The Effect of Education on Mortality and Health: Evidence from a Schooling Expansion in Romania

Ofer Malamud
Northwestern University
NBER and CESifo

Andreea Mitrut
University of Gothenburg
and UCLS

Cristian Pop-Eleches
Columbia University
BREAD and NBER

October 2020

Abstract*

This paper examines a schooling expansion in Romania that increased educational attainment for successive cohorts born between 1945 and 1950. We use a difference-in regression discontinuities (D-RD) empirical strategy based on school entry cutoff dates to estimate impacts on mortality using 1994-2016 Vital Statistics data, self-reported health in the 2011 Romanian Census, and hospitalizations from 1997-2017 in-patient registers. We find that the schooling reform led to significant increases in years of schooling but did not affect mortality, hospitalizations, or self-reported health. These estimates provide new evidence on the causal effect of education on mortality and health outside of high-income countries and at lower margins of educational attainment.

* We would especially like to thank Andreea Balan-Cohen for her work on the schooling reform in Romania for a different project that is still in progress. Andreea Mitrut gratefully acknowledge support from Jan Wallanders and Tom Hedelius Fond. We have benefited from comments by Doug Almond, Robert Kaestner and Bash Mazumder, as well as participants at the ERMAS 2017, the CHERP conference at the Federal Reserve Bank of Chicago, and the NBER Health Economics Spring 2018 program meeting. All errors are our own.

1. Introduction

There is substantial evidence showing that more educated people have better health and longer life expectancies. However, whether this correlation reflects a causal relationship remains an open question. A number of recent papers have used changes in compulsory schooling requirements to identify the causal impact of schooling on mortality in the United States (Lleras-Muney, 2005; Mazumder, 2008), the United Kingdom (Clark and Royer, 2013; Davies, et al. 2016), France (Albouy and Lequien, 2009), the Netherlands (van Kippersluis, et al., 2011), Sweden (Meghir, et al., 2018), and Taiwan (Kan, 2016). While this empirical approach can be compelling, the findings have been mixed and sometimes contradictory, even when based on the same educational expansions. Moreover, most of these studies are focused on high income countries where compulsory schooling laws usually affect students already enrolled in secondary school. As a result, we know relatively little about the causal effect of education on health and mortality in low or middle-income countries, and at lower margins of educational attainment.

This paper examines the impact of a schooling expansion in Romania during the late 1950s and early 1960s, which sought to provide all students with at least 7 years of compulsory education. We show that successive cohorts of individuals, born between 1945 and 1950, who were affected by this schooling expansion, experienced rising educational attainment. We first consider a regression discontinuity (RD) design at the day level to compare individuals born just before the school entry cutoff of January 1 to those born just after, who were almost identical in age but began school later and therefore had greater opportunities to extend their education. However, since students born immediately before and after January 1 were also the oldest and youngest in their respective cohorts, we draw on cohorts born after the systematic schooling expansion had concluded, and utilize a difference-in-regression discontinuity (D-RD) design to separate the effect of increased education from that of relative age and other confounding shocks.

Using the complete count (100%) Romanian Census in 1992, we demonstrate that Romania's schooling expansion led to significant increases in years of schooling for the affected cohorts born between 1945 and 1950. This is driven by an increase in the fraction of students who continue beyond the four years of primary school to complete an additional three years of lower secondary education, with some students continuing even further onto upper secondary schools.

To calculate our mortality and health outcomes we use high-quality data from Vital Statistics records, which are rarely available in middle and low-income countries. Detailed information on deaths from 1994 and 2016 indicate that the schooling expansion did not reduce the mortality of affected cohorts up to the age of 71. Nor are there reductions in mortality from more specific causes of death. We also examine two health outcomes that may affect quality of life and life expectancy: the total number of days spent in hospital (overall and by specific cause of hospitalization) based on individual Romanian Inpatient registers from 1997 to 2017, and a measure of self-reported health problems using data from the 2011 Romanian Census. For both of these outcomes, the estimated effects are small and insignificant, suggesting that the schooling reform had no discernable impacts on health.

While we have good measures of mortality and health, we lack data on socio-economic outcomes such as income or earnings. We examine several outcomes available in the 1992 Census, such as employment or occupational skills, but these do not vary much in Romania's highly centralized economic system in the immediate period following the fall of Communism. The results for these socio-economic outcomes, as well as fertility, suggest positive effects but are not robust across specifications.

