# Does Less Education Harm Health? Evidence from a Natural Experiment in a Developing Country

Khalil, Islam and Mallick, Debdulal and Nicholas, Aaron

Deakin University

January 2023

# Does Less Education Harm Health? Evidence from a Natural Experiment in a Developing Country*

**Islam Khalil** [a] **(ikhalil@deakin.edu.au)**
**Debdulal Mallick** [a, b] **(dmallic@deakin.edu.au)**
**Aaron Nicholas** [a] **(aaron.nicholas@deakin.edu.au)**

**First Draft: January 2023**
**This Draft: March 2023**

**Abstract**

We investigate the health outcome effects of a *reduction* in years of schooling in Egypt in 1988, a policy change that moves in the opposite direction in relation to the extant literature. We exploit this policy change as a natural experiment and employ a fuzzy regression discontinuity design to investigate a wide range of objectively measured health outcomes and behaviors. Despite the policy's adverse effect on students' years of schooling and ability to complete educational milestones, there is no effect on any health outcomes. Results also suggest that the reduction in years of schooling had no effect on labor market outcomes, thus providing a rationale for the lack of effect on health outcomes. Our results raise the possibility of a threshold model for understanding the relationship between education and health: while improvements in education may improve health, the inverse need not be true.

**Keywords:** Education; Health; Natural experiment; Fuzzy regression discontinuity.

**JEL Classification Codes:** I12; I20; C99.

[a] Department of Economics, Deakin University, 70 Elgar Road, Burwood, VIC 3125, Australia.
[b] Corresponding author.

*We would like to acknowledge helpful comments by Daniel Avdic, Cahit Guven, David Johnston, Wang Sheng Lee, Yan Liang, Iqbal A. Syed and the participants at the 8th Annual Workshop on Natural Experiments in History and the Brown Bag Seminar both at the Deakin Business School, and the 34th Ph.D. Conference in Economics and Business at Monash University. Usual disclaimers apply.

**Does Less Education Harm Health? Evidence from a Natural Experiment in a Developing Country**

## 1. Introduction

The education-health nexus is a policy-relevant subject that has long been an area of active research in economics and medical science. Although there is consensus about a positive correlation between education and health outcomes, evidence for a causal effect remains mixed. The literature has focused on exploring the beneficial effect of increasing years of compulsory schooling and is typically limited to the context of developed countries (see, Lochner, 2011 for a review and Hamad et al., 2018 for a meta-analysis). This paper, using a natural experiment in Egypt that involved lowering both the years of primary schooling and the total years of schooling,[1] investigates the causal effect of decreased education on several dimensions of objectively measured health outcomes and health related behaviors.

The effect of education on health is, a priori, ambiguous. On the one hand, a higher level of education can increase the efficiency with which health-related information is acquired and processed (e.g. eating nutritious food, seeking appropriate treatment when sick, purchasing health insurance) (Goldman and Smith, 2002; Glied and Lleras-Muney, 2003), contribute to optimal allocation of time in health-related behaviors (such as exercise) (Rosenzweig and Schultz, 1982) and also increase (life-time) income that might also positively affect health (Grossman, 1972).[2] On the other hand, higher education increases the opportunity cost of time spent on health-related activities such as physical exercise, and may also lead to sedentary (white-color, less manual) occupations that might result in less physical activities (Grossman, 1972). Individuals with less schooling can enter the job market at an earlier age and may also be more likely to select into occupations that require more physical activities. In addition, obtaining greater health knowledge may not be enough to improve health if adopting better practices is not affordable, especially in developing countries (Fang et al., 2012; Montez and Friedman, 2015). Therefore, the causal link between education and health, is likely to be ambiguous, especially in developing countries.

Empirical findings on the causal effect of increased education on health outcomes have been inconclusive. Studies that found positive effects of education include Lleras-Muney (2005) in

[1] By *years of schooling*, we mean *years of education* at all levels from primary school to university.

[2] Cutler and Lleras-Muney (2010), using the US and UK data, estimated that income, health insurance, and family background can account for about 30 percent, knowledge and measures of cognitive ability explain an additional 30 percent and social networks account for another 10 percent of the education gradient (the relationship between education and health behaviors).

lowering mortality (in USA), Powdthavee (2010) in reducing the likelihood of developing hypertensions (in UK), Brunello, Fabbri and Fort (2013) in reducing the likelihood of being overweight but only for women (in nine EU countries), and Huang (2015) on self-rated health, cognition, fitness and in lowering the likelihood of smoking (in China). In contrast, studies that found no effect of education include Braakmann (2011) and Clark and Royer (2013) on mortality, self-rated health, diabetes, smoking and alcohol consumption (in UK), Xie and Mo (2014) in self-rated health, hypertension, obesity and smoking (in China) and Ali and Elsayed (2018) in child health (in Egypt). Some studies also found mixed effects. For example, Dursun, Cesur and Mocan (2018) found positive effect on self-rated health and healthy weight among females, no impact on smoking but increased obesity among males (in Turkey). Bratti, Cottini and Ghinetti (2022) found positive effect on healthy eating, and physical activity and also in lowering obesity but also found increased prevalence of certain unhealthy behaviors, such as greater smoking and drinking prevalence and higher cigarettes consumption (in Italy).

Given the above ambiguity, it is likely that the effects of education on health vary across time and space. Health outcomes may be impacted differently depending on the initial level of education and the extent of change in the level of education. Although our paper is not intended to reconcile the conflicting results, it contributes to the extant literature in several important ways.

First, we investigate effects on health outcomes resulting from a *reduction* in years of schooling – a policy change that moves in the opposite direction in relation to the extant literature. We exploit an exogenous policy change in which the Egyptian government in 1988 reduced the length of primary schooling from six to five years. Second, assuming all students complete at least compulsory schooling, the policy affected the total years of schooling of *all* students. This contrasts with the literature where policies affect the years of compulsory schooling: students planning to complete more than compulsory schooling remain unaffected. Third, we consider a wide range of health outcomes which are objectively measured by the Demographic and Health Surveys (DHS) Program staff (funded by the United States Agency for International Development).[3] The surveys took measurements of individuals' biomarkers such as weight, height, and blood pressure, and collected blood specimens from each respondent to test for hepatitis and diabetes as well as collected information of self-reported health related behaviors such as smoking (See Section 4 for details). Finally, we study a developing country, a context

[3] Subjective health information may suffer from possible biases such as recall bias and self-anchoring effects (i.e., respondents evaluating their health status according to different benchmarks). Furthermore, when correlated with education, measurement error may lead to serious bias in education–health relationship. Dursun, Cesur and Mocan (2018) argue that highly educated women, probably more aware of the consequences, tend to misreport their body weight by answering survey questions in a way that others perceive favorably.

that is relatively unexplored.[4] Studying the effect of education on health in developing countries is crucial from a policy perspective because of their low level of both human capital and health outcomes.

Our study is based on a unique policy change in the schooling system in Egypt.[5] In Egypt, schooling rules state that children turning six years old before October shall attend primary school in the same year. In 1988, the government reduced the length of primary school from six to five years (effectively reducing compulsory education from nine to eight years and pre-university education from 12 to 11 years) because of public budget constraints. Therefore, individuals born in October 1977 comprise the first cohort to attend the 5-year primary school per the 1988 policy change. The identifying assumption is that, given the policy change, the conditional expectation of years of schooling is a smooth function with respect to the month-year of birth through the cut-off point in October 1977. Any discontinuity in years of schooling can then be attributed to the change in the length of schooling brought about by this policy change. Using this exogenous variation in years of schooling, we examine the causal effect of education on health outcomes. Our identification is based on a fuzzy regression discontinuity (RD) design in which total years of schooling is instrumented by the assignment rule into the treatment or control groups due to the exogenous policy change (see Section 3 for details).

We utilize data from two waves of large nationally representative Health Issues Surveys (HIS) conducted on behalf of the Ministry of Health and Population in 2008 and 2015. We find that the change in policy not only decreased total years of schooling but also the likelihood of attending secondary school (the new grade 8-11 after the policy change) significantly. However, this policy change more adversely affected years of schooling of males. Females were not significantly affected perhaps because of the counter-balancing effect resulting from concurrent increased attention to female education in developing countries including Egypt.

Our results on health outcomes are as follows: OLS regressions that do not account for the endogeneity of the years of schooling show some mixed effects on different health outcomes. Notably, additional years of schooling reduces the probability of being underweight and of being a smoker, while also increasing the probability of being overweight and of not having weight

[4] Xie and Mo (2014) and Huang (2015) in the context of China and Dursun, Cesur and Mocan (2018) in the context of Turkey are perhaps the only exceptions. However, these studies use self-rated health and/or self-reported weight and height information.

[5] Egypt fares very poorly in some health dimensions. Excluding some small nations, it ranks third in the world in obesity prevalence (28.9%) and first in obesity among females (37.5%). It also has the highest hepatitis C virus infection rate in the world (WHO, 2014; Elgharably et al., 2017).

within the 'normal' BMI range. However, once we account for endogeneity through a fuzzy RD design, we do not find any effect of lowering years of schooling on any of the health outcomes or health related behaviors (the likelihood of being: underweight, normal weight, overweight obese; of having hypertension, hepatitis, diabetes and of being a smoker). Our benchmark results are based on the optimal bandwidth for each outcome variable (Imbens and Lemieux, 2008; Jacob et al., 2012). The statistical insignificance of the years of schooling coefficient are broadly robust to reduced-form estimations, a wide range of alternative bandwidths, choice of control variables, and for different age groups based on separate analyses of 2008 and 2015 surveys. Most importantly, the results are robust in the male sub-sample for which the effect of the policy reduced the likelihood completing even the first year of secondary schooling. The statistical insignificance is not driven by a weak instrument, since the *F*-statistics for the restricted male sample are well above the cut-off value of 10.

One of the important channels through which education affects health is income. Hence, we supplement our findings with an analysis of Egypt's Labour Market Panel Surveys. Adopting a similar regression discontinuity design, we find that the reduction in years of schooling had no effect on wages and the probability of being employed, thus providing a rationale for the lack of effect on health outcomes.

Our results highlight a potential asymmetry: while improvements in education may improve health outcomes, the inverse need not be true. This suggests a threshold model understanding of the causal relationship of education on health. For example, in Egypt, if the initial level of education is below the threshold beyond which education has direct health benefits, reductions in education are unlikely to have health effects. This is consistent with the notion that there is less focus on science in education in Egypt. Foundations of medical knowledge that include topics such as human biology, the reproductive system, the immune system and causes of some diseases are taught in grades 10 and 11 as an optional subject, which are chosen by a small number of high-performing and motivated students who want to pursue a medical and science degree at university.

Our paper is methodologically related to Silles (2009), Powdthavee (2010) and Clark and Royer (2013) in the context of the UK, and Brunello, Fabbri and Fort (2013) in the context of the EU, where the fuzzy RD is used to estimate the causal effect of education on health. Lleras-Muney (2005) used the sharp RD in the context of the USA. Ali and Elsayed (2018) use the same Egyptian policy change as ours in an RD framework when investigating the effect of parental education on child health outcomes. They too find little to no effects and attribute it to the low level of education and inferior quality of primary schooling. There are also several studies that employed policy change as the IV (but not based on RD) to estimate a causal effect that include Braakmann (2011) in the context of the UK, Etilé and Jones (2011) in the context of France, Xie and Mo (2014) and Huang (2015) in the context of China, Dursun, Cesur and Mocan (2018) in the context of Turkey and Bratti, Cottini and Ghinetti (2022) in the context of Italy. Nonetheless, our paper differs in its contribution due to the nature of the policy change, the comprehensiveness of the health outcomes and the developing country context.

The rest of the paper proceeds as follows. In Section 2, we introduce the institutional background of the policy change in Egypt that form the natural experiment. In Section 3, we explain the identification and empirical strategy using this natural experiment. We describe the data and validity of the identifying assumptions in Sections 4 and 5, respectively. The results are discussed in Section 6. Finally, we conclude in Section 7.

## 2. Institutional background

Egypt's education system is the largest in the Middle East and North Africa region, with about 22.7 million students enrolled in pre-university education as of 2019.[6] The pre-university schooling comprises three levels spanning 12 years: primary school (grades 1–6), preparatory school (grades 7–9), and secondary school (grades 10–12). Primary and preparatory education are compulsory (totaling 9 years). Education is almost free in state-run institutions at all levels. Most students (91%) are enrolled in either public schools run by the government or private schools (these two are mainstream schooling). Both school systems fall under the supervision of, and follow the policy guidelines set by, the Ministry of Education. The rest are enrolled in either international or special religious schools, *Al-Azhar*, with their own system and curriculum (Elbadawy, 2015). After completing compulsory education, a student chooses between two tracks depending on their scores in preparatory school leaving exam at grade 9. The first track terminates in grade 9 with students moving to technical-vocational secondary school for three or five years, while the second track follows the standard secondary school (grade 12), a more prestigious track that often leads students to university.

Since the early 1980s, enrolment rates have increased substantially along with a public employment guarantee scheme for graduates and large-scale investments in infrastructure and school construction (Hansen and Radwan, 1982; Salehi-Isfahani, Hassine and Assaad, 2014).[7] In

[6] Among them 58% are in primary school and 42% are in secondary school (UNESCO Institute for Statistics (UIS) database, 2021). Source: http://sdg4-data.uis.unesco.org (Accessed on December 20, 2021).
[7] That was manifested in sharp increases in the literacy rate from one-third to two-thirds between 1976 and 2017 (UIS database, 2021).

