j$$

$$\ln y_j=\beta_0+\beta_1 I_{over18j}+\beta_2 aweek_j+\beta_3 aweek_j * I_{over18j}+\mu_j \qquad (15)$$

where $j=1, 2, ..., 128;$ d_j and y_j are weekly counts of insured visits and total visits respectively, $aweek_j$ is age in weeks equals $\frac{1}{64}$ if the visits are made in the first week after turning 18, equals $\frac{2}{64}$ if the visits are made in the second week after turning 18, etc., $aweek_j$ equals 0 if the visits are made on the 18^{th} birthday, $aweek_j$ equals $-\frac{1}{64}$ if the visits are made in the first week before turning 18, etc., $I_{over18j}$ is an indicator variable equal to one if observation j is older than 18 and ε_j, μ_j are random errors with $E(\varepsilon_j)=0$, $E(\mu_j)=0$ and $Cov(\varepsilon_j, \mu_j)\neq 0$. The dependent variables are the weekly percent of uninsured visits $1-\frac{d_j}{y_j}$ and the natural log of weekly number of visits $\ln y_j$. Since the errors affecting percentage of uninsured visits and total visits in each week are likely to be correlated, the use of SUR is justified.

Using estimated coefficients from equation (15), the limit expectations in equation (11) can be recovered as follows:

$$\lim_{a\downarrow 18} E[U_i|A_i=a]=\hat{\alpha}_0+\hat{\alpha}_1$$

$$\lim_{a\downarrow 18} E[Y_i|A_i=a]=\exp(\hat{\beta}_0+\hat{\beta}_1)$$

$$\lim_{a\uparrow 18} E[Y_i|A_i=a]=\exp(\hat{\beta}_0)$$

$$\lim_{a\uparrow 18} E[U_i|A_i=a]=\hat{\alpha}_0$$

where $\hat{\alpha}_2$, $\hat{\alpha}_3$, $\hat{\beta}_2$ and $\hat{\beta}_3$ drop out because $aweek=0$ at the limit of age 18. Therefore, the first stage estimator for the increment in the proportion of uninsured at age 18 becomes:

$$\frac{(\hat{\alpha}_0+\hat{\alpha}_1)\exp(\hat{\beta}_0+\hat{\beta}_1)}{\exp(\hat{\beta}_0)}+1-\frac{\exp(\hat{\beta}_0+\hat{\beta}_1)}{\exp(\hat{\beta}_0)}-\hat{\alpha}_0 \qquad (16)$$

In the second step we estimate the limit expectations in equation (13) using the following equation:

$$\ln y_j = \beta_0 + \beta_1 I_{over\,18\,j} + \beta_2 aweek_j + \beta_3 aweek_j * I_{over\,18\,j} + \mu_j \qquad (17)$$

which turns out to be identical to the second equation from (15). Therefore, the reduced form estimator for the percentage change in hospital visits for individuals turning 18 is:

$$\frac{\exp(\hat{\beta}_0 + \hat{\beta}_1) - \exp(\hat{\beta}_0)}{\exp(\hat{\beta}_0)} = \exp(\hat{\beta}_1) - 1 \qquad (18)$$

where $\hat{\beta}_2$, $\hat{\beta}_3$ drop out in the limit.

Based on equation (14), it is straightforward to see that the moral hazard effect y_1 is identified by dividing the effect of turning 18 on hospital visits, i.e. the reduced form estimator from equation (18), by the effect of turning 18 on insurance coverage, i.e. the first stage estimator from equation (16). Thus, the IV estimator can be written as:

$$\tau_{IV} = \frac{\exp(\hat{\beta}_1) - 1}{\frac{(\hat{\alpha}_0 + \hat{\alpha}_1)\exp(\hat{\beta}_0 + \hat{\beta}_1)}{\exp(\hat{\beta}_0)} + 1 - \frac{\exp(\hat{\beta}_0 + \hat{\beta}_1)}{\exp(\hat{\beta}_0)} - \hat{\alpha}_0} \qquad (19)$$

4 RESULTS

The first-step SUR results of equation (15) are presented in table 2. The top panel indicates that the percentage of uninsured visits increases after people turn 18 by about 5%. This represents the short-run effect of loss of insurance. Splitting the sample by gender reveals exactly the same pattern of behavior for young men and women.