Our main findings indicate that more education does not help individuals avoid or postpone deaths during middle and old age. This is consistent with the null results in the most recent papers by Clark and Royer (2013) and Meghir et al. (2018) for the United Kingdom and Sweden. However, to the best of our knowledge, this is the first paper to provide compelling estimates for the causal effect of education on mortality outside of high-income countries and at lower margins of educational attainment. We do not interpret these estimates as an argument against further educational expansions in the developing world. But they do suggest the need to be more circumspect about the potential for such expansions to improve health and increase life expectancy, at least at lower margins of educational attainment.

The paper is organized as follows. Section 2 reviews the related literature. Section 3 provides a background of the Romanian educational system and the schooling expansion. Section 4 describes the data. Section 5 explains the empirical strategy. Section 6 presents the results, and Section 7 concludes.

2. Related Literature

This section reviews some of the previous literature estimating the causal impact of education on health and mortality. We begin with a discussion of studies that take advantage of changes in compulsory schooling requirements. Then we describe some of the alternative empirical approaches used for identifying the causal effect of education at higher margins of educational attainment. For more detailed reviews of these and other studies, see Grossman (2006), Mazumder (2012) and Galama et al. (2020).

For the United States, Lleras-Muney (2005) uses Census data to examine the impact of changes in compulsory schooling laws between 1915 and 1939 that affected students over 14 years of age. Her instrumental variables (IV) estimates indicate that an additional year of schooling leads to significant declines in the probability of dying in the next 10 years. In a follow-up study, Mazumder (2008) notes that these results are not robust to including state-specific trends but presents evidence from the Survey of Income and Program Participation (SIPP) showing positive impacts of education on self-reported health status. Relatedly, Black et al. (2016) argue that virtually all of the variation in mortality rates is captured by cohort effects and state effects, making it difficult to reliably estimate the effects of changing educational attainment due to state-level changes in compulsory schooling.[1]

For the United Kingdom, Clark and Royer (2013) use changes to British compulsory schooling laws in 1947 and 1972 that increased the minimum school leaving age from 14 to 15 and then from 15 to 16. Their regression discontinuity (RD) design does not provide strong evidence for an impact of education on mortality or other health outcomes. Davies et al. (2016) re-examine the 1972 change in compulsory schooling using UK Biobank data and find a statistically significant decline in mortality but their results are somewhat sensitive to functional form.

Other studies are mostly focused on other European countries: For Sweden, Meghir et al. (2018) do not find improvements in mortality and other health measures for affected cohorts following an educational reform in Sweden that raised the number of years of compulsory schooling from 7/8 to 9, eliminated early selection based on academic ability, and introduced a national curriculum. Arendt (2005) and Albouy and

[1] In a paper that considers the effect of school quality on health, Aaronson et al. (2017) find that childhood exposure to Rosenwald schools in the Jim Crow south increased life expectancy, after accounting for the negative effects of migration.

Lequien (2009) also find no statistically significant impact of compulsory school reforms on health outcomes in Denmark or mortality in France, respectively. Yet van Kippersluis et al. (2011) do find that increasing compulsory school beyond grade 6 in the Netherlands leads to significant reduction in mortality in old age.

Outside of the United States and Europe, Kan (2016) studies Taiwan, which looks like a developed country by most measures, and finds that the extension of compulsory education from 6 to 9 years reduced men's mortality rate but did not affect women's mortality. In a paper contemporaneous with ours, Dursun et al. (2018) examine the effect of a Turkish schooling expansion on measures of health, but not on mortality,

A different set of studies use draft avoidance behavior in the United States during the Vietnam War to estimate the impact of college education on mortality and health outcomes. Buckles et al. (2016) show that the increased college going among men in cohorts associated with greater draft avoidance also leads to lower mortality in subsequent years. Grimard and Parent (2007) and de Walque (2007) use a similar identification strategy to estimate impacts on smoking behavior and find evidence suggesting that more education reduces the take-up of smoking and current smoking. That the causal impact of education on mortality at the margin of a college education appears to differ from the impact at the margin of compulsory schooling suggests that looking at another margin of educational attainment could be informative as well.

In our own review of the literature, and in those by Grossman (2006) and Galama et al. (2020), we have not found any papers that provide compelling causal estimates for the impact of education on mortality in low and middle-income countries and at lower margins of schooling.[2]

[2] We have limited our review of the literature to the effects of education on own health and mortality. A separate literature has explored the impact of parental education on child health outcomes; e.g. McCrary and Royer (2011) and Chou et al. (2010).