1988, due to budget constraints compounded by the increase in enrolment rates, the Egyptian government reduced the length of primary schooling by one year from six to five years (hence, compulsory schooling reduced from nine to eight years and pre-university schooling reduced from 12 to 11 years). This decision was intended to meet the increasing enrolment rates and to reduce class sizes in over-crowded primary schools, some of which were running on multiple shifts. In October 1988, students who completed grade 6, along with those who completed grade 5, proceeded to grade 7. However, on eliminating grade 6, some adjustments were made in the curriculum to accommodate the new schooling system (Abdel-kareem, 2009). The old grade 6 curriculum was shortened and merged with the grade 4-5 curricula to develop a revised grade 4-5 curricula. Therefore, students leaving primary school by the end of the threshold school year (1988–1989) formed two groups—those who attended grade 5 learned a revised (shortened) version of the old grade 5-6 curricula, while those in grade 6 learned the original grade 6 curriculum.

Insert Table 1 here

The 1988 policy change was *reversed* in 2004 and the pre-1988 schooling system was re-introduced in response to criticisms of the earlier change as the government struggled to cope with international standards (Abdel-kareem, 2009). Table 1 summarizes the schooling systems under different policy regimes. We omit the effects of the second policy change (i.e. the reversal) in our analysis since most of the treated individuals are too young to have completed secondary schooling at the time of the latest survey (i.e. in 2015).[8]

## 3 Identification and estimation strategy

To investigate the causal effect of the reduction in years of schooling on health outcomes, we employ the Regression Discontinuity (RD) design using the 1988 policy change. In Egypt, schooling rules require that a child who turns six before October shall enroll in primary school in the same year. Therefore, individuals born before October 1977 attended a 6-year primary

[8] The first cohort affected by the policy reversal are those born in October 1993. Even if we employ a relatively shorter bandwidth of ±50 months (≈4 years), the cohorts included in the analysis would be those born between 1989 and 1997 (aged 18–26 years) at the time of the latest 2015 round of survey (see, Section 4 for details about the surveys). In this round of survey data, 34.4% complete secondary education by the age of 18 and another 15% complete by the age of 22.

school system, while those born on or after October 1977 attended a 5-year primary school system.

The identifying assumption is that, given the policy change, the conditional expectation of years of schooling is a smooth function with respect to the month-year of birth through the cut-off point in October 1977. Any discontinuity in years of schooling could be attributed to this policy change. We use month-year of birth as the *assignment* (alternatively termed *running* or *forcing*) variable determining the type of primary schooling (5-year vs. 6-year) an individual was enrolled in. However, law enforcement regarding compulsory schooling is not strict in Egypt; some individuals never attended primary school (Assaad, Levison and Dang, 2010). The rule that a child attends primary school at age six before October is also not strict in that some children may attend primary school even up to six months before the official school start age. Furthermore, the policy change did not affect students enrolled in international and special religious schools, which constitute 9% of the population (information on the type of school attended by an individual was not collected). In addition, if an individual born before October 1977 who was entitled to 6-year primary schooling repeated a class for any reason, they would now have 5-years of primary schooling because of the policy change. Given the possibility of overlap between the control and treatment groups, we utilize a fuzzy RD design.

The fuzzy RD design is based on instrumental variable (IV) estimation, where the cut-off in the assignment variable ($D_i$ = 1 after the policy change; 0 = otherwise) is the instrument for the (probability) of receipt of treatment. The following two equations summarize our estimation strategy.

$$S_i = \gamma_0 + \gamma_1 D_i + f(X_i) + \mathbf{Z}_i'\boldsymbol{\gamma}_2 + \omega_i \quad (1)$$

$$H_i = \beta_0 + \beta_1 \hat{S}_i + f(X_i) + \mathbf{Z}_i'\boldsymbol{\beta}_2 + \mu_i \quad (2)$$

where $H_i$ and $S_i$ represent a health outcome and years of total schooling of individual $i$ .

Equation (1) is the first-stage equation of the IV estimation. The assignment variable, $D_i$, in equation (1) is binary which equals 1 for an individual born on or after October 1977 (treatment group) and 0 otherwise (control group), which is the instrument for years of schooling in the first-stage regression. In equation (2), $\hat{S}_i$ is the fitted value of years of schooling obtained from the first-stage regression. $X_i$ is the re-centered assignment variable that is number of months an individual's month-year of birth deviates from the October 1977 cut-off (centered to be zero at the cut-off), and $f(X_i)$ is the local polynomial in $X_i$. $Z_i$ is a vector of control variables. At the individual level, it includes age and age-squared, marital status and gender. The random assignment of individuals into treatment and control groups due to the exogenous policy change should ensure that they are similar in all respects except their age (treatment members younger than control members). Given that health conditions depend on the age of an individual, which is also correlated with years of schooling by construction, the main control variable is age (and its square). Years of schooling also determines an individual's life-time income, which is a channel through which education might impact on health. We use an index for wealth at the household level at the time of the survey as a proxy for life-time income (Rutstein and Johnson, 2004).[9] Additionally, we include for a rural-urban dummy and dummies for regions to account for geographical variations.[10]

The coefficient of interest is $\beta_1$, which is the Local Average Treatment Effect (LATE) of years of schooling on health outcomes. The standard errors are clustered by season-year of birth (Lee and Card, 2008).[11]

An important methodological issue concerns deciding on the bandwidth size around the discontinuity before and after the cut-off date of October 1977, which involves balancing between precision and bias. A large bandwidth means a larger sample size and more precise estimates at the cost of increasing the bias. Using the cross-validation and plug-in procedures to choose the optimal bandwidth as suggested by, among others, Imbens and Lemieux (2008) and Jacob, Zhu, Somers and Bloom (2012), we obtain a bandwidth between ±39 and ±79 months (≈ 3.25 and 6.6 years) for different outcome variables. Our benchmark estimation involves using the optimal bandwidth for each outcome variable, although this leads to different sample size across the outcome variables. For robustness, we use two common bandwidths for all outcome variables which are the minimum and maximum in the range concerning all outcome variables (±40 and ±80 months). Finally, we employ nonparametric local polynomial regressions approximating $f(Xi)$ in quadratic forms.

[9] The wealth index is a composite measure of a household's cumulative living standard calculated by the DHS. The index is calculated by the principal component method using easy-to-collect data on i) a household's ownership of selected assets, such as televisions and bicycles, ii) materials used for housing construction, and iii) types of water access and sanitation facilities. The DHS reports the index as categorical into five quintiles. For details, see https://dhsprogram.com/topics/wealth-index/index.cfm. This index is highly correlated with measures of household wealth based on income and expenditure data (Filmer and Prichett, 2001; Rutstein, 1999).

[10] Given that the assignment variable is month-year of birth, we can only include either age or year of survey but not both, because either variable can be recovered by a linear combination of the other with the assignment variable.

[11] The results (not reported) remain strongly robust for alternative clustering, such as, at month-year, and at the primary sampling unit level (village/neighborhood in the rural/urban areas).

## 4. Data and descriptive statistics

The dataset comprises the two waves of the Health Issues Surveys (HIS) in 2008 and 2015 conducted on behalf of the Ministry of Health and Population of Egypt. The HIS is a special survey conducted to evaluate the prevalence of a number of chronic diseases such as diabetes and hepatitis.[12] The HIS surveyed a representative sample of the general population for those aged between 15 and 59 years, yielding a total of 12,892 males and 15,787 females across both surveys. The surveys collected information about individuals' biomarkers such as weight, height, and blood pressure, and collected blood specimens from each respondent to test for hepatitis and diabetes.[13] The health measurements were performed by the DHS staff.

The health indicators include height, weight, chronic diseases such as diabetes, hypertension, and hepatitis as well as self-reported health behaviors such as smoking. Using the measures of height and weight, we calculate BMI,[14] which is then classified into four categories—underweight (<18.5), normal (18.5–24.9), overweight (25.0–29.9) and obese (≥30) following the World Health Organization definition. The surveys also collected information about individuals' demographic characteristics, educational attainment, and month-year of birth necessary for this analysis.

Insert Table 2 here

Table 2 shows the descriptive statistics for the key variables for individuals born ±60 months around the cut-off.[15] There is no significant difference in gender composition (% female), household-level wealth index and percentage of individuals living in urban areas. A significant five year difference in age reflects the bandwidth of ±60 months around the policy cut-off and highlights the importance of controlling for age differences in the analysis. The significant

[12] HIS is part of the Demographic and Health Surveys (DHS) Program which is funded by the United States Agency for International Development (USAID) (https://dhsprogram.com/pubs/pdf/FR313/FR313.pdf). The DHS program typically carries out surveys that focus on maternal health for women of reproductive age (15-49 years). However, the Egyptian government and the DHS program made a special arrangement to implement the HIS to assess health problems that are widespread in Egypt – mainly hypertension, hepatitis C virus and diabetes.

[13] For the full testing protocols, see Ministry of Health and Population, Egypt (2015).

[14] BMI (Body Mass Index) = weight in kilograms/(height in meters^2).

[15] Since the optimal bandwidth varies across health outcomes, a common bandwidth (approximately the median of all optimal bandwidths) is used here to compare the observable characteristics of the treatment and control groups. The comparison is robust to alternative bandwidths.

difference in marital status can also be attributed to the age difference. These results also hold when the 2008 and 2015 surveys are analyzed independently (Appendix Table A.1).

**5. Validity of the identification strategy**

The effect of one-year reduction in the length of primary school on educational attainment is shown in Figure 1 by the distribution of total years of schooling averaged by month-year of birth for cohorts within ±60 months of the October 1977 cut-off corresponding to the 1988 policy. It shows a clear downward jump in average years of schooling for those born after October 1977.

Insert Figure 1 and Tables 3/3A/3B here

Table 3 presents the results of the regression of the policy cut-off, $D_i$, on years of schooling. On average, schooling fell by 0.7 (0.9) years as identified by the RD within the bandwidth of ±40 (±80) months, and this effect is statistically significant at the 1% level.

One concern, however, is that the reduction in the number of available years of primary schooling from six to five years was accompanied by a rearrangement of education curricula that might have resulted in little to minimal changes in actual educational content. This is particularly important in a developing country context, where the marginal returns to curriculum revision may be higher. If true, this would weaken our identification strategy.

We therefore consider if the policy impacted more than just the total years spent in school. From a purely mechanical perspective, reducing the number of available years of primary schooling reduces the total years spent in education without affecting the likelihood of completing education milestones such as preparatory or secondary school. However, if the policy actually impacted educational content, students may be less prepared for higher level studies and may subsequently be less likely to complete higher level milestones such as preparatory, secondary or even tertiary education.

We explore this possibility in columns (2) to (6), which show the effects of the policy on the likelihood of completing at least the first year of preparatory school all the way to the first year of tertiary education. We see that the strongest effects appear in columns (4) and (5): there is a significant reduction in the likelihood of completing at least the first year of secondary school, or secondary school as a whole (columns 4 and 5, 1% significance for both ±40 and ±80 months bandwidths).[16] The equivalent effects for preparatory and tertiary education are smaller in

[16] Similar result is also reported by Ali and Gurmu (2018).

magnitude and significant only at the ±80 months bandwidth. Similar patterns are observed when the sample is disaggregated according to gender (Tables 3A and 3B for male and female, respectively); however, effect sizes are notably smaller for females, and significant for only ±80 months bandwidths. Overall, there is evidence that the policy reduced the likelihood of students completing milestones in their education, thus supporting the notion that the reduction in total years of schooling in column (1) reflect an actual decline in educational quality, and this adverse effect is more pronounced for males.[17]

## 6. Results

Before examining any causal effect of the reduction in the years of schooling on health outcomes from the fuzzy RD design, we first provide a graphical analysis of any discontinuities in the health outcomes associated with the reduction in schooling due to the policy change.

Insert Figure 2 here

Figure 2 show discontinuities in some health outcomes. More specifically, the likelihood of being overweight increases but that of obesity and of not having weight within the 'normal' BMI range decreases. Regarding the incidences of chronic disease such as hypertension and diabetes there is an upward jump, but a downward jump in the incidence of hepatitis.

### 6.1 OLS results

There are several studies in the education-health literature based on OLS estimation.[18] While OLS estimation is severely limited in its ability to allow for causal inference,[19] we place our

---

[17] As mentioned earlier, after completing compulsory education, a student can choose either the technical-vocational secondary school or the standard secondary school that often leads students to university. We do not have information about whether students have switched between the tracks as a result of the policy.

[18] Examples include, among others, Meara, Richards and Cutler (2008), Miech and Shanahan (2000), Freedman, Martin, Schoeni, and Cornman (2008), Cutler and Lleras-Muney (2006) and Adjiwanou, Bougma, and Legrand (2018). These studies documented an association between education and different health outcomes such as depression, disability, heart disease, diabetes, days lost due to morbidity, modern contraceptive use and antenatal care among women.