Table 2

Seemingly unrelated regression results: 64-week window

Variable: Dependent	Variable: Independent	Total	Male	Female
$1-\frac{d_j}{y_j}$	$I_{over18j}$	0.051 (0.023)**	0.051 (0.023)**	0.051 (0.023)**
	$aweek_j$	0.000 (0.023)	0.000 (0.023)	0.000 (0.023)
	$aweek_j * I_{over18j}$	0.018 (0.033)	0.018 (0.033)	0.018 (0.033)
	Constant	0.000 (0.016)	0.000 (0.016)	0.000 (0.016)
$\ln y_j$	$I_{over18j}$	-0.485 (0.149)***	-0.488 (0.149)***	-0.485 (0.149)***
	$aweek_j$	0.504 (0.146)***	0.491 (0.145)***	0.504 (0.146)***
	$aweek_j * I_{over18j}$	-0.548 (0.209)***	-0.536 (0.208)**	-0.548 (0.209)***
	Constant	2.994 (0.104)***	2.997 (0.103)***	2.994 (0.104)***

*Note: The dependent variables in the regressions are proportion of uninsured visits and log of visits at each age in weeks. Standard errors are in parentheses. *** – 1% significance level; ** – 5% significance level; * – 10% significance level.*

We expect the coefficient reflecting the long-term effect of loss of insurance, $aweek_j * I_{over18j}$, to be negative. This would indicate that with the passage of time, the percentage of uninsured people should decrease, as those who lost the supplemental insurance after the 18th birthday would gradually become more inclined to purchase the coverage. However, our estimate of this coefficient is positive, though not significantly different from zero, indicating that, at least during the first year after reaching the age of 18, the percentage of young adults with supplemental insurance will not change significantly. This seems to indicate that all those that have supplemental insurance beyond the 18th birthday have it by default, i.e. by maintaining their full-time student status rather than by actually purchasing the policy.

The bottom panel of table 2 results indicate that the number of hospital visits decreases by about 32% (exp(0.485)-1) as young adults cross the 18th birthday threshold. Again, there is hardly any difference in the behavior between genders. In this equation, the coefficient on $aweek_j * I_{over18j}$ is negative and statistically significant indicating that the long-term effect of loss of insurance is negative, at least within the first 64 weeks, given that the number of visits further decreases with time (age).

The second-step OLS regression results of the reduced form equation (17) are displayed in table 3. The coefficients are the same as those obtained using SUR (bottom panel of table 2) but the SUR results are more efficient.

Table 3
OLS regression results: 64-week window

Variable		Total	Male	Female
Dependent	**Independent**			
$\ln y_j$	$I_{over18j}$	-0.485 (0.152)***	-0.488 (0.151)***	-0.485 (0.152)***
	$aweek_j$	0.504 (0.148)***	0.491 (0.147)***	0.504 (0.148)***
	$aweek_j * I_{over18j}$	-0.548 (0.212)**	-0.536 (0.211)**	-0.548 (0.212)**
	Constant	2.994 (0.106)***	2.997 (0.105)***	2.994 (0.106)***

Note: The dependent variable in the regression is log of visits at each age in weeks. Standard errors are in parentheses.

The results presented in tables 2 and 3 were obtained with hospital users-only data and as such are not generally valid. We are interested in results that reflect the population as a whole. Table 4 summarizes these results. The asymptotically correct (bias corrected) estimates of the increment in the percentage of uninsured at age 18 (based on equation 16) are displayed in the left-hand column. We estimated a 42% increase in the percentage of uninsured in the whole population, with negligible differences between the female and the male population.