3. Background on Education in Romania

3.1 Historical context

During the post-war period, the structure and the organization of education in Romania was largely based on the model in the Soviet Union as codified by Decree No. 175 of 1948 (Braham, 1972). There were several types of schools offering elementary and secondary education. First, there were 4-year primary schools that offered compulsory primary education from grades 1 through 4. Second, there were 7-year general schools, called gymnasiums, which offered the same first four grades as primary schools, but also grades 5-7 (that were later extended to include grade 8). These gymnasiums corresponded to the lower secondary education according to the International Standard Classification of Education (ISCED), and still referred as such today.[3] However, not all localities provided lower secondary education, especially in the rural areas.[4]

Third, there was the upper secondary education which included: *vocational and technical schools*, often operated under the supervision of the large state enterprises or collective agricultural farms, and *(academic) high schools* that provided education (grades 8 to 11), and prepared students for the baccalaureate exam which was a prerequisite for entry into higher education (i.e. university).[5] These upper secondary options were available only to graduates of gymnasiums, i.e. those who completed grade 7 of lower secondary schools.

[3] See https://eacea.ec.europa.eu/national-policies/eurydice/content/romania_en

[4] Children who did not have a school offering grades 5 and above in their locality would have had to commute to a nearby town. The cost of commuting prevented many families from sending their children to acquire additional education, especially in rural areas where children were also expected to help their families with agriculture/farm related activities and household chores.

[5] In some large cities, the government provided 11-year schools, which offered grades 1 through 11 (4 grades of primary + 3 grades of gymnasium + 4 grades of high school) in one school building. Note, there were no differences in curricula for grades 1 to 4 whether in a primary school, a gymnasium, or an 11-year school; nor any differences for grades 5 to 7 in a gymnasium or in an 11-year school.

Upon completing their education, individuals entered a labor market characterized by a centralized wage-setting process, with standard rules based on occupations and industry (Andrén and Andrén, 2015). Similar to other communist countries, the highly centralized political system maintained small wage differentials that did vary with the workers' education, experience and occupation (or industry). Many individuals with lower levels of education worked on collective agricultural farms that paid fixed wages. Thus, while wages did differ by occupation or experience, returns to education under Communism were substantially lower than in free market economies.

During the late 1940s and early 1950s, Romania's government was focused on providing basic literacy education for all ages.[6] By the mid-1950s, it turned its attention towards increasing enrollment beyond the first four grades. According to Giurescu et al. (1971, p. 351), the five-year plan of 1955-1960 specified that the extension of compulsory schooling to 7 years was to be given special attention by the party and government. Thus, the directives of the Communist Party's Second Congress of 1955, which outlined the second five year plan, envisioned a "situation under which, by 1960-1961, the fifth grade would enroll 90 percent of the 4-year school graduates; and under which, according to the Third Five Year plan, the 7-year school would be universal and compulsory." (Braham, 1963).

The extension of compulsory schooling from 4 to 7 years meant that lower secondary schools (i.e. gymnasiums) had to take in a surplus of children who graduated primary school and who otherwise would not have continued their education. "At first, only the first four grades were made compulsory, but villages and rural communities having 7-year schools were required by virtue of Decision No. 1035/1958 to make the 7

[6] The Ministry of Education organized literacy courses lasting 1-2 years for people aged 14-55 and, in some cases, authorized the establishment of two-year elementary schools for literacy.

year schooling period universal beginning with the 1958-1959 academic year" (Braham, 1963).[7] Nevertheless, the process to provide graduates of 4-years schools access to lower secondary education was not immediate and was constrained by a lack of enough schools offering 7 years of compulsory schooling.[8] Filipescu and Oprea (1972) confirm the gradual process of expanding education at the lower secondary level. They explain that the expansion of 7-year compulsory education began in 1956 within towns and larger villages that already had schools beyond the 4th grade, and that it gradually expanded until it was close to universal by 1961-1962.