[19] Both education and health are affected by individual heterogeneities or other socioeconomic factors related to family background (e.g., parental income or education) and time-discounting orientation (Fuchs, 1982), cognitive and non-cognitive skills (Carneiro, Crawford, and Goodman, 2007), genetics (Braveman, Egerter and Williams (2011)). Poor health in early life can result in low educational attainment and can be compounded if earlier health problems affect future health (Alderman, Behrman, Lavy and Menon, 2001; Case, Fertig and Paxson, 2005; Lochner, 2011).

paper in the context of this extant literature by also providing OLS estimates where the health outcome is regressed on years of schooling. These results also supplement the causal interpretation of the IV results with a broader picture based on correlations.

$$H_i = \alpha_0 + \alpha_1 S_i + f(X_i) + \mathbf{Z}_i'\boldsymbol{\alpha}_2 + \varepsilon_i . \qquad (3)$$

To allow for comparability with our benchmark IV results presented in the next section, we restrict the sample size to the one that is determined by the optimal bandwidth for each outcome variables, employ a quadratic polynomial in $f(X_i)$ and include all control variables. The OLS results are presented in Table 4. One additional (less) year of schooling decreases (increases) the likelihood of being underweight and smoker (both current and ever smoker). But it also increases the likelihood of being overweight, and decreases the likelihood of having weight within the 'normal' BMI range. These mixed results are also reported elsewhere in the literature (for review, see Lochner, 2011; Hamad et al., 2018). However, the magnitudes of the effects are very low. For example, the largest magnitude is the coefficient on smoking (both current and ever) at -0.008 suggesting that one additional (less) year of schooling decreases (increases) the likelihood of smoking by 0.8 percent.

When estimated separately for male and female sub-samples, there are some differences in the effect on different health outcomes. For example, one additional (less) year of schooling increases (decreases) the likelihood of obesity for males but decreases (increases) it for females, making the net effect null in the full sample. It decreases the likelihood of smoking only for males by approximately 2.1%. The coefficient for normal weight is significant for only males.

Insert Table 4 here

### 6.2 IV results using the fuzzy regression discontinuity

The IV results are based on estimations of equations (1) and (2). The benchmark results, reported in Table 5, use the optimal bandwidth for each health outcome, a quadratic polynomial in $f(X_i)$ and include all control variables.

We first discuss the first-stage regression results of instrumental validity specific to each health outcome investigated in the second-stage regressions (first-stage regression estimates differ across health outcomes because of different optimal bandwidths, and therefore sample size). The coefficients on the instrument are negatively significant at the 1% level in all first-stage regressions,[20] and the F-statistics are larger than the cut-off value of 10 (except for underweight and hepatitis for which these are still above 9), suggesting that the instrument is strongly correlated with years of schooling. Nonetheless, we employ Limited Information Maximum Likelihood (LIML) estimation for underweight and hepatitis, a variant of the 2SLS estimation, which provides robust inference with weak instruments (Stock, Wright and Yogo, 2002).

The second-stage results show that, for all health outcomes, the coefficients on (predicted) years of schooling are statistically insignificant at any conventional level (the largest of all t-statistics is 0.9 in absolute value). We do not discuss the sign of the coefficients (direction of the effect) because of very large confidence intervals. The results suggest that reduction in the years of schooling due to the policy change does not have any significant effect on the health outcomes under consideration.

To check if there were significant effects mediated by our controls (such as life-time income or other demographic characteristics), we exclude the control variables (except age and its square). The results, presented in Appendix Table A2, show that although the magnitudes of the coefficients for some outcome variables slightly increase, their statistical insignificances remain almost unaltered. The results are similar even excluding age and its square, and also using alternative combinations of the control variables (not reported). This suggests that the insignificance cannot be explained by the mediating role of any of the control variables, including wealth.

To check if the previous insignificant results are robust to alternative bandwidths, we now re-estimate using common bandwidths—minimum ±40 and maximum ±80 months (approximately ±3 and ±7 years, respectively). For the minimum bandwidth, the age cohort will be 28-41 years, and for the maximum bandwidth it will be 31-45 years.[21] The results estimated using all controls are presented in Table 6. Note that previously since optimal bandwidth varied across the health outcomes, the age cohorts also differed accordingly. Estimations using a common bandwidth also ensure the same age cohorts for all health outcomes. Although instrumental validity improves (the first-stage F-statistics further increase and become well above

[20] These first-stage regression coefficients correspond to results in the first columns in Tables 3/3A/3B. However, those results (not specific to any outcome variable) are based on different bandwidths, so the sample sizes were different.

[21] For the minimum bandwidth, the age cohort is 28-34 years [(2008 – 1977) ± 3 = 28-34] in the 2008 survey and 35-41 years [(2015 – 1977) ± 3] in the 2015 survey. So the range is 28-41 years. The maximum bandwidth, the only change in previous calculation is ± 7.

the cut-off value of 10 in all cases) for the maximum bandwidth, the insignificance of the effect of years of schooling in the second-stage regressions persists in both sets of regressions.[22]

Insert Tables 5-6 here

Since the change in the policy has less adversely affected females in their educational attainment, the insignificant results in the full sample may be driven by females. The results for the independent male and female samples using different combinations of the bandwidths are presented in Tables 7-10.

Insert Tables 7-10 here

In the male sub-sample with the optimal bandwidth, the F-statistics in the first-stage regressions are larger than the cut-off value of 10 for all health outcomes suggesting that the instrument is now strongly correlated with the endogenous regressor. However, there is no effect of the reduction in the years of schooling on any of the health outcomes in the second-stage results (Table 7). When re-estimated using the maximum bandwidth (80 months), only the coefficient on obesity becomes statistically significant at the 10% level. Quantitatively, obesity increases by approximately 5.1% for one more year of schooling (Table 8). This significance is perhaps because of the larger sample size from the wider bandwidth (which is supported by the results using the minimum bandwidth where the coefficient is again insignificant) but is also more prone to bias because of larger deviation from the cut-off point. Importantly, this significance is not robust to the choice of the control variables (Appendix Table A3).

On the other hand, in the female sub-sample using the optimal bandwidth, the F-statistics in the first-stage regressions are not only very low (at around 2), but also the coefficients on $D_i$ are statistically insignificant. The irrelevance of the instrument corroborates that the change in policy did not have effects on education for the female; therefore, no significant effect is expected in the second-stage regressions, which is also found in the estimated results (Table 9). The lack of significances persists for the maximum bandwidth as well (Table 10).

[22] We also estimated using several common bandwidths ranging between 40 and 110 (not reported). The insignificance persists.

### 6.3 Reduced form results

To supplement our core results based on the IV estimations, we estimate the reduced form equation in which the outcome variables are regressed directly on the instrument, $D_i$, which is shown in the equation below.

$$H_i = \delta_0 + \delta_1 D_i + f(X_i) + \mathbf{Z}_i'\boldsymbol{\delta}_2 + \eta_i . \qquad (4)$$

The reduced form estimates capture the overall effect that the change in the policy may have had on health. Notably, our IV estimation captures the effects of education on health only through changes in total years of schooling. The policy, however, may have had large impacts on students whose total years of schooling remain barely unaffected; for example the policy could have encouraged substitution from 'standard' secondary school to vocational secondary school. The reduced form estimates may pick-up overall effects of the policy that the more restrictive IV estimation was unable to. The reduced form estimate is more efficient than the IV estimate, so we expect greater statistical significance.

Since $D_i$ =1 is interpreted as a reduction in education, the reduced form coefficient should be interpreted in the opposite direction vis-à-vis the results from the IV estimation. The results show, as in the IV estimations, that there is no effect of the policy change on health outcomes using both the optimal bandwidth (Table 11) and the common minimum or maximum bandwidths (Appendix Table A5). In the male sub-sample, the insignificances hold for the optimal bandwidths (Table 12). When the maximum bandwidth is chosen (Table 13), the coefficient on obesity (= -0.055) becomes significant at the 10% level. The likelihood of being underweight also decreases by 1.2% (significant at the 10%) level. The comparable IV coefficient was very close at 0.011 but with a t-statistics of 1.57. In the female sub-sample, the change in policy increases the likelihood of being underweight and of having hepatitis (significant at the 10% level) (Appendix Table A6). However, (interestingly) that the effect disappears in both cases for the maximum bandwidth but persists in the case of underweight for the minimum bandwidth (Appendix Table A7). As mentioned earlier, in the female sub-sample, the policy did not have significant effect on education. Overall, the IV and reduced-form results are broadly similar.

Insert Tables 11-13 here

### 6.4 Separate estimation of 2008 and 2015 surveys and other checks

Due to the timing of the two surveys, relatively younger individuals were included in 2008 than in 2015. For example, for ±40 months bandwidth, the age cohorts in the 2008 and 2015 surveys are 28-34 and 35-41 years, respectively. Although the previous results are robust with and without controlling for age (and its square), separate analyses allow us to investigate the heterogeneous effects of education on health outcomes, if any, for different age groups. If there are opposing effects for different age groups, pooling data may cancel each other. In addition, the results for different age groups also compare medium- and long-run effects on health outcomes.

We estimate the benchmark specification using the optimal bandwidths (which differ across the two datasets and the pooled dataset), including all control variables. The results, presented in Tables 14 and 15, respectively, show that there is no effect of the policy on any of the health outcomes for either age cohorts. Given that now we deal with a smaller sample, it is not surprising that the estimates when male and female subsamples were estimated separately were insignificant (results not reported).

Insert Tables 14 and 15 here

We conduct several procedures to verify the insignificance of our estimated coefficients. We re-estimated the baseline results using different sample restrictions, such as excluding those with no formal education,[23] restricting the sample to individuals with at least primary education, and with at least secondary education. We also use alternative specifications such as a linear function in $f(X_i)$, triangular kernel estimation with larger weights for observations close to the cut-off and bias-corrected robust confidence interval suggested by Calonico, Cattaneo and Titiunik (2014). The pattern of no-impact on health outcomes remains unchanged.[24]

### 6.5 Exploring possible channels

While we find that the policy reduced education levels, primarily among male students, we find no significant causal effects of the reduction in years of schooling on any of the health outcomes and health related behaviors under consideration.

In the introduction, we discussed three possible channels through which education affects health outcomes: i) income, ii) knowledge that enhances (or lessens) the efficiency with which health-related information is acquired and processed, and iii) allocation of time in health-related behaviors. Among these, the income channel is arguably the most important. Hence,

[23] A sizeable segment of individuals in the HIS sample had no formal education (18%). Approximately 51.5% of those with no education were born before the cut-off of October 1977.

[24] These results are not reported (available from the authors upon request).

one of the primary candidates for a lack of effect of reduction in years of schooling (as induced by the policy) on health outcomes might be a lack of effect of the same on income. Although the effect of income on health outcomes is ambiguous for the reasons discussed in the introduction, we first explore whether the policy change had any impact on income.

The Health Issues Surveys (HIS), although containing rich and objective information about health outcomes, do not contain information about labor market outcomes. The latter information was collected in the Egypt Labour Market Panel Surveys (ELMPS), which were conducted in 2012 and 2018.[25] ELMPS are nationally representative surveys conducted by the Economic Research Forum in cooperation with the Central Agency for Public Mobilization and Statistics of the Egyptian government (OMADI, 2019a). The surveys collected information about employment, wages, birth (month-year) and schooling. However, since they do not collect data about health outcomes, the effect of education on both health and labor market outcomes cannot be estimated for the same individuals as the samples are different in the HIS and ELMPS. We therefore use the 2012 and 2018 ELMPS to investigate the effect of the 1988 policy on labor market outcomes alone.

In the sample, information about employment status is available for 35,905 individuals out of which 8.3% are unemployed.[26] However, information about the income of self-employed, employers, or unpaid family members (37% of the employed) was not reported. Therefore, income data consists of wage income only. Although most of the sample individuals had only one job, a small fraction of them (approximately 9%) had multiple (part-time) jobs. We use wage income from both primary job and all jobs, consisting of 19,423 individuals.

Our econometric approach is identical to the benchmark approach employed to estimate the effect of the policy on health outcomes. We alternatively use the minimum and maximum bandwidths of ±40 and ±80 months, as in the case of health outcomes.[27] Our outcome variables are employment status and the logarithm of monthly real wage.

The results for the full sample are presented in columns (1)-(3) in Table 16, and disaggregated by male and female in columns (4)-(6) and (7)-(9), respectively. The results for the first-stage regressions are similar to the ones obtained using the HIS data. Specifically, the

[25] The first two rounds of survey were conducted in 1988 and 1998, respectively; however, individuals affected by the policy change were too young to enter the job market at the time of the survey.

[26] In 2012 and 2018, 15,809 and 20,096 individuals were interviewed, respectively, out of which 13,888 individuals were interviewed in both rounds.

[27] The optimal bandwidths would now be different from those in the case of health outcomes, which also varied across health outcomes. Using the same minimum and maximum bandwidths ensures consistency.

coefficients on $D_i$ (the dummy for the policy) in all regressions of years of schooling are negative and statistically significant (at least at the 5% level) in the full and male sub-sample using both bandwidths but insignificant in the female sub-sample using the minimum bandwidth. The F-statistics are at least as large as the cut-off value of 10 (the strict cut-off value is 9.86) in both the full sample (except for employment as the outcome variable using the minimum bandwidth) and the male sub-sample but are much smaller than the cut-off value in the female sub-sample. These results corroborate the previous results using the HIS data that the policy has lowered years of schooling primarily for males. In all second-stage regressions—the full sample and gender-disaggregated sub-samples—the coefficients on the years of schooling, though positive, are insignificant. This suggests no effect of reducing the years of schooling on either wage income or the likelihood of being employed. The results are robust to using the minimum and maximum bandwidths as well as the alternative dataset from the Harmonized Labor Force Surveys (HLFS).[28] Assaad, Aydemir, Dayioglu and Kirdar (2016) found a negative, but very small relative to that in other developing countries, effect of the same policy on wage in Egypt. However, the authors used only the 2012 round of the ELMPS data.