Table 4

Percent change in uninsured (first stage) and total visits (reduced form) and the moral hazard (IV): 64-week window

Sample	First stage	Reduced form	IV estimator
Total	0.416 (0.086)***	-0.385 (0.093)***	-0.924 (0.054)***
Male	0.418 (0.086)***	-0.386 (0.093)***	-0.924 (0.054)***
Female	0.416 (0.086)***	-0.385 (0.093)***	-0.924 (0.054)***

Note: Standard errors are in parentheses.

The percentage change in hospital visits at the age 18 (based on equation 18) is displayed in the middle column of table 4. The three segments of the population all experience a close to 38.5% decrease in the number of hospital visits as they cross their 18^{th} birthday threshold.

The right-hand-side column of table 4 shows the estimates of the moral hazard effect, i.e. the reduction in hospital visits for individuals who visited the hospital before the age of 18 but lost their health insurance after reaching the age of 18. There is a 92% decrease in visits in this group of people. This figure reflects both increased probability of becoming uninsured and the drop in the numbers of hospital visits.

How can one interpret these results? Relying on the summary statistics from table 1, we see that the number of uninsured visits in the *O-18* group is 57, which indicates a 92% reduction in the number of visits due to loss of insurance. Without the moral hazard effect, the number of visits would have been 57/(1-0.924)=750. In other words, 750-57=693 visits never happened because people who lost insurance either never came to the hospital or came less frequently.

We can further compare the actual work load of the hospital staff with the hypothetical workload that would have occurred if nobody lost the insurance after passing their 18^{th} birthday. Assuming the hospital is opened 6 days a week, the actual workload of 17-19 year olds was 18.1 visits per day (1,883 visits/104 days), whereas the hypothetical workload would have been 24.8 visits per day ((1,883+693)/104), a substantial 37% increase.

Another way to look at the result is to take the total number of visits in the *Y-18* group (1,063, table 1) and apply a 38.5% decrease in hospital visits based on the reduced form estimate (middle column, table 4). This would give 409 visits. Taking the number of insured visits in the *Y-18* group (which is also 1,063 because all visits are insured) and applying a 41.6% increase in uninsured visits from the first stage estimates (table 4) yields 442 as the number of uninsured visits. Finally dividing 409/442=0.92 gives us the population-wide estimate of the reduction in the number of visits due to the loss of insurance. This figure coincides with the instrumental variables estimate of moral hazard (table 4, right-hand column).

Since the data used in the estimation covers only a 4-month period, we need to multiply the number of visits that never happened by 3 to obtain an annual estimate of moral hazard at the hospital level. This calculation gives an estimate of 2,079 visits per year.

Finally, the reduction in the number of visits can be converted into a monetary measure of moral hazard by using the average cost per visit (HRK 157.76). This yields around HRK 328,000 (USD 63,200) worth of savings at the hospital level per year.[8] Assuming that the convergence region for the regional hospital in question coincides with the population of the county where the hospital is located, we can obtain a monetary estimate of the moral hazard per capita. Because our estimates of moral hazard are based on a 1.24-year window (64 weeks), we approximate the relevant number of people who passed the age of 18 by the entire 18 year-old cohort plus 24% of the 19 year-old cohort. Based on the 2011 Census, the population of 18-year-olds in the county was 1,401 and the population of 19-year-olds was 1,497, so the moral hazard per person is calculated as 328,000/(1401+0.24*1497) = HRK 186.

How economically significant is this effect at the national level? Based on the 2011 Census figures, the population of 18-year-olds in Croatia was 47,960 and the population of 19-year-olds was 50,790. This gives an estimate of 60,150 young adults who crossed the 18th birthday threshold during the 1.24-year window and the total moral hazard cost of approximately HRK 11.2 million. Therefore, the total cost of extending the supplemental insurance for all young adults in the country from the current age 18 to age 19.4 would consist of two components: (a) the direct loss in collected premiums for supplemental insurance, and (b) the indirect cost of moral hazard. Assuming the highest premium for supplemental insurance of HRK 130 per month gives us a total cost of HRK 7.8 million per year in lost premiums plus the indirect cost of moral hazard of HRK 11.2 million per year, for the total of HRK 19 million.