3.2 Patterns over time

We can document some of these changes using aggregate administrative data on enrollment from the Annual Statistics of the Socialist Republic of Romania. Figure 1 shows the number of students graduating from lower secondary schools (gymnasiums) between 1951 and 1969. During this period, graduation from these lower secondary schools increased sharply from 116,698 in 1959 to 329,739 in 1963 and stayed at similar levels through the late 1960s. Braham (1963) confirms that during the period 1960-1965 the State boosted the funds for education, which led to the construction of over 15,000 classrooms (of which about 70% were in rural areas). The number of teachers employed also increased, especially between the 1960-61 and 1967-68 academic years, with the largest increase of 70 percent in grades 5 to 11 (Braham, 1972). Accordingly, the overall the pupil-teacher ratio remained largely similar during this period.[9]

[7] Decision No. 1383/1948 had made the first four grades compulsory but it was only enforced in 1955.
[8] Braham (1963) notes that "with rural communities retaining the 4-year compulsory level, the lack of detailed planning to elevate their schools to the 7-year compulsory level has left an irregular pattern of schooling in the provinces".
[9] Our own calculations using the 1967 Romanian Yearbook Statistics reveals quite stable student-teacher ratios for the academic years 1060-61 to 1966-66: an average of 28 students per teacher for grades 1 to 4,

Further evidence for these dramatic changes can be observed at the cohort level. By law, students entered grade 1 in September of the year following the calendar year in which they reached 6 years of age. Thus, the cohort born in 1945 was 6 years of age in 1951, entered first grade in the fall of 1952, entered fifth grade in the fall of 1956 and would have graduated with 7 years of schooling in the spring of 1959. This cohort should be the first cohort that could have been affected by the policy reform. Similarly, the cohort born in 1947 was the first cohort to have potentially benefited from the 1958 Government Decision that made 7-year of schooling compulsory. Finally, the cohort that entered fifth grade in 1961-1962, which according to Filipescu and Oprea (1972) was the first cohort to have achieved universal 7-years of compulsory education, was born in 1950.[10]

Figure 2 shows the highest education level completed by year of birth in the Romanian Census of 1992. For our main cohorts of interest, born between 1945 and 1950, we see a sharp decline in the proportion of individuals with primary education and a sharp increase in the proportion of individuals who have secondary education. For the cohort born in 1950, the fraction with only a primary education was below 8 percent, suggesting that 7 years of compulsory schooling was nearly universal. For cohorts born after 1950, the levels of primary and secondary schooling are much more stable with only gradual changes over time. It is also apparent that cohorts born before 1944 experienced large increases in educational attainment which were mainly driven by the successful literacy campaigns mentioned earlier.

---

and about 21 students per teacher for grades 5 to 7 (or 8). We do not have yearly data for the cohorts in school before the 1960-61 academic year.

[10] Additional sources indicate that enrolments in fifth grade included 99 percent of the fourth-grade graduates by the 1961-1962 school year, confirming that the 7 years of compulsory schooling became universal by this time (Braham, 1963).

It is important to note that graduation from lower secondary schools opened up opportunities for further educational attainment at the upper secondary level, and potentially even higher education.[11] In other words, Romania's schooling expansion enabled individuals to continue beyond primary education. Therefore, in Appendix Figure 1, we plot the "residual" percent of individuals born between 1944 and 1955 who completed primary education or less by their month of birth, after accounting for calendar month of birth effects. Consistent with Figure 2, we observe the large decrease in the proportion of individuals who have only primary education or less among those born between 1945 and 1950. More importantly for our empirical strategy, these declines in primary education occur discontinuously, with disproportionately large decreases for those born after January 1. At the same time, no declines are visible for the cohorts born between 1951 and 1953, which we will use as controls in our empirical strategy. The patterns in this figure suggest that we can use detailed information on date of birth to estimate the impact of the schooling expansion using a regression discontinuity design.

To summarize, the graphical evidence shown here is broadly consistent with the historical record of educational reforms in Romania. Education levels past the first 4 years of primary schooling expanded from the 1956-1957 school-year and by 1961-1962, enrollment in lower secondary education was nearly universal. Thus, the schooling expansion affected cohorts born starting in 1945 and was essentially completed for cohorts born after 1950. Finally, though we do not have a direct measure of changes in the quality of schooling during this period, Braham (1972) confirms that the curriculum

[11] The communist regime wished to train "politically reliable workers," especially in vocational and technical schools that admitted only students who completed 7 years of compulsory education. They often offered the possibility of taking evening or correspondence courses such that workers could attend them without leaving the field of production.

and the educational standards in general schools did not experience any major changes until 1969.[12] Moreover, in communist Romania, the school curriculum and standards were centrally planned and standardized across schools (see also Braham, 1972).