Insert Table 16 here

One possible reason for the lack of effect of reduction in years of schooling on labor market outcomes may be that the policy was implemented nation-wide and hence employers wanting to hire from specific cohorts had no choice but to hire graduates with less years of schooling. There were no significant change in wages due to downward stickiness of nominal wages. However, exploration of the reasons for these findings is beyond the scope of the objective of this paper.[29]

---

[28]The Harmonized Labor Force Surveys (HLFS) (2006-2017) conducted by Egypt's Central Agency for Public Mobilization and Statistics (OMADI, 2019b) constitute the other primary source of data on the labor market in Egypt. Unfortunately these surveys do not collect information about month-year of birth to accurately identify the cut-off due to the education policy change in 1988. Nonetheless, using age information (which was not always accurately reported and also recorded only as an integer), we calculated the birth year, and conducted the same exercise using six rounds of survey data (2007-09 and 2014-16) for the comparable HIS periods (2008 and 2015). The results are qualitatively similar to what we obtain in the ELMPS data. However, the first-stage F-statistics are much smaller than the cut-off value of 10 even in the full and male sub-sample, which are necessary to ensure the validity of the instrument. Therefore, we do not report these results.

[29] There is a large literature on the effect of education on labor market outcomes. See Harmon, Oosterbeek and Walker (2003) and Dickson and Harmon (2011) for nice surveys and the references therein.

Another rationale for our core results arises as a counterargument against the knowledge channel: less education need not translate into overall poorer learning (Elbadawy, 2014). If the existing education was already of poor quality,[30] the change in policy had minimal further negative effects on learning. Assaad, Aydemir, Dayioglu and Kirdar (2016) and Ali and Elsayed (2018) also found no effect of this policy change on literacy. However, we detected adverse effects of the policy on non-primary educational milestones (e.g. completing the first year of secondary schooling), suggesting that there were some negative effects on overall learning.

We propose a narrower, albeit related, explanation: the reduction in education was not associated with a reduction of specific knowledge about human health and causes of diseases. In Egypt, foundations of medical knowledge that includes topics such as human biology, the reproductive system, the immune system, and causes of some diseases are introduced in grade 10 and 11 (grade 11 and 12 before the policy change in 1988 and after the policy reversal in 2004).[31] Biology, in which these topics are included, is an optional subject in high school and only high-performing students who plan to pursue a medical and science degree at the university level choose to study this subject. Although the change in policy decreased the likelihood of completing high school in general, it is highly likely that these high-performing students were not among those affected. Therefore, the results can be attributed to a weak initial base of health-related education in Egypt, which had little scope to be affected by the change in policy.

## 7. Conclusion

In this paper, we investigate the effect of an unconventional education policy change in Egypt in 1988 that lowered the length of primary schooling by one year. As primary school is technically compulsory in Egypt, this resulted in widespread reduction in the total years of schooling, particularly for males. By exploiting this exogenous policy change as a natural experiment, we employ a fuzzy RD to estimate the effect on a multitude of objectively measured health outcomes and health behaviors in later stages of life. We do not find any effect on the health outcomes and health behaviors under consideration. The results are consistent across different specifications and multiple robustness checks. A likely explanation can be the lack of

[30] In 2017, Egypt ranked 129 out of 137 in education quality (source: World Economic Forum, The Global Competitiveness Report 2017--2018). Source: https://www3.weforum.org/docs/GCR2017-2018/05FullReport/TheGlobalCompetitivenessReport2017%E2%80%932018.pdf (Accessed on January 25, 2022).

[31] Source: https://science3sec.moe.gov.eg/#p=344 (p. 313-344), Concepts of Biology for Third-grade Secondary School, Ministry of Education [Egypt].

effects on labor market outcomes as a result of the reduction in years of schooling induced by the policy, and the lack of broad-access education on health and diseases.

These results should be generalized with caution, as the policy change resulted in a very specific variation to education. The impacts on health, for example, could have been significant if the policy reduced *secondary* education, or *increased* primary education. More research is needed to better understand the education-health nexus in developing countries.

**References**

Abdel-kareem, Noha (2009), “Sun'a Alqarar fi Assyasat Al-taelimiah: Al'atraf Al-faeilah Walaliat [Decision-Making in Educational Policy: Actors and Mechanisms],” Al-Dar Al-Masriah Al-Lubnaniah, Cairo, Egypt.

Adjiwanou, Vissého, Moussa Bougma and Thomas LeGrand (2018), “The Effect of Partners' Education on Women's Reproductive and Maternal Health in Developing Countries,” *Social Science & Medicine*, 197, 104-115.

Alderman, Harold, Jere R Behrman, Victor Lavy and Rekha Menon (2001), “Child Health and School Enrollment: A Longitudinal Analysis,” *Journal of Human Resources*, 36 (1), 185-205.

Ali, Fatma Romeh M. and Mahmoud A. A. Elsayed (2018), “The Effect of Parental Education on Child Health: Quasi-Experimental Evidence from a Reduction in the Length of Primary Schooling in Egypt,” *Health Economics*, 27 (4), 649-662.

Ali, Fatma Romeh M and Shiferaw Gurmu (2018), “The Impact of Female Education on Fertility: A Natural Experiment from Egypt,” *Review of Economics of the Household,* 16 (3), 681-712.

Assaad, Ragui, Abdurrahman Aydemir, Meltem Dayioglu and Guray Kirdar (2016), “Returns to Schooling in Egypt,” Economic Research Forum Working Paper No. 1000, Cairo, Egypt.

Assaad, Ragui, Deborah Levison and Hai-Anh Dang (2010), “How Much Work is too Much? Effects of Child Work Hours on Schooling–the Case of Egypt,” in Akee, Randall, Eric Edmonds and Kostantinos Tatsiramos, eds., Child Labor and the Transition between School and Work, 53-97, Emerald, Bingley.

Behrman, Jere R, Hans-Peter Kohler, Vibeke Myrup Jensen, Dorthe Pedersen, Inge Petersen, Paul Bingley and Kaare Christensen (2011), “Does More Schooling Reduce Hospitalization and Delay Mortality? New Evidence Based on Danish Twins,” *Demography*, 48 (4), 1347-1375.

Braakmann, Nils (2011), “The Causal Relationship between Education, Health and Health Related Behaviour: Evidence from a Natural Experiment in England,” *Journal of Health Economics*, 30 (4), 753-763.

Bratti, Massimilano, Elena Cottini and Paolo Ghinetti (2022), “Education, Health and Health-Related Behaviors: Evidence from Higher Education Expansion,” Institute for the Study of Labour (IZA) Working Paper No. 15035, Bonn, Germany.

Braveman, Paula, Susan Egerter and David R Williams (2011), “The Social Determinants of Health: Coming of Age,” *Annual Review of Public Health*, 32 (1), 381-398.

Brunello, Giorgio, Daniele Fabbri and Margherita Fort (2013), “The Causal Effect of Education on Body Mass: Evidence from Europe,” *Journal of Labor Economics,* 31 (1), 195-223.

Calonico, Sebastian, Matias D. Cattaneo and Rocio Titiunik (2014), "Robust Nonparametric Confidence Intervals for Regression-Discontinuity Designs," *Econometrica*, 82 (6), 2295-2326.

Carneiro, Pedro, Claire Crawford and Alissa Goodman (2007), "The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes," Centre for the Economics of Education, LSE, London, UK.

Case, Anne, Angela Fertig and Christina Paxson (2005), "The Lasting Impact of Childhood Health and Circumstance," *Journal of Health Economics*, 24 (2), 365-389.

Clark, Damon and Heather Royer (2013), "The Effect of Education on Adult Mortality and Health: Evidence from Britain," *American Economic Review,* 103 (6), 2087-2120.

Cutler, David M and Adriana Lleras-Muney (2010), "Understanding Differences in Health Behaviors by Education," *Journal of Health Economics,* 29 (1), 1-28.

Dickson, Matt and Colm Harmon (2011), "Economic returns to education: what we know, what we don't know, and where we are going—some brief pointers," *Economics of Education Review*, 30 (6), 1118-1122.

Dursun, Bahadır, Resul Cesur and Naci Mocan (2018), "The Impact of Education on Health Outcomes and Behaviors in a Middle-Income, Low-Education Country," *Economics & Human Biology,* 31, 94-114.

Elbadawy, Asmaa (2014), "Education in Egypt: Improvements in Attainment, Problems with Quality and Inequality," Economic Research Forum Working Paper No. 854, Cairo, Egypt.

Elgharably, Ahmed, Asmaa I Gomaa, Mary ME Crossey, Peter J Norsworthy, Imam Waked and Simon D Taylor-Robinson (2017), "Hepatitis C in Egypt–Past, Present, and Future," *International Journal of General Medicine,* 10, 1-6.

Etilé, Fabrice and Andrew M Jones (2011), "Schooling and Smoking among the Baby Boomers–An Evaluation of the Impact of Educational Expansion in France," *Journal of Health Economics,* 30 (4), 811-831.

Fang, Hai, Karen N Eggleston, John A Rizzo, Scott Rozelle and Richard J Zeckhauser (2012), "The Returns to Education in China: Evidence from the 1986 Compulsory Education Law," NBER Working Paper No. 18189, Cambridge, MA.

Filmer, Deon and Lant H Pritchett (2001), "Estimating Wealth Effects without Expenditure Data—Or Tears: An Application to Educational Enrollments in States of India," *Demography,* 38 (1), 115-132.

Freedman, Vicki A, Linda G Martin, Robert F Schoeni and Jennifer C Cornman (2008), "Declines in Late-Life Disability: The Role of Early-and Mid-Life Factors," *Social Science & Medicine*, 66 (7), 1588-1602.

Fuchs, Victor R (1982), "Time Preference and Health: An Exploratory Study," in Fuchs, Victor R., ed., Economic Aspects of Health, 93-120, University of Chicago Press, Chicago.

Glied, Sherry and Adriana Lleras-Muney (2003), "Health Inequality, Education and Medical Innovation," NBER Working Paper No. 9738, Cambridge, MA.

Goldman, Dana P and James P Smith (2002), "Can Patient Self-Management Help Explain the SES Health Gradient?" *Proceedings of the National Academy of Sciences*, 99 (16),10929-10934.

Grossman, Michael (1972), "On the Concept of Health Capital and the Demand for Health," *Journal of Political Economy*, 80 (2), 223-255.

Hamad, Rita, Holly Elser, Duy C Tran, David H Rehkopf and Steven N Goodman (2018), "How and Why Studies Disagree about the Effects of Education on Health: A Systematic Review and Meta-Analysis of Studies of Compulsory Schooling Laws," *Social Science & Medicine,* 212, 168-178.

Harmon, Colm, Hessel Oosterbeek and Ian Walker (2003), "The returns to education: Microeconomics," *Journal of Economic Surveys*, 17 (2), 115-156.

Hansen, Bent and Samir Radwan (1982), "Employment Planning in Egypt: An Insurance Policy for the Future," *International Labour Review,* 121, 535.

Huang, Wei (2015), "Understanding the Effects of Education on Health: Evidence from China," Institute for the Study of Labour (IZA) Working Paper No. 9225, Bonn, Germany.

Imbens, Guido W and Thomas Lemieux (2008), "Regression Discontinuity Designs: A Guide to Practice," *Journal of Econometrics,* 142 (2), 615-635.

Jacob, Robin, Pei Zhu, Marie-Andrée Somers and Howard Bloom (2012), "A Practical Guide to Regression Discontinuity," MDRC, New York.

Lee, David S and David Card (2008), "Regression Discontinuity Inference with Specification Error," *Journal of Econometrics*, 142 (2), 655-674.

Lleras-Muney, Adriana (2005), "The Relationship between Education and Adult Mortality in the United States," *Review of Economic Studies*, 72 (1), 189-221.

Lochner, Lance (2011), "Nonproduction Benefits of Education: Crime, Health, and Good Citizenship," in Wolter, Stefan C. and Paul Ryan, eds., Handbook of the Economics of Education, 183-282, Elsevier, Amsterdam, The Netherlands.

Meara, Ellen R, Seth Richards and David M Cutler (2008), "The Gap Gets Bigger: Changes in Mortality and Life Expectancy, by Education, 1981–2000," *Health Affairs,* 27 (2), 350-360.

Miech, Richard Allen and Michael J Shanahan (2000), "Socioeconomic Status and Depression over the Life Course," *Journal of Health and Social Behavior*, 41 (2), 162-176.

Ministry of Health and Population [Egypt], El-Zanaty and Associates [Egypt] and ICF International (2015), Egypt Demographic and Health Survey 2014, (Cairo, Egypt; Rockville, Maryland, USA).

Montez, Jennifer Karas and Esther M Friedman (2015), "Educational Attainment and Adult Health: Under What Conditions Is the Association Causal?" *Social Science & Medicine*, 127, 1-7.

OAMDI, 2019a. Egypt Labor Market Panel Surveys (ELMPS), http://erf.org.eg/data-portal/, Version 2.0 of Licensed Data Files; ELMPS 2018. Egypt: Economic Research Forum.

OAMDI, 2019b. Harmonized Labor Force Surveys (HLFS), http://erf.org.eg/data-portal/, Version 1.0 of Licensed Data Files; LFS 2017- Central Agency for Public Mobilization and Statistics, Egypt: Economic Research Forum.

Powdthavee, Nattavudh (2010), "Does Education Reduce the Risk of Hypertension? Estimating the Biomarker Effect of Compulsory Schooling in England," *Journal of Human Capital*, 4 (2), 173-202.