5 CONCLUSION

In this paper we implemented fuzzy regression discontinuity design to estimate the moral hazard effect in the health care consumption of young adults drawing on invoice data from one small regional hospital in Croatia. The challenge in estimating moral hazard with such a dataset is that we only observe people who actually used medical care services within the time period covered by the data. To deal with this sample selection bias, we estimate the causal effect of insurance on medical care consumption using the 18th birthday as an instrumental variable. The 18th birthday represents a threshold at which young adults will, by default, lose the full

[8] A more precise monetary estimate of the moral hazard effect would be possible by looking at the change in the cost structure of visits before and after crossing the age 18 threshold. The average cost per visit for the *Y-18* group is HRK 132.90, for *O-18* group HRK 189.99, and for both groups combined HRK 157.76. However, since our estimation procedure relies on the visits data and not the cost data, this calculation would be less reliable.

coverage unless they remain full time students or decide to buy the coverage on their own. The estimation uses weekly counts of hospital visits. Results based on the 64 weeks window estimation show that there is an 92% decrease in hospital visits due to loss of insurance among individuals who did visit the hospital before age 18 and lost their insurance after passing the age of 18.

As an illustration, the above numbers can be related to an alternative program recently promulgated by the Croatian government. Given a severe youth unemployment problem[9], starting January 1, 2015, the government has launched a program whereby the employers are excused from paying the payroll taxes of 17.2% (health insurance earmarked tax rate of 15% plus some other taxes of 2.2%) for a period of up to 5 years for every new employee below the age of 30. For an employee with an average net salary of HRK 5,800 per month the total savings for the employer amount to HRK 11,971 per year. Of course, the amount of money saved by an employer represents a loss to HZZO or the state budget. Relying on the data published by the Croatian Employment Service (Hrvatski zavod za zapošljavanje, HZZ), the total unemployment in the 15-29 age cohort amounts to 102,483 persons. Assuming an optimistic scenario whereby the newly introduced measure will reduce youth unemployment by about 10% or 10,250 people per year, the total cost to the HZZO or the state budget can be estimated at HRK 122.7 million per year.

A comparison of benefits of the two mentioned programs would require a more elaborate modeling and simulation, which is outside the scope of this paper. Nevertheless, the above example highlights several interesting points. First, the cost of moving the threshold for free supplemental insurance from 18 to 19.4 years would cost less than one sixth of the employment stimulus package currently in place. Second, a significant part of that cost is the cost of moral hazard and not the budgetary costs due to lost insurance premiums, hence, the actual impact on hospitals' workloads in terms of increased number of visits could be significant. However, the national health effect of such a program could be substantial. Recall that losing the supplemental insurance at the age of 18 caused a reduction in the number of hospital visits. Some of those lost visits could have profound impacts on early disease detection and prevention and could actually save a lot of money for the health insurance system in the future.

[9] According to Eurostat, youth unemployment in the fourth quarter of 2013 was a staggering 48.6%.