4. Data

Our main sample consists of individuals born in Romania between 1945 and 1953.[13] Those born from 1945-1950 were enrolled in the affected grades during the period of schooling expansion, while those born from 1951-1953 were enrolled after the expansions had already been completed. As we explain in our discussion of the empirical strategy in section 5, we use the three subsequent cohorts born immediately after the end of the schooling expansion to account for the independent effect of relative age, to address the possibility of school-cohort specific shocks, and the misreporting of births.[14] Moreover, as shown in Figure 2, these cohorts did not experience any large changes in the level of the highest completed education.

We put together information on these cohorts from several different datasets. We use the complete count (100 percent) 1992 Romanian Census, when individuals were 38 to 47 years of age, to estimate the impact of the schooling reform on the level of completed education, certain labor market outcomes, and conduct specification checks of our empirical strategy. Two features make this dataset especially useful for our analysis: First, with over 300,000 observations in each yearly birth cohort, we have substantial

[12] In 1968, a new reform introduced a 10-year compulsory school but this does not affect our cohorts of interest.

[13] Strictly speaking, because our empirical strategy considers individuals born within 180 days of January 1 for the years 1945 to 1953, we include those born between mid-1944 and mid-1953.

[14] The three subsequent cohorts born immediately after the end of the schooling expansion are most similar in age to the cohorts affected by the schooling expansion and offer sufficiently large samples. However, our results are mostly unchanged if we use two, four, five, or six subsequent cohorts as our comparison group.

power to employ a regression discontinuity design. Second, there is detailed information about the day, month, and year of birth so we can identify the discontinuity induced by the policy within a narrow window.

The 1992 Census provides detailed information about the highest level of completed education for each respondent according to the following categories: none, primary, lower secondary (gymnasium), upper-secondary (measured separately as academic high school, vocational, and technical schools), and university or higher education. For simplicity, we impute years of schooling by assigning the number of years associated with each level of education.[15] This serves as our main measure of schooling when estimating the impact of the schooling expansion. The Census also has information on socio-economic characteristics of our respondents, such as gender, ethnicity, and region of birth. We use these variables to validate our research design. Furthermore, it contains information on labor force participation and occupational status (for those employed) as well as the fertility of women, which serve as useful ancillary outcomes. We can also reconstruct household composition to look at spousal schooling if spouses live in the same household.[16]

Panel A of Table 1 presents summary statistics for the individuals in cohorts born between 1945 and 1953. However, we also collapse the data to the day-of-birth level (3,240 observations) because that is the relevant unit of analysis for our regression discontinuity design. The average age at the time of the 1992 census is 42.5 years and the

[15] We also use data collected by the Romanian National Statistics Institute in 1995 and 1996 with reports of actual years of schooling (rather than educational attainment) in order to validate our imputed measure of years of schooling. These data come from surveys based on the 1994 World Bank's Living Standards Measurement Studies (LSMS) for Romania.

[16] Unfortunately, it is not feasible to examine the intergenerational transmission of education using the 1992 Census because (i) we do not observe completed education for all children, and especially for the youngest cohorts in our control years, (ii) we cannot measure schooling for children that no longer live with their parents, especially for the oldest cohorts in our treatment years, and (iii) the impacts on fertility raise concerns about changes in the composition of births.

fraction of female respondents is almost exactly half. Almost 90 percent of the sample is ethnic Romanian, with 7.7 percent ethnic Hungarians, and about 1.2 percent are Roma. The average years of schooling in our sample is 9.9, which is imputed based on the highest level of education completed. The employment rate is just below 85 percent while the average level of occupational skill is based on the classification suggested by the International Labour Office (ILO) ranging from 1 to 4.[17] Note that only 3 percent of the sample is unemployed, and most of the other non-employed individuals were housewives (women working in household and agricultural activities). The average number of children born to women in our sample is 2.3.

We use the 1994-2016 Vital Statistics Mortality files (VSM) to estimate the impact of the schooling expansion on mortality. These individual-level data cover the universe of deceased persons in Romania with detailed information on the day of birth/death and the main cause of death, as well as some socio-economic characteristics.[18] Thus, we can observe mortality for the cohorts used in our analysis between the ages of 41 and 71 by day and year of birth.[19] We compute mortality by day of birth as follows: (i) we sum the number of deaths at each day of birth from 1945 to 1953 over the period 1994-2016; (ii) we estimate the population at risk by calculating the number of people alive in 1992 at each day of birth from 1945 to 1953; then we take the ratio of (i) to (ii). This yields a mortality rate by day of birth which is at the finest level of our running variable.