Rosenzweig, Mark R and T Paul Schultz (1982), "The Behavior of Mothers as Inputs to Child Health: The Determinants of Birth Weight, Gestation, and Rate of Fetal Growth," in Fuchs, Victor R., ed., Economic Aspects of Health, 53-92. University of Chicago Press, Chicago.

Rutstein, Shea (1999), "Wealth Versus Expenditure: Comparison between the DHS Wealth Index and Household Expenditures in Four Departments of Guatemala," ORC Macro, Calverton, Maryland.

Rutstein, Shea O and Kiersten Johnson (2004), "The DHS Wealth Index. DHS Comparative Reports No. 6," ORC Macro, Calverton, Maryland.

Salehi-Isfahani, Djavad, Nadia Belhaj Hassine and Ragui Assaad (2014), "Equality of Opportunity in Educational Achievement in the Middle East and North Africa," *The Journal of Economic Inequality,* 12 (4), 489-515.

Silles, Mary A. (2009), "The Causal Effect of Education on Health: Evidence from the United Kingdom," *Economics of Education Review*, 28 (1), 122-128.

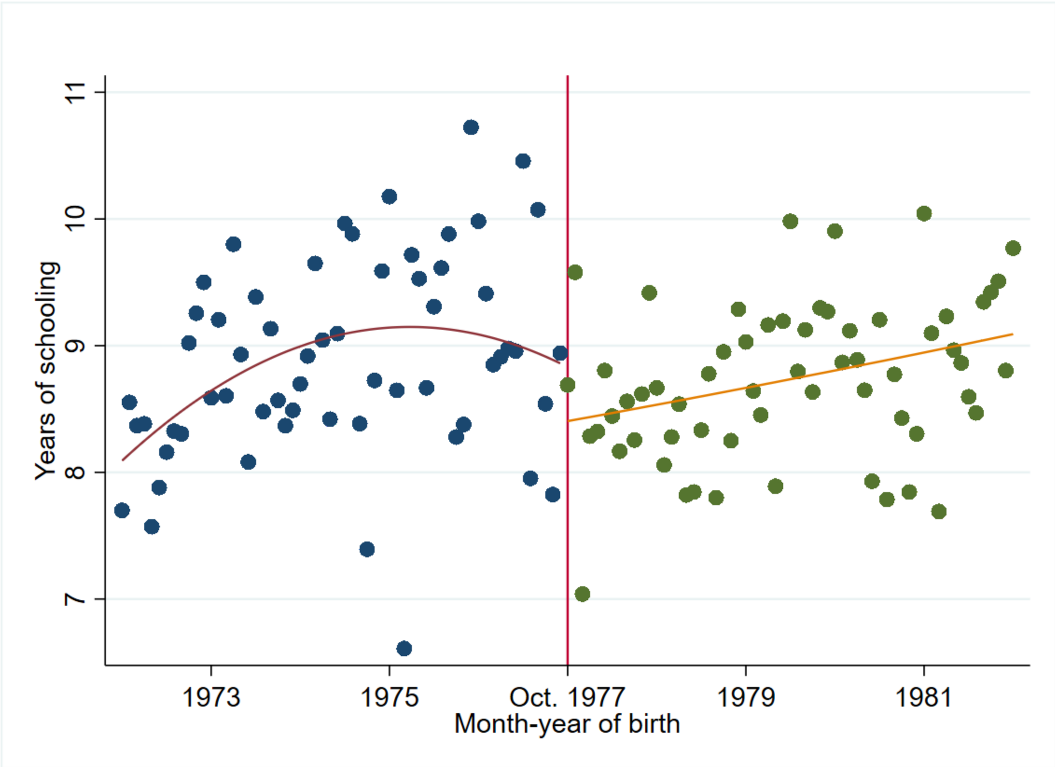
Stock, James H., Jonathan H. Wright, and Motohiro Yogo (2002), "A Survey of Weak Instruments and Weak Identification in Generalized Method of Moments," *Journal of Business & Economic Statistics,* 20 (4), 518-529.

WHO, World Health Organization (2014), Global Status Report on Noncommunicable Diseases 2014, World Health Organization.

Xie, Shiqing and Taiping Mo (2014), "The Impact of Education on Health in China," *China Economic Review,* 29, 1-18.

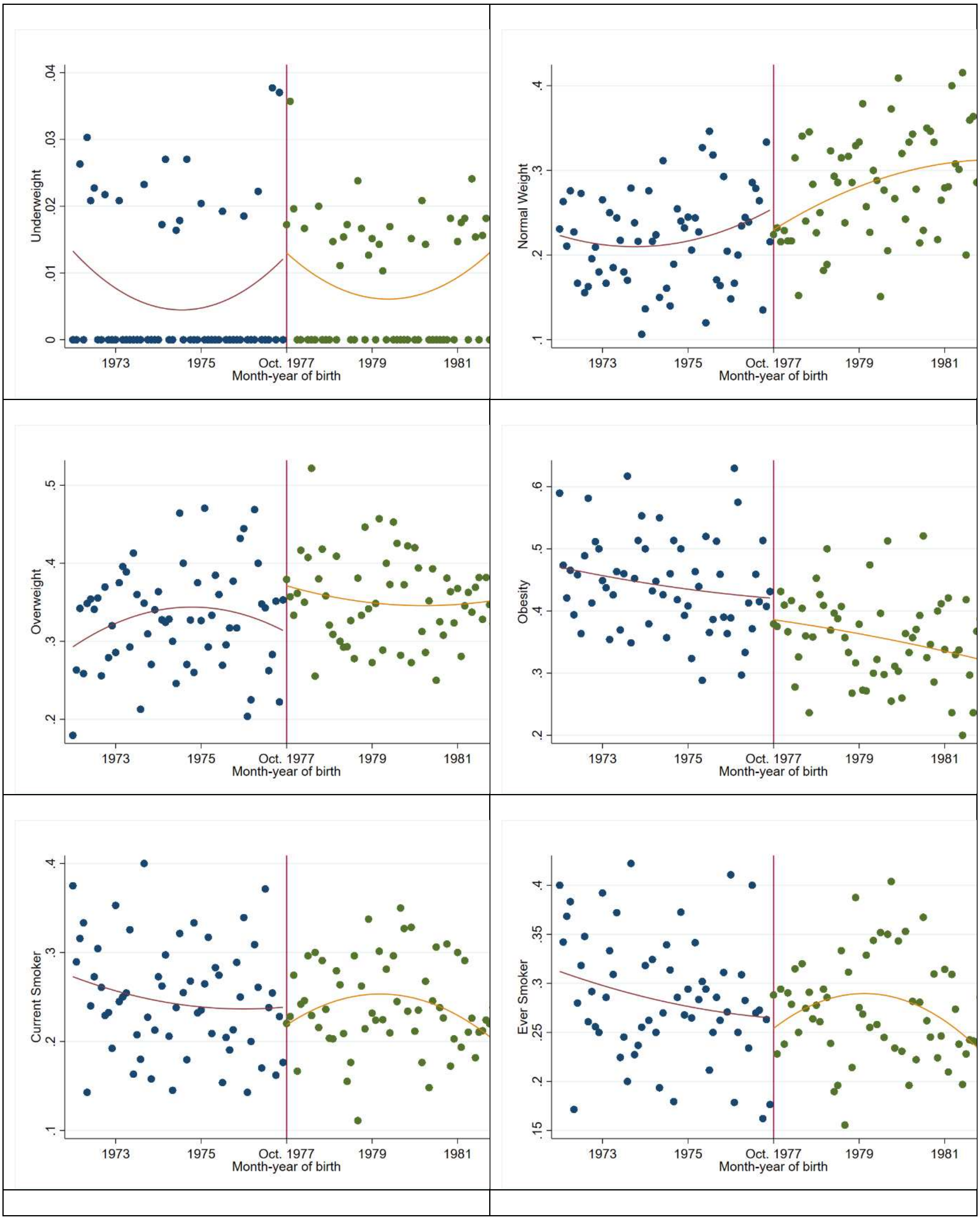
## Figures and Tables

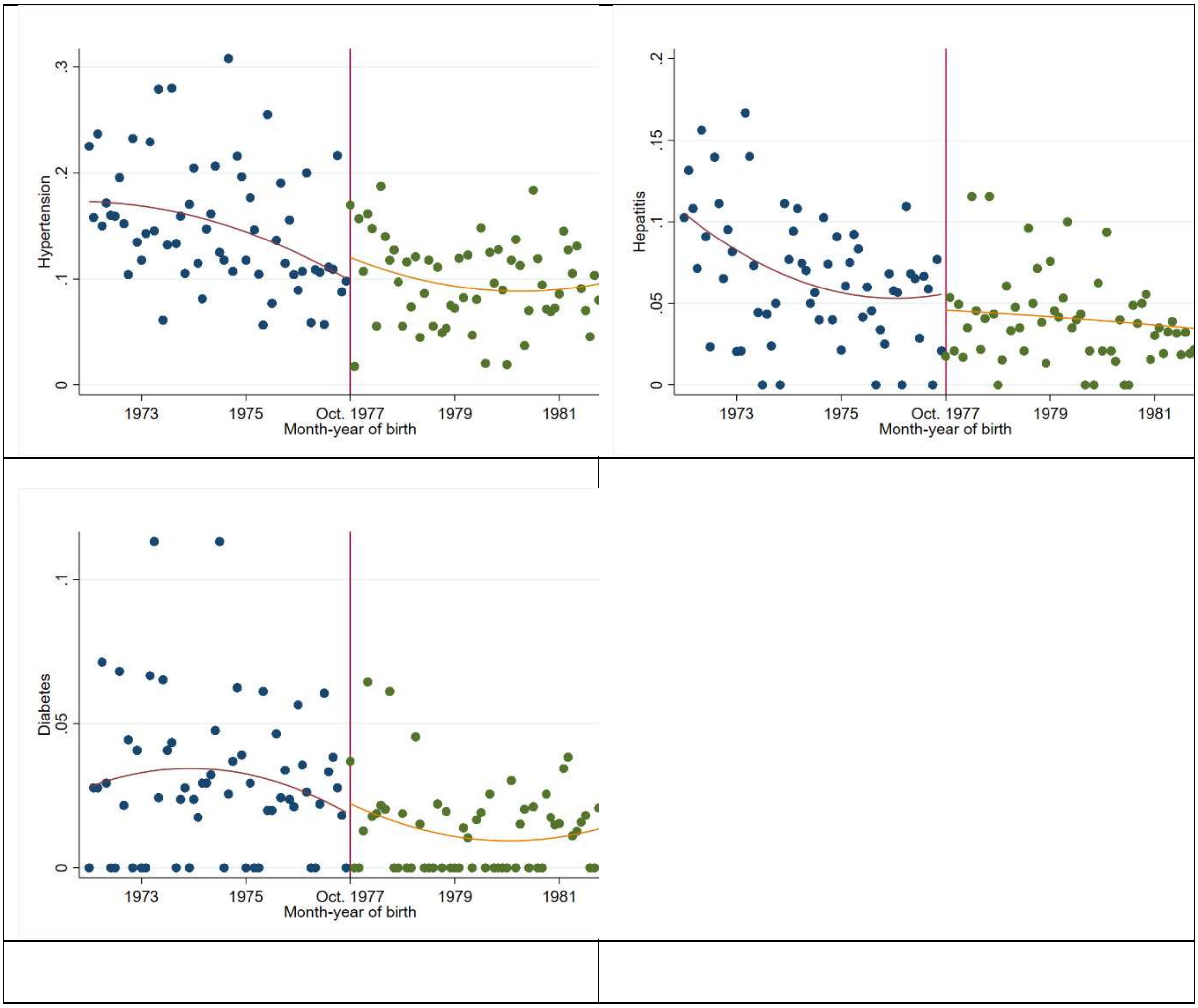
Figure 1: Discontinuity in total years of schooling




*Note:* A dot represents the average for those born in a specific month-year. A linear polynomial was fitted on each side of the cut-off point and 60 months bandwidth is used.

Figure 2: The discontinuity in health outcomes and health behaviors






*Note:* A dot represents the average for those born in a specific month-year. A linear polynomial was fitted on each side of the cut-off point and 60 months bandwidth is used.

Table 1: Length of schooling under different policy regimes.

| **Sequence of schooling (years)** | **Pre-1988 and post-2003 (sequence of years)*** | **1989-2003 (sequence of years)**** |
|---|---|---|
| Primary (compulsory) | 1-6 | 1-5 |
| Preparatory (compulsory) | 7-9 | 6-8 |
| Secondary | 10-12 | 9-11 |
| *Total* | *12* | *11* |
| | | |

*Before the policy change in 1988 and after the policy reversal in 2004.
** Period during which the 1988 policy change was effective.

Table 2: Descriptive statistics

| | Total | | | Female | | | Male | | |
|---|---|---|---|---|---|---|---|---|---|
| | All | Control | Treatment | All | Control | Treatment | All | Control | Treatment |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| Age | 33.116 | 36.499 | 31.436*** | 32.903 (12.286) | 36.436 | 31.300*** | 33.378 | 36.577 | 31.606 |
| Gender (% Female) | 0.550 | 0.550 | 0.556 | ---- | ---- | ---- | ---- | ---- | ---- |
| Married (%) | 0.654 | 0.908 | 0.844*** | 0.689 | 0.899 | 0.884 | 0.612 | 0.919 | 0.793*** |
| Urban (%) | 0.463 | 0.466 | 0.466 | 0.455 | 0.469 | 0.454 | 0.472 | 0.463 | 0.480 |
| Wealth index | 3.096 | 3.124 | 3.161 | 3.078 | 3.089 | 3.137 | 3.118 | 3.166 | 3.191 |
| N | 28637 | 2991 | 3736 | 15761 | 1645 | 2075 | 12876 | 1346 | 1661 |

*** (**) [*] Difference between treatment and control groups significant at the 1% (5%) [10%] level. Columns 1, 4, and 7 consider all observations in the 2008 and 2015 surveys. Observations in other columns are those within 60 months bandwidth.