REFERENCES

1. Anderson, M., Dobkin, C. and Gross, T., 2012. The Effect of Health Insurance Coverage on the Use of Medical Services. *American Economic Journal: Economic Policy,* 4(1), pp. 1-27. doi: 10.1257/pol.4.1.1
2. Cameron, A. C. [et al.], 1998. A Microeconometric Model of the Demand for Health Care and Health Insurance in Australia. *Rev. of Economic Studies,* (1), 85-106.
3. Card, D., Dobkin, C. and Maestas, N., 2009. Does Medicare Save Lives. *The Quarterly Journal of Economics,* 124(2), pp. 597-636. doi: 10.1162/qjec.2009.124.2.597
4. Chen, S. and van der Klaauw, W., 2008. The work disincentive effects of the disability insurance program in the 1990s. *Journal of Econometrics,* 142, pp. 757-784. doi: 10.1016/j.jeconom.2007.05.016
5. Coulson, N. E. [et al.], 1995. Estimating the Moral-Hazard Effect of Supplemental Medical Insurance in the Demand for prescription Drugs by the Elderly. *American Economic Review,* 85(2), pp. 122-126.
6. Cutler, D. and Zeckhauser, R., 2000. "The Anatomy of Health Insurance" In: A. J. Culyer and J. P. Newhouse, eds. *Handbook of Health Economics Vol. 1A*. Amsterdam: North Holland. doi: 10.1016/S1574-0064(00)80170-5
7. Einav, L. [et al.], 2013. Selection on Moral Hazard in Health Insurance. *The American Economic Review,* 103(1), pp. 178-219. doi: 10.1257/aer.103.1.178
8. Hahn, J., Todd, P. and van der Klaauw, W., 2001. Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design. *Econometrica,* 69(1), pp. 201-209. doi: 10.1111/1468-0262.00183
9. Holly, A. [et al.], 1998. An Econometric Model of Health Care Utilization and Health Insurance in Switzerland. *European Economic Review,* 42, pp. 513-522. doi: 10.1016/S0014-2921(98)00003-8
10. Imbens, G. W. and Kalyanaraman, K., 2012. Optimal Bandwidth Choice for the Regression Discontinuity Estimator. *Review of Economic Studies,* 79(3), pp. 933-959. doi: 10.1093/restud/rdr043
11. Imbens, G. W. and Lemieux, T., 2008. Regression Discontinuity Designs: A Guide to Practice. *Journal of Econometrics,* 142(2), pp. 615-635. doi: 10.1016/j.jeconom.2007.05.001
12. Lee, D. S. and Lemieux, T., 2010. Regression Discontinuity Designs in Economics. *Journal of Economic Literature,* 48, pp. 281-355. doi: 10.1257/jel.48.2.281
13. Liu, X., Nestic, D. and Vukina, T., 2012. Estimating Adverse Selection and Moral Hazard Effects with Hospital Invoices Data in a Government-Controlled Healthcare System. *Health Economics,* 21(8), pp. 883-901. doi: 10.1002/hec.1756
14. Paully, M., 1968. The Economics of Moral Hazard: Comment. *The American Economic Review,* 58(3), pp. 531-537.
15. Savage, E. and Wright, D. J., 2003. Moral Hazard and Adverse Selection in Australian Private Hospitals: 1989-1990. *Journal of Health Economics,* 22, pp. 331-359. doi: 10.1016/S0167-6296(02)00104-2

16. Schwartz, K. and Schwartz, T., 2008. Uninsured Young Adults: A Profile and Overview of Coverage Options. *The Kaiser Commission on Medicaid and the Uninsured* Issue Paper.
17. van der Klaauw, W., 2002. Estimating the Effect of Financial Aid Offers on College Enrollment: A Regression-Discontinuity Approach. *International Economic Review,* 43(4), pp. 1249-1287. doi: 10.1111/1468-2354.t01-1-00055
18. Vera-Hernandez, A. M., 1999. Duplicate Coverage and Demand for Health Care. The Case of Catalonia. *Health Economics,* 8, pp. 579-598. doi: 10.1002/(SICI)1099-1050(199911)8:7<579::AID-HEC478>3.0.CO;2-P
19. World Bank, 2015. Overview. Available at: [http://www.worldbank.org/en/country/croatia/overview].
20. Yörük, B. K., 2015. Health Insurance Coverage and Risky Health Behaviors among Young Adults. *CESIFO Working paper*, No. 5495. Available at: [https://ag-econ.ncsu.edu/wp-content/uploads/2015/11/FWB_2015_Final.pdf].