Our calculation of the mortality rate could differ from the true mortality because of migration in and out of Romania. However, the number of migrants into and out of

[17] See more on the classification of occupational skill at: https://www.ilo.org/wcmsp5/groups/public/---dgreports/---dcomm/---publ/documents/publication/wcms_172572.pdf.
[18] The information on day of birth and death is from official records (death certificates, identity cards).
[19] Lleras-Muney (2005) and Clark and Royer (2013) suggest that the largest effects of education on mortality occur before the age of 64. Life expectancy in Romania was 69.5 years in 1994, 74.2 in 2011, and 75.5 years in 2016.

Romania was very small prior to 1992 because of closed borders during Communism. So the denominator described in (ii) above, based on the complete count in the 1992 Census, is likely to be an accurate measure of the population at risk. Moreover, the VSM files include all people deceased abroad as long as they still have a Romanian residence and/or citizenship. Therefore, our mortality files should account for the majority of the Romanian migrants abroad who are temporary emigrants and do not change their permanent residence.[20] Still, we will directly examine the potential for bias due to migration by checking whether the schooling expansion affects the probability of migration.

The VSM files also provide detailed information on the main cause of death (ICD codes) so we can look separately at deaths associated with circulatory diseases and cancer. These are the two most important causes of death in Romania, accounting for 44.6% and 26.5% respectively of all deaths. Following Meghir et al. (2018) we also classify diseases according to the epidemiological literature as preventable and treatable; preventable causes of death may reflect health behaviors while the treatable causes of death may be related to access to healthcare.[21]

Panel B of Table 1 shows the overall mortality rate and the mortality rate by category for our main sample. Approximately 25 percent of our sample died between 1994-2016. The largest category of deaths was associated with circulatory diseases which account for 10 percentage points, followed by cancer at 7.4 percentage points; preventable deaths accounted for 5.7 percentage points, while treatable diseases only for 3.8 percentage points.

[20] According to Statistics Romania these emigrants are the vast majority (over the 95%) of emigrants.

[21] We use the ICD 10 codes for defining cancer, circulatory diseases and treatable and preventable causes of death. See the Notes at the end of the Table 5 for more information.

We use the 1997-2017 National Inpatient Registers to calculate the number of nights spent in hospital care by day of birth. The National Inpatient register contains individual-level data on duration and ICD codes for all hospital stays in Romanian hospitals starting in January 1, 1997. Based on 7,892,000 hospital entries for our cohorts of interest, we calculate that individuals in our cohorts aged 54 to 72 spent an average of 24.7 days in hospital, as shown in Panel C of Table 1.

Finally, in the 2011 Romanian Census all respondents are asked whether they have any health-related problems that may affect their daily life (at work, school, at home, etc.). Thus, we can compute a measure of self-reported health for individuals who survived until 2011. Approximately 7.7 percent of people in our cohorts of interest reported having such problems. Those who answered affirmatively were given a set of six follow-up questions – whether they were (i) visually, (ii) hearing, or (iii) movement impaired, (iv) whether they had any memory or concentration problems, (v) self-care or (vi) difficulties in communication with their peers.

5. Empirical Strategy

We are primarily interested in the effect of education on mortality and other health outcomes, which can be expressed most simply as follows:

$$Y_i = \beta' X_i + \sigma S_i + \varepsilon_i \tag{1}$$

where Y_i is a measure of mortality or health for individual i, S_i is a measure of schooling, X_i represents observed individual characteristics, and ε_i represents unobserved factors (such as ability or motivation) that influence mortality and health. Since the unobserved factors may also be correlated with schooling and therefore bias our estimates of σ, we use the schooling expansion in Romania to generate exogenous variation in the level of schooling.