Table 3: The effect of the policy change on educational attainment

| | Years of schooling | Completed at least 7 (6) years of schooling[a] | Completed at least 9 (8) years of schooling[b] | Completed at least 10 (9) years of schooling[c] | Completed at least 12 (11) years of schooling[d] | Completed at least 13 (12) years of schooling[e] |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 40 months bw | | | | | | |
| $D_i$ | -0.702*** | -0.003 | -0.033 | -0.056*** | -0.059*** | -0.024 |
| | (0.228) | (0.020) | (0.026) | (0.019) | (0.020) | (0.015) |
| | | | | | | |
| $R^2$ | 0.211 | 0.141 | 0.143 | 0.145 | 0.143 | 0.100 |
| N | 4511 | 4511 | 4511 | 4511 | 4511 | 4511 |
| | | | | | | |
| 80 months bw | | | | | | |
| $D_i$ | -0.903*** | -0.012 | -0.032* | -0.048*** | -0.056*** | -0.031** |
| | (0.176) | (0.015) | (0.017) | (0.014) | (0.015) | (0.013) |
| | | | | | | |
| $R^2$ | 0.215 | 0.144 | 0.148 | 0.154 | 0.153 | 0.109 |
| N | 8971 | 8971 | 8971 | 8971 | 8971 | 8971 |

[a]Year 7(6) is the first year of preparatory schooling before (after) the policy change. [b]Year 9(8) is the final year of compulsory schooling before (after) the policy change; [c]Year 10(9) is the first year of secondary school before (after) the policy change; [d]12(11) is the total years of pre-university schooling before (after) the policy change. [e]Year 13(12) is the first year of tertiary (university) education before (after) the policy change.

$D_i$ =1 if treatment (born on or after October 1977), = 0 if control (born before October 1977).

Estimation of equation (1) using non-parametric local quadratic polynomial regressions. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses clustered at season and year-of-birth. *** p<0.01, ** p<0.05, * p<0.10.

Table 3A: The effect of the policy change on educational attainment (Male)

| | Years of schooling | Completed at least 7 (6) years of schooling | Completed at least 9 (8) years of schooling | Completed at least 10 (9) years of schooling | Completed at least 12 (11) years of schooling | Completed at least 13 (12) years of schooling |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 40 months bw | | | | | | |
| $D_i$ | -1.102*** | -0.014 | -0.033 | -0.067* | -0.070* | -0.048* |
| | (0.243) | (0.031) | (0.041) | (0.036) | (0.038) | (0.027) |
| | | | | | | |
| $R^2$ | 0.161 | 0.087 | 0.095 | 0.100 | 0.096 | 0.094 |
| N | 2047 | 2041 | 2047 | 2047 | 2047 | 2047 |
| | | | | | | |
| 80 months bw | | | | | | |
| $D_i$ | -1.156*** | -0.008 | -0.019 | -0.056* | -0.061** | -0.048** |
| | (0.216) | (0.023) | (0.028) | (0.028) | (0.028) | (0.021) |
| | | | | | | |
| $R^2$ | 0.152 | 0.082 | 0.093 | 0.097 | 0.097 | 0.097 |
| N | 3981 | 3981 | 3981 | 3981 | 3981 | 3981 |

See Table 3 for notes.

Table 3B: The effect of the policy change on educational attainment (Female)

| | Years of schooling | Completed at least 7 (6) years of schooling | Completed at least 9 (8) years of schooling | Completed at least 10 (9) years of schooling | Completed at least 12 (11) years of schooling | Completed at least 13 (12) years of schooling |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| 40 months bw | | | | | | |
| $D_i$ | -0.282 | 0.019 | -0.017 | -0.037 | -0.040 | -0.009 |
| | (0.370) | (0.031) | (0.031) | (0.030) | (0.031) | (0.018) |
| | | | | | | |
| $R^2$ | 0.229 | 0.168 | 0.172 | 0.171 | 0.171 | 0.110 |
| N | 2464 | 2464 | 2464 | 2464 | 2464 | 2464 |
| | | | | | | |
| 80 months bw | | | | | | |
| $D_i$ | -0.645*** | -0.011 | -0.035* | -0.037** | -0.047** | -0.017 |
| | (0.234) | (0.020) | (0.019) | (0.019) | (0.019) | (0.015) |
| | | | | | | |
| $R^2$ | 0.238 | 0.172 | 0.182 | 0.186 | 0.186 | 0.124 |
| N | 4990 | 4990 | 4990 | 4990 | 4990 | 4990 |

See Table 3 for notes.

Table 4: OLS Estimation: The association between education and health.

| | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
|---|---|---|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
| **All sample** | | | | | | | | | |
| Years of Schooling | -0.001** | -0.003*** | 0.004*** | 0.000 | -0.009*** | -0.008*** | -0.002 | -0.001 | -0.000 |
| | (0.000) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.001) | (0.000) |
| Bandwidth | 48.80 | 68.81 | 66.99 | 60.13 | 71.07 | 61.08 | 50.96 | 83.23 | 73.13 |
| Mean of outcome | 0.007 | 0.261 | 0.343 | 0.394 | 0.241 | 0.278 | 0.118 | 0.052 | 0.020 |
| $R^2$ | 0.016 | 0.102 | 0.026 | 0.144 | 0.392 | 0.417 | 0.033 | 0.023 | 0.018 |
| N | 5212 | 7386 | 7218 | 6546 | 7918 | 6812 | 5559 | 8836 | 7806 |
| **Males** | | | | | | | | | |
| Years of Schooling | -0.000 | -0.011*** | 0.005** | 0.006*** | -0.021*** | -0.021*** | -0.000 | -0.001 | 0.001 |
| | (0.000) | (0.002) | (0.002) | (0.002) | (0.003) | (0.002) | (0.002) | (0.001) | (0.000) |
| Bandwidth | 57.77 | 61.93 | 60.24 | 71.03 | 47.53 | 54.69 | 61.31 | 56.66 | 72.19 |
| Mean of outcome | 0.008 | 0.335 | 0.398 | 0.254 | 0.533 | 0.593 | 0.122 | 0.061 | 0.019 |
| $R^2$ | 0.021 | 0.076 | 0.021 | 0.061 | 0.062 | 0.079 | 0.032 | 0.027 | 0.020 |
| N | 2791 | 2969 | 2922 | 3422 | 2339 | 2712 | 3053 | 2684 | 3450 |
| | | | | | | | | | |
| **Females** | | | | | | | | | |
| Years of Schooling | -0.001** | 0.000 | 0.005** | -0.004** | 0.000 | 0.000 | -0.002** | 0.000 | -0.001 |
| | (0.000) | (0.002) | (0.002) | (0.002) | (0.000) | (0.001) | (0.001) | (0.001) | (0.001) |
| Bandwidth | 45.58 | 50.27 | 52.63 | 55.74 | 79.04 | 68.82 | 61.67 | 52.35 | 67.45 |
| Mean of outcome | 0.007 | 0.189 | 0.296 | 0.505 | 0.006 | 0.023 | 0.116 | 0.043 | 0.021 |
| $R^2$ | 0.030 | 0.080 | 0.032 | 0.121 | 0.012 | 0.044 | 0.042 | 0.025 | 0.022 |
| N | 2697 | 2987 | 3134 | 3310 | 4919 | 4192 | 3759 | 3037 | 3924 |
| | | | | | | | | | |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses clustered at season and year-of-birth. Robust standard errors in parentheses are clustered at year-of-birth by season-of-birth. *** p<0.01, ** p<0.05, * p<0.10.

Table 5: IV estimation using fuzzy RD: The effect of education on health outcomes and health behaviors (optimal bandwidth)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| | | | | | | | | | |
| Years of schooling | -0.002 | 0.019 | -0.015 | 0.015 | -0.007 | -0.025 | -0.005 | 0.008 | 0.007 |
| | (0.011) | (0.022) | (0.029) | (0.029) | (0.009) | (0.016) | (0.025) | (0.010) | (0.007) |
| | | | | | | | | | |
| Mean of outcome | 0.007 | 0.261 | 0.343 | 0.394 | 0.241 | 0.278 | 0.118 | 0.052 | 0.020 |
| First-stage F-stat | 9.370 | 19.26 | 18.94 | 14.68 | 24.43 | 17.45 | 9.150 | 23.58 | 25.30 |
| First-stage Coef. | -0.612*** | -0.809*** | -0.826*** | -0.744*** | -0.911*** | -0.813*** | -0.611*** | -0.849*** | -0.881*** |
| | (0.200) | (0.184) | (0.190) | (0.194) | (0.184) | (0.195) | (0.202) | (0.175) | (0.175) |
| N | 5212 | 7386 | 7218 | 6546 | 7918 | 6812 | 5559 | 8836 | 7806 |
| Bandwidth | 48.80 | 68.81 | 66.99 | 60.13 | 71.07 | 61.08 | 50.96 | 83.23 | 73.13 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses clustered at season and year-of-birth. Robust standard errors in parentheses are clustered at year-of-birth by season-of-birth. *** p<0.01.

Table 6: IV estimation using fuzzy RD: The effect of education on health outcomes and health behaviors (minimum and maximum bandwidths)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| **40 months bw** | | | | | | | | | |
| Years of schooling | -0.001 | 0.038 | -0.049 | 0.012 | 0.010 | 0.001 | -0.017 | -0.002 | 0.004 |
| | (0.011) | (0.029) | (0.053) | (0.038) | (0.016) | (0.019) | (0.020) | (0.020) | (0.012) |
| | | | | | | | | | |
| Mean of outcome | 0.008 | 0.251 | 0.346 | 0.396 | 0.243 | 0.278 | 0.115 | 0.051 | 0.020 |
| First-stage F-stat | 8.511 | 8.511 | 8.511 | 8.511 | 9.023 | 9.023 | 9.023 | 7.747 | 7.486 |
| First-stage Coef. | -0.685*** | -0.685*** | -0.685*** | -0.685*** | -0.700*** | -0.700*** | -0.700*** | -0.667*** | -0.651*** |
| | (0.235) | (0.235) | (0.235) | (0.235) | (0.233) | (0.233) | (0.233) | (0.240) | (0.238) |
| N | 4303 | 4303 | 4303 | 4303 | 4412 | 4412 | 4412 | 4185 | 4249 |
| **80 months bw** | | | | | | | | | |
| Years of schooling | 0.003 | 0.011 | -0.033 | 0.019 | -0.003 | -0.015 | 0.010 | 0.003 | 0.001 |
| | (0.006) | (0.019) | (0.028) | (0.023) | (0.010) | (0.013) | (0.016) | (0.011) | (0.008) |
| | | | | | | | | | |
| Mean of outcome | 0.010 | 0.266 | 0.340 | 0.384 | 0.240 | 0.277 | 0.118 | 0.051 | 0.021 |
| First-stage F-stat | 22.07 | 22.07 | 22.07 | 22.07 | 24.61 | 24.61 | 24.61 | 20.51 | 24.35 |
| First-stage Coef. | -0.831*** | -0.831*** | -0.831*** | -0.831*** | -0.865*** | -0.865*** | -0.865*** | -0.811*** | -0.824*** |
| | (0.177) | (0.177) | (0.177) | (0.177) | (0.174) | (0.174) | (0.174) | (0.179) | (0.167) |
| N | 8629 | 8629 | 8629 | 8629 | 8859 | 8859 | 8859 | 8383 | 8494 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth. *** p<0.01.

Table 7: IV estimation using fuzzy RD (Male): The effect of education on health outcomes and health behaviors (optimal bandwidth)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| | | | | | | | | | |
| Years of schooling | 0.010 | 0.003 | -0.043 | 0.041 | -0.007 | -0.028 | 0.004 | 0.008 | 0.008 |
| | (0.008) | (0.023) | (0.036) | (0.029) | (0.023) | (0.022) | (0.015) | (0.019) | (0.006) |
| | | | | | | | | | |
| Mean of outcome | 0.008 | 0.335 | 0.398 | 0.254 | 0.533 | 0.593 | 0.122 | 0.061 | 0.019 |
| First-stage F-stat | 20.47 | 25.90 | 25.26 | 28.96 | 19.54 | 27.46 | 37.35 | 20.15 | 40.11 |
| First-stage Coef. | -1.060*** | -1.211*** | -1.138*** | -1.213*** | -0.988*** | -1.176*** | -1.345*** | -1.058*** | -1.337*** |
| | (0.234) | (0.238) | (0.227) | (0.225) | (0.224) | (0.225) | (0.220) | (0.236) | (0.211) |
| N | 2791 | 2969 | 2922 | 3422 | 2339 | 2712 | 3053 | 2684 | 3450 |
| Bandwidth | 57.77 | 61.93 | 60.24 | 71.03 | 47.53 | 54.69 | 61.31 | 56.66 | 72.19 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth. *** p<0.01, * p<0.10.