5.1 A regression discontinuity (RD) design

The schooling expansion that we study in this paper occurred over a five-year period from 1956 to 1961 and affected those born between 1945 and 1950. Since the government rapidly expanded access to schooling during this period, a child born just after the school entry cutoffs of January 1 of 1945, 1946, 1947, 1948, 1949, and 1950, would have benefited from the additional schools slots created by the government over the course of a year, as compared to a child born just before January 1 who would have been part of an earlier cohort. We can estimate these differences in schooling across successive cohorts during the period of schooling expansion using a regression discontinuity (RD) design:

$$S_i = \beta_0 + \alpha AFTER_i + f(day_i) + \varepsilon_i \qquad (2)$$

where S_i is our measure of (imputed) completed schooling for individual i, $AFTER_i$ is an indicator for individuals born just after January 1, and $f(day_i)$ is a parametric or non-parametric function of the day of birth which serves as our running variable. For simplicity, our preferred specifications do not include any covariates except for a constant β_0, although including them does not affect our results. We stack the discontinuities from 1945-1950 to estimate the average impact of the educational reforms for the affected cohorts. Thus, the coefficient on α is an estimate for the effect of being born just after the school entry cutoff on schooling; in other words, it represents the "first-stage" effect of the schooling expansion on completed schooling.

We can also estimate a version of equation (2) using our main outcome variables, such as mortality and health, as dependent variables. In this case, the corresponding estimates represent the "reduced-form" effects of the schooling expansion on mortality and health. Given that children born just before and after January 1 should have very similar background characteristics, we expect that our regression discontinuity design (if correctly specified) to yield causal estimates for the effect of the schooling expansion on these outcomes. If we also assume that the exclusion restriction holds, i.e. that being born after the school entry cutoff affects mortality and health only through years of schooling, the ratio of the reduced-form and first stage coefficients provides an estimate for the impact of education on mortality.

However, children born just after the school entry cutoff are generally the oldest children in their school cohort. Therefore, the exclusion restriction will not hold if relative age has an independent effect on health or mortality.[22] Our estimates may also be confounded by *school* cohort-specific shocks affecting health and mortality that are correlated with the increase in schooling generated by the schooling expansion. For example, if labor market conditions at entry affect later health and mortality, and these are improving over time, those born just after the school entry cutoff will benefit more than those who are born before it and enter the labor market earlier. Finally, as we document later, there may be some differential reporting of births around the first of every month (and especially around January 1), which cannot be addressed with a standard regression discontinuity design.

[22] See Bedard and Dhuey (2006) for a discussion on the long-term effects of the relative age effects induced by cutoff date for school eligibility. Cascio and Schanzenbach (2016) also provide evidence on the impacts of relative age in Tennessee while Black, Devereux and Salvanes (2011) estimate the effect of starting school younger in Norway.

5.2 A difference-in-regression discontinuities (D-RD) design

In order to account for a (stable) independent effect of relative age on health and mortality, to address certain confounding school cohort-specific shocks, and to deal with differential reporting around the January 1st cutoff, we use individuals who were born just before and after the school entry cutoff during 1951 and 1953, when lower-secondary education was already compulsory (and universal) for all the primary school graduates, as a comparison group (i.e. these cohorts form our "control years"). We can do this by estimating an analogous regression model to equation (2) for the discontinuities in the control years, and then comparing the impact of being born just after the school entry cutoff in treatment years to control years. This is our preferred specification and it can be estimated directly with the following "difference-in-discontinuities" (D-RD) regression model:

$$S_i = \beta_0 + \gamma AFTER_i + \tau TREAT_i + \delta AFTER_i * TREAT_i + f(day_i) + \varepsilon_i \qquad (3)$$

where $TREAT_i$ is an indicator for individuals born during years of schooling expansion 1945-1950, $AFTER_i$ is defined as before, and $f(day_i)$ now includes the interactions of our running variable with both $TREAT$ and $AFTER$, allowing for different relationships between the outcome and day-of-birth both before and after the threshold, and in the treatment and control periods. [23] In this specification, the coefficient on the interaction term, δ, yields the impact of being born just after the school entry cutoff during treatment years over and above the effect in control years that did not experience a compulsory schooling expansion, which is captured by γ. Given our model, the effect of being born

[23] This specification is similar to ones used by other recent papers which estimate a difference in RD discontinuities across cohorts. Grembi et al. (2016) provide a more formal presentation of the standard assumptions underlying this setting.

just after the school entry cutoff during the treatment years, coefficient α in equation (2), is equivalent to the sum of γ and δ.