Table 8: IV estimation using fuzzy RD (Male): The effect of education on health outcomes and health behaviors (minimum and maximum bandwidths)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| **40 months bw** | | | | | | | | | |
| Years of schooling | 0.019 | 0.016 | -0.079 | 0.043 | 0.013 | -0.002 | -0.018 | 0.008 | -0.000 |
| | (0.013) | (0.035) | (0.060) | (0.054) | (0.025) | (0.030) | (0.021) | (0.027) | (0.011) |
| | | | | | | | | | |
| Mean of outcome | 0.007 | 0.324 | 0.401 | 0.268 | 0.531 | 0.593 | 0.112 | 0.060 | 0.020 |
| First-stage F-stat | 9.360 | 9.360 | 9.360 | 9.360 | 15.18 | 15.18 | 15.18 | 9.721 | 13.75 |
| First-stage Coef. | -0.929*** | -0.929*** | -0.929*** | -0.929*** | -1.025*** | -1.025*** | -1.025*** | -0.854*** | -0.963*** |
| | (0.304) | (0.304) | (0.304) | (0.304) | (0.263) | (0.263) | (0.263) | (0.274) | (0.260) |
| N | 1940 | 1940 | 1940 | 1940 | 1996 | 1996 | 1996 | 1900 | 1941 |
| **80 months bw** | | | | | | | | | |
| Years of schooling | 0.011 | -0.017 | -0.046 | 0.051* | -0.005 | -0.027 | 0.024 | 0.018 | 0.005 |
| | (0.007) | (0.024) | (0.033) | (0.031) | (0.017) | (0.022) | (0.021) | (0.016) | (0.008) |
| | | | | | | | | | |
| Mean of outcome | 0.012 | 0.343 | 0.396 | 0.250 | 0.532 | 0.595 | 0.121 | 0.061 | 0.020 |
| First-stage F-stat | 23.13 | 23.13 | 23.13 | 23.13 | 28.51 | 28.51 | 28.51 | 24.31 | 27.00 |
| First-stage Coef. | -1.066*** | -1.066*** | -1.066*** | -1.066*** | -1.129*** | -1.129*** | -1.129*** | -1.060*** | -1.093*** |
| | (0.222) | (0.222) | (0.222) | (0.222) | (0.211) | (0.211) | (0.211) | (0.215) | (0.210) |
| N | 3830 | 3830 | 3830 | 3830 | 3940 | 3940 | 3940 | 3746 | 3815 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth. *** p<0.01, ** p<0.05.

Table 9: IV estimation using fuzzy RD (Female): The effect of education on health outcomes and health behaviors (optimal bandwidth)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| | | | | | | | | | |
| Years of schooling | -0.054 | 0.017 | -0.015 | 0.046 | 0.004 | 0.000 | -0.008 | -0.068 | 0.007 |
| | (0.088) | (0.197) | (0.135) | (0.162) | (0.009) | (0.017) | (0.051) | (0.094) | (0.024) |
| | | | | | | | | | |
| Mean of outcome | 0.007 | 0.189 | 0.296 | 0.505 | 0.006 | 0.023 | 0.116 | 0.043 | 0.021 |
| First-stage F-stat | 0.582 | 0.266 | 0.577 | 0.552 | 6.745 | 3.671 | 1.597 | 0.777 | 2.240 |
| First-stage Coef. | -0.267 | -0.165 | -0.235 | -0.216 | -0.616*** | -0.508* | -0.336 | -0.269 | -0.402 |
| | (0.351) | (0.319) | (0.310) | (0.291) | (0.237) | (0.265) | (0.266) | (0.306) | (0.269) |
| N | 2697 | 2987 | 3134 | 3310 | 4919 | 4192 | 3759 | 3037 | 3924 |
| Bandwidth | 45.58 | 50.27 | 52.63 | 55.74 | 79.04 | 68.82 | 61.67 | 52.35 | 67.45 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth. *** p<0.01.

Table 10: IV estimation using fuzzy RD (Female): The effect of education on health outcomes and health behaviors (minimum and maximum bandwidths)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| **40 months bw** | | | | | | | | | |
| Years of schooling | -0.041 | 0.106 | -0.021 | -0.044 | 0.019 | 0.014 | -0.008 | -0.012 | 0.021 |
| | (0.055) | (0.142) | (0.099) | (0.108) | (0.038) | (0.038) | (0.057) | (0.032) | (0.047) |
| | | | | | | | | | |
| Mean of outcome | 0.008 | 0.191 | 0.300 | 0.500 | 0.005 | 0.018 | 0.118 | 0.044 | 0.020 |
| First-stage F-stat | 0.981 | 0.981 | 0.981 | 0.981 | 0.804 | 0.804 | 0.804 | 1.090 | 0.557 |
| First-stage Coef. | -0.386 | -0.386 | -0.386 | -0.386 | -0.335 | -0.335 | -0.335 | -0.419 | -0.290 |
| | (0.389) | (0.389) | (0.389) | (0.389) | (0.374) | (0.374) | (0.374) | (0.401) | (0.388) |
| N | 2363 | 2363 | 2363 | 2363 | 2416 | 2416 | 2416 | 2285 | 2308 |
| **80 months bw** | | | | | | | | | |
| Years of schooling | -0.008 | 0.050 | -0.018 | -0.025 | 0.004 | 0.007 | -0.004 | -0.018 | -0.005 |
| | (0.011) | (0.047) | (0.042) | (0.046) | (0.009) | (0.014) | (0.024) | (0.022) | (0.014) |
| | | | | | | | | | |
| Mean of outcome | 0.008 | 0.204 | 0.297 | 0.492 | 0.006 | 0.022 | 0.116 | 0.044 | 0.022 |
| First-stage F-stat | 6.157 | 6.157 | 6.157 | 6.157 | 6.745 | 6.745 | 6.745 | 5.452 | 5.496 |
| First-stage Coef. | -0.602** | -0.602** | -0.602** | -0.602** | -0.616*** | -0.616*** | -0.616*** | -0.565** | -0.553** |
| | (0.243) | (0.243) | (0.243) | (0.243) | (0.237) | (0.237) | (0.237) | (0.242) | (0.236) |
| N | 4799 | 4799 | 4799 | 4799 | 4919 | 4919 | 4919 | 4637 | 4679 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth. *** p<0.01, ** p<0.05.

Table 11: Reduced form estimation: The effect of the policy change on health outcomes and health behaviors (optimal bandwidth)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| | | | | | | | | | |
| $D_i$ | 0.002 | -0.015 | 0.012 | -0.011 | 0.007 | 0.020 | 0.003 | -0.007 | -0.006 |
| | (0.007) | (0.017) | (0.023) | (0.021) | (0.009) | (0.013) | (0.016) | (0.009) | (0.006) |
| | | | | | | | | | |
| Mean of outcome | 0.007 | 0.261 | 0.343 | 0.394 | 0.241 | 0.278 | 0.118 | 0.052 | 0.020 |
| N | 5212 | 7386 | 7218 | 6546 | 7918 | 6812 | 5559 | 8836 | 7806 |
| Bandwidth | 48.8 | 68.81 | 66.99 | 60.13 | 71.07 | 61.08 | 50.96 | 83.23 | 73.13 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth.

Table 12: Reduced form estimation (Male): The effect of the policy change on health outcomes and health behaviors (optimal bandwidth)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| | | | | | | | | | |
| $D_i$ | -0.011 | -0.003 | 0.049 | -0.050 | 0.007 | 0.033 | -0.006 | -0.008 | -0.010 |
| | (0.008) | (0.028) | (0.040) | (0.034) | (0.023) | (0.026) | (0.021) | (0.020) | (0.009) |
| | | | | | | | | | |
| Mean of outcome | 0.008 | 0.335 | 0.398 | 0.254 | 0.533 | 0.593 | 0.122 | 0.061 | 0.019 |
| N | 2791 | 2969 | 2922 | 3422 | 2339 | 2712 | 3053 | 2684 | 3450 |
| Bandwidth | 57.77 | 61.93 | 60.24 | 71.03 | 47.53 | 54.69 | 61.31 | 56.66 | 72.19 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth. * p<0.10.

Table 13: Reduced form estimation (Male): The effect of the policy change on health outcomes and health behaviors (minimum and maximum bandwidths)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| **40 months bw** | | | | | | | | | |
| $D_i$ | -0.018 | -0.015 | 0.073 | -0.040 | -0.013 | 0.002 | 0.018 | -0.007 | 0.000 |
| | (0.011) | (0.034) | (0.051) | (0.048) | (0.026) | (0.031) | (0.022) | (0.024) | (0.011) |
| | | | | | | | | | |
| Mean of outcome | 0.007 | 0.324 | 0.401 | 0.268 | 0.531 | 0.593 | 0.112 | 0.060 | 0.020 |
| N | 1940 | 1940 | 1940 | 1940 | 1996 | 1996 | 1996 | 1900 | 1941 |
| **80 months bw** | | | | | | | | | |
| $D_i$ | -0.012* | 0.018 | 0.049 | -0.055* | 0.006 | 0.030 | -0.027 | -0.019 | -0.005 |
| | (0.007) | (0.026) | (0.033) | (0.031) | (0.020) | (0.025) | (0.022) | (0.016) | (0.009) |
| | | | | | | | | | |
| Mean of outcome | 0.012 | 0.343 | 0.396 | 0.250 | 0.532 | 0.595 | 0.121 | 0.061 | 0.020 |
| N | 3830 | 3830 | 3830 | 3830 | 3940 | 3940 | 3940 | 3746 | 3815 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth. * p<0.10.

Table 14: IV estimation using fuzzy RD (2008 survey): The effect of education on health outcomes and health behaviors (optimal bandwidth)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| | | | | | | | | | |
| Years of schooling | 0.005 | 0.028 | -0.006 | -0.023 | -0.004 | -0.006 | -0.006 | 0.008 | -0.001 |
| | (0.006) | (0.024) | (0.029) | (0.018) | (0.013) | (0.018) | (0.015) | (0.012) | (0.004) |
| | | | | | | | | | |
| Mean of outcome | 0.014 | 0.344 | 0.358 | 0.285 | 0.216 | 0.239 | 0.080 | 0.057 | 0.007 |
| First-stage F-stat | 20.07 | 19.01 | 17.56 | 19.01 | 21.36 | 16.42 | 21.67 | 14.75 | 17.79 |
| First-stage Coef. | -1.183*** | -1.218*** | -1.171*** | -1.218*** | -1.124*** | -1.002*** | -1.189*** | -1.122*** | -1.169*** |
| | (0.264) | (0.279) | (0.279) | (0.279) | (0.243) | (0.247) | (0.255) | (0.292) | (0.277) |
| N | 3986 | 3588 | 3552 | 3588 | 4662 | 5178 | 4117 | 3473 | 3610 |
| Bandwidth | 81.34 | 74.86 | 73.93 | 74.83 | 93.08 | 102.6 | 82.03 | 75.41 | 79.11 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth. * p<0.10.

Table 15: IV estimation using fuzzy RD (2015 survey): The effect of education on health outcomes and health behaviors (optimal bandwidth)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| | | | | | | | | | |
| Years of schooling | -0.003 | 0.002 | -0.139 | 0.061 | -0.029 | -0.048 | 0.027 | -0.024 | 0.023 |
| | (0.009) | (0.058) | (0.166) | (0.050) | (0.024) | (0.031) | (0.027) | (0.023) | (0.019) |
| Mean of outcome | 0.006 | 0.194 | 0.329 | 0.469 | 0.259 | 0.301 | 0.150 | 0.046 | 0.031 |
| First-stage F-stat | 5.031 | 2.245 | 1.518 | 5.779 | 5.697 | 5.697 | 6.653 | 6.272 | 4.384 |
| First-stage Coef. | -0.556** | -0.385 | -0.308 | -0.569** | -0.654** | -0.654** | -0.628*** | -0.602** | -0.572** |
| | (0.248) | (0.257) | (0.250) | (0.237) | (0.274) | (0.274) | (0.244) | (0.240) | (0.273) |
| N | 3903 | 3489 | 3433 | 4814 | 3790 | 3790 | 5054 | 3987 | 3760 |
| Bandwidth | 64.45 | 57.52 | 56.32 | 80.39 | 60.62 | 60.53 | 81.76 | 66.45 | 60.25 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth. *** p<0.01, ** p<0.05, * p<0.10.

Table 16: IV estimation using fuzzy RD: The effect of education on wage and employment (minimum and maximum bandwidths)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| | All | | | Male | | | Female | | |
| VARIABLES | Wage (log)[1] | Wage (log)[2] | Employed[3] | Wage (log)[1] | Wage (log)[2] | Employed[3] | Wage (log)[1] | Wage (log)[2] | Employed[3] |
| **40 months bw** | | | | | | | | | |
| Years of Schooling | 0.049 | 0.022 | 0.001 | 0.047 | 0.020 | 0.000 | 0.113 | 0.095 | 0.009 |
| | (0.032) | (0.030) | (0.019) | (0.034) | (0.027) | (0.014) | (0.128) | (0.141) | (0.056) |
| | | | | | | | | | |
| Mean of outcome | 7.570 | 7.594 | 0.618 | 7.613 | 7.639 | 0.945 | 7.349 | 7.359 | 0.235 |
| First-stage F-stat | 10.33 | 9.945 | 6.063 | 15.84 | 14.84 | 10.88 | 1.045 | 0.824 | 1.428 |
| First-stage Coef. | -1.273*** | -1.251*** | -0.781** | -1.338*** | -1.327*** | -1.219*** | -0.987 | -0.872 | -0.490 |
| | (0.396) | (0.397) | (0.317) | (0.336) | (0.344) | (0.369) | (0.965) | (0.960) | (0.410) |
| N | 3645 | 3679 | 8469 | 3061 | 3092 | 4567 | 584 | 587 | 3902 |
| **80 months bw** | | | | | | | | | |
| Years of Schooling | 0.011 | -0.006 | 0.005 | 0.001 | -0.018 | -0.003 | 0.060 | 0.050 | 0.026 |
| | (0.025) | (0.023) | (0.012) | (0.028) | (0.027) | (0.012) | (0.050) | (0.049) | (0.027) |
| | | | | | | | | | |
| Mean of outcome | 7.563 | 7.590 | 0.606 | 7.606 | 7.636 | 0.947 | 7.342 | 7.352 | 0.230 |
| First-stage F-stat | 20.47 | 20.73 | 17.27 | 22.63 | 21.55 | 18.70 | 5.100 | 5.294 | 6.244 |
| First-stage Coef. | -1.189*** | -1.195*** | -0.842*** | -1.135*** | -1.137*** | -0.993*** | -1.594** | -1.619** | -0.709** |
| | (0.263) | (0.262) | (0.203) | (0.239) | (0.245) | (0.230) | (0.706) | (0.704) | (0.284) |
| N | 7365 | 7437 | 17402 | 6166 | 6230 | 9129 | 1199 | 1207 | 8273 |

*Notes*. Each coefficient in a column represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth. *** p<0.01, ** p<0.05, * p<0.10.

[1]Monthly (real) wage from primary occupation. [2]Monthly (real) wage from all occupations. [3]Out of the labor force is also counted as unemployed.

## Appendix Tables

**Appendix Table A.1: Descriptive statistics by survey year**

| | 2008 | | | 2015 | | |
|---|---|---|---|---|---|---|
| | All | Control | Treatment | All | Control | Treatment |
| Age (SD) | 29.81 (2.95) | 32.639 | 27.580*** | 36.69 (2.89) | 39.443 | 34.456*** |
| Gender (% Female) | 0.567 | 0.556 | 0.574 | 0.542 | 0.546 | 0.539 |
| Married (%) | 0.820 | 0.895 | 0.761*** | 0.913 | 0.918 | 0.909 |
| Urban (%) | 0.422 | 0.420 | 0.425 | 0.500 | 0.501 | 0.498 |
| Wealth index | 3.047 | 3.036 | 3.055 | 3.220 | 3.190 | 3.244 |
| N | 2,935 | 1294 | 1641 | 3,792 | 1697 | 2095 |

***Difference between treatment and control groups significant at the 1% level. Observations within 60 months bandwidth.

Appendix Table A2: IV estimation using fuzzy RD: The effect of education on health outcomes and health behaviors (optimal bandwidth)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| | | | | | | | | | |
| Years of schooling | -0.003 | 0.012 | -0.015 | 0.021 | -0.014 | -0.033 | -0.004 | 0.008 | 0.006 |
| | (0.009) | (0.018) | (0.027) | (0.031) | (0.014) | (0.021) | (0.023) | (0.010) | (0.006) |
| | | | | | | | | | |
| Mean of outcome | 0.007 | 0.261 | 0.343 | 0.394 | 0.241 | 0.278 | 0.118 | 0.052 | 0.020 |
| First-stage F-stat | 6.368 | 14.11 | 14.05 | 10.47 | 17.66 | 11.66 | 5.595 | 15.16 | 16.97 |
| First-stage Coef. | -0.681** | -0.890*** | -0.911*** | -0.812*** | -1.023*** | -0.877*** | -0.657** | -0.862*** | -0.958*** |
| | (0.270) | (0.237) | (0.243) | (0.251) | (0.244) | (0.257) | (0.278) | (0.221) | (0.233) |
| N | 5212 | 7386 | 7218 | 6546 | 7918 | 6812 | 5559 | 8836 | 7806 |
| Bandwidth | 48.80 | 68.81 | 66.99 | 60.13 | 71.07 | 61.08 | 50.96 | 83.23 | 73.13 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control only for age and age-squared. Robust standard errors in parentheses are clustered at season and year-of-birth. *** p<0.01, ** p<0.05.

Appendix Table A3: IV estimation using fuzzy RD (Male): The effect of education on health outcomes and health behaviors (minimum and maximum bandwidths)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| **40 months bw** | | | | | | | | | |
| Years of schooling | 0.020 | 0.021 | -0.084 | 0.044 | 0.012 | -0.003 | -0.019 | 0.009 | -0.001 |
| | (0.014) | (0.040) | (0.060) | (0.057) | (0.027) | (0.032) | (0.021) | (0.028) | (0.012) |
| | | | | | | | | | |
| Mean of outcome | 0.007 | 0.324 | 0.401 | 0.268 | 0.531 | 0.593 | 0.112 | 0.060 | 0.020 |
| First-stage F-stat | 8.619 | 8.619 | 8.619 | 8.619 | 14.44 | 14.44 | 14.44 | 9.846 | 11.96 |
| First-stage Coef. | -0.871*** | -0.871*** | -0.871*** | -0.871*** | -0.962*** | -0.962*** | -0.962*** | -0.848*** | -0.883*** |
| | (0.297) | (0.297) | (0.297) | (0.297) | (0.253) | (0.253) | (0.253) | (0.270) | (0.255) |
| N | 1940 | 1940 | 1940 | 1940 | 1996 | 1996 | 1996 | 1900 | 1941 |
| **80 months bw** | | | | | | | | | |
| Years of schooling | 0.011 | -0.015 | -0.043 | 0.047 | -0.007 | -0.026 | 0.022 | 0.017 | 0.004 |
| | (0.007) | (0.022) | (0.030) | (0.030) | (0.016) | (0.021) | (0.019) | (0.015) | (0.007) |
| | | | | | | | | | |
| Mean of outcome | 0.012 | 0.343 | 0.396 | 0.250 | 0.532 | 0.595 | 0.121 | 0.061 | 0.020 |
| First-stage F-stat | 21.49 | 21.49 | 21.49 | 21.49 | 27.48 | 27.48 | 27.48 | 23.47 | 25.65 |
| First-stage Coef. | -1.122*** | -1.122*** | -1.122*** | -1.122*** | -1.208*** | -1.208*** | -1.208*** | -1.131*** | -1.159*** |
| | (0.242) | (0.242) | (0.242) | (0.242) | (0.230) | (0.230) | (0.230) | (0.234) | (0.229) |
| N | 3830 | 3830 | 3830 | 3830 | 3940 | 3940 | 3940 | 3746 | 3815 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control only for age and age-squared. Robust standard errors in parentheses are clustered at season and year-of-birth. *** p<0.01, ** p<0.05.

Appendix Table A4: IV estimation using fuzzy RD (Female): The effect of education on health outcomes and health behaviors (minimum and maximum bandwidths)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| **40 months bw** | | | | | | | | | |
| Years of schooling | -0.019 | 0.035 | -0.009 | -0.007 | 0.007 | 0.005 | -0.003 | -0.008 | 0.009 |
| | (0.018) | (0.044) | (0.042) | (0.044) | (0.013) | (0.014) | (0.025) | (0.015) | (0.016) |
| | | | | | | | | | |
| Mean of outcome | 0.008 | 0.191 | 0.300 | 0.500 | 0.005 | 0.018 | 0.118 | 0.044 | 0.020 |
| First-stage F-stat | 3.142 | 3.142 | 3.142 | 3.142 | 2.456 | 2.456 | 2.456 | 2.641 | 2.298 |
| First-stage Coef. | -0.882* | -0.882* | -0.882* | -0.882* | -0.793 | -0.793 | -0.793 | -0.839 | -0.782 |
| | (0.497) | (0.497) | (0.497) | (0.497) | (0.506) | (0.506) | (0.506) | (0.516) | (0.516) |
| N | 2363 | 2363 | 2363 | 2363 | 2416 | 2416 | 2416 | 2285 | 2308 |
| **80 months bw** | | | | | | | | | |
| Years of schooling | -0.006 | 0.036 | -0.014 | -0.015 | 0.003 | 0.006 | -0.003 | -0.016 | -0.003 |
| | (0.008) | (0.035) | (0.032) | (0.035) | (0.007) | (0.011) | (0.018) | (0.019) | (0.011) |
| | | | | | | | | | |
| Mean of outcome | 0.008 | 0.204 | 0.297 | 0.492 | 0.006 | 0.022 | 0.116 | 0.044 | 0.022 |
| First-stage F-stat | 6.507 | 6.507 | 6.507 | 6.507 | 6.524 | 6.524 | 6.524 | 4.599 | 5.944 |
| First-stage Coef. | -0.781** | -0.781** | -0.781** | -0.781** | -0.794** | -0.794** | -0.794** | -0.669** | -0.747** |
| | (0.306) | (0.306) | (0.306) | (0.306) | (0.311) | (0.311) | (0.311) | (0.312) | (0.306) |
| N | 4799 | 4799 | 4799 | 4799 | 4919 | 4919 | 4919 | 4637 | 4679 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control only for age and age-squared. Robust standard errors in parentheses are clustered at season and year-of-birth. *** p<0.01, ** p<0.05, * p<0.10.

Appendix Table A5: Reduced form estimation: The effect of the policy change on health outcomes and health behaviors (minimum and maximum bandwidths)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| **40 months bw** | | | | | | | | | |
| $D_i$ | 0.001 | -0.026 | 0.034 | -0.009 | -0.007 | -0.001 | 0.012 | 0.002 | -0.003 |
| | (0.008) | (0.017) | (0.032) | (0.026) | (0.010) | (0.014) | (0.017) | (0.013) | (0.008) |
| | | | | | | | | | |
| Mean of outcome | 0.008 | 0.251 | 0.346 | 0.396 | 0.243 | 0.278 | 0.115 | 0.051 | 0.020 |
| N | 4303 | 4303 | 4303 | 4303 | 4412 | 4412 | 4412 | 4185 | 4249 |
| **80 months bw** | | | | | | | | | |
| $D_i$ | -0.003 | -0.009 | 0.027 | -0.015 | 0.002 | 0.013 | -0.009 | -0.002 | -0.001 |
| | (0.005) | (0.016) | (0.020) | (0.018) | (0.009) | (0.011) | (0.013) | (0.009) | (0.006) |
| | | | | | | | | | |
| Mean of outcome | 0.002 | 0.266 | 0.340 | 0.384 | 0.240 | 0.277 | 0.118 | 0.052 | 0.021 |
| N | 8629 | 8629 | 8629 | 8629 | 8859 | 8859 | 8859 | 8383 | 8494 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth. *** p<0.01, ** p<0.05, * p<0.10.

Appendix Table A6: Reduced form estimation (Female): The effect of the policy change on health outcomes and health behaviors (optimal bandwidth)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| | | | | | | | | | |
| $D_i$ | 0.014* | -0.003 | 0.004 | -0.010 | -0.002 | -0.000 | 0.003 | 0.018* | -0.003 |
| | (0.008) | (0.032) | (0.032) | (0.034) | (0.005) | (0.009) | (0.018) | (0.011) | (0.009) |
| | | | | | | | | | |
| Mean of outcome | 0.007 | 0.189 | 0.296 | 0.505 | 0.006 | 0.023 | 0.116 | 0.043 | 0.021 |
| N | 2697 | 2987 | 3134 | 3310 | 4919 | 4192 | 3759 | 3037 | 3924 |
| Bandwidth | 45.58 | 50.27 | 52.63 | 55.74 | 79.04 | 68.82 | 61.67 | 52.35 | 67.45 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth.

Appendix Table A7: Reduced form estimation (Female): The effect of the policy change on health outcomes and health behaviors (minimum and maximum bandwidth)

| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) | (9) |
|---|---|---|---|---|---|---|---|---|---|
| VARIABLES | Underweight | Normal weight | Overweight | Obese | Current smoker | Ever smoker | Hypertension | Hepatitis | Diabetes |
| **40 months bw** | | | | | | | | | |
| $D_i$ | 0.016* | -0.041 | 0.008 | 0.017 | -0.006 | -0.005 | 0.003 | 0.005 | -0.006 |
| | (0.009) | (0.030) | (0.039) | (0.038) | (0.008) | (0.010) | (0.020) | (0.011) | (0.012) |
| | | | | | | | | | |
| Mean of outcome | 0.008 | 0.191 | 0.300 | 0.500 | 0.005 | 0.018 | 0.118 | 0.044 | 0.020 |
| N | 2363 | 2363 | 2363 | 2363 | 2416 | 2416 | 2416 | 2285 | 2308 |
| **80 months bw** | | | | | | | | | |
| $D_i$ | 0.005 | -0.030 | 0.011 | 0.015 | -0.002 | -0.005 | 0.003 | 0.010 | 0.003 |
| | (0.006) | (0.024) | (0.025) | (0.027) | (0.005) | (0.008) | (0.015) | (0.011) | (0.008) |
| | | | | | | | | | |
| Mean of outcome | 0.008 | 0.204 | 0.297 | 0.492 | 0.006 | 0.022 | 0.116 | 0.044 | 0.022 |
| N | 4799 | 4799 | 4799 | 4799 | 4919 | 4919 | 4919 | 4637 | 4679 |

*Notes*. Each coefficient in column and panel represents a separate regression. All models use a quadratic polynomial in month-year of birth centered around the cut-off. All models control for age and its square, gender, marital status, wealth index, dummies for region of residence, and urban status. Robust standard errors in parentheses are clustered at season and year-of-birth. *** p<0.01, ** p<0.05, * p<0.10.