As before, we can estimate a "reduced form" version of equation (3) for our main outcome variables. We can also take the ratio of these "reduced-form" and "first stage" coefficients to generate an estimate for the impact of education on mortality. Alternatively, we can estimate a two stage least squares (2SLS) model where we instrument for schooling with the interaction of AFTER*TREAT. The main identification assumptions underlying this empirical strategy are threefold: (i) that the relative age effects are stable across treatment and control years, and (ii) that school-cohort specific shocks, other than the schooling expansion, are balanced across treatment and control years, and (iii) that the misreporting of births is similar for both treatment and control discontinuities.[24]

We have no reason to expect the relative age effects to vary between treatment and control years. Nor are we aware of school-cohort specific shocks, other than the schooling expansion, that would affect school quality or entry into the labor market. As mentioned earlier, the labor market in Communist Romania was strictly controlled and highly regulated so there were few differences in labor market opportunities for students entering the labor market over short time horizons. Nevertheless, while we estimate and report 2SLS estimates, we do not consider them as our main specification because of the possibility that there were changes in school quality that are not captured by our imputed

[24] These factors can be made explicit by writing equation (1) as $Y_{ic} = \beta_0 + \eta B_i + \rho_c R_{ic} + \theta_c + \sigma S_i + \varepsilon_{ic}$ where Y_{ic} is now our outcome for individual *i* in cohort *c*, B_i are background characteristics, R_{ic} is relative age, and θ_c are school-cohort specific shocks. Our RD design assumes that B_i is smooth around the cutoff. Our D-RD design further assumes that ρ_c and $\theta_c - \theta_{c-1}$ (the differences in shocks across successive school cohorts) are similar between treatment and control years.

measure of years of schooling.[25] Instead, we follow Clark and Royer (2013) and Meghir et al. (2019) by focusing on the reduced-form effects of our school expansion policy.

5.3 The RD specifications

A key consideration when implementing a regression discontinuity design is the functional form of the forcing variable, $f(day_i)$. We estimate our impacts using local linear regressions with a triangular kernel as suggested by Hahn, Todd, and van der Klaauw (2001). While many recent papers adopt the Calonico, Cattaneo and Titiunik's (2014) optimal bandwidth procedure, this appears to yield excessively small bandwidths in our setting.[26] Instead, we present our findings for a fixed bandwidth of 180 days on either side of the cutoff for every specification, while plotting the RD estimates for our main outcomes over a broad range of alternative bandwidths (180 to 30 days).[27] We also confirm that our main results are robust to using parametric specifications that include higher order polynomials such as linear, quadratic and cubic trends in day of birth.

To avoid the issues associated with clustering on a discrete running variable, we collapse the data to the day of birth level and estimate our regressions with heteroskedastic-robust standard errors (Lee and Card, 2008; Kolesar and Rothe, 2018). Thus, as implemented in our regressions, the subscripts i in equations (1), (2), and (3) refer to *cohorts* of individuals born on a particular day of birth. However, we also verify that our results hold when we estimate our regressions at the individual-level.

[25] Relatedly, Stephens and Unayama (2019) discuss some issues when using instrumental methods with imputed endogenous variables. The recent review by Galama et al. (2020) also mentions that when using education reforms as IVs to study mortality and health the results are, with few exceptions, very imprecise.
[26] The estimates using CCT bandwidths generally suggest larger impacts of the schooling expansions on educational attainment and socio-economic outcomes but continue to indicate null effects on health and mortality. Thus, our choice of specification yields more conservative estimates than the CCT bandwidths.
[27] Our standard errors become extremely large for bandwidths below 30, especially in the donut specifications described below.

A common specification check for the regression discontinuity design is to verify that the density of observations is continuous around the cutoff (McCrary, 2008). When we examine the density, we find substantial heaping on January 1 and on some of the days immediately preceding it. This can be seen in Panels A and B of Appendix Figure 2, which plot the density around the January 1 cutoff (normalized to day 0) or the first week of January (normalized to week 0). We believe that this heaping is mainly due to delays in the reporting of births that occurred during the holiday period between Christmas and New Year's Day when government offices were closed.[28] Indeed, Panels C and D of Appendix Figure 2 show similar patterns for our control years.[29]

As mentioned earlier, we account for the issue of heaping around January 1 with our D-RD specifications that use both sets of years. Accordingly, there are less visible discontinuities in the density in Panels E and F which difference the impacts of the discontinuities in the control years from those in the treatment years.[30] We further attempt to deal with this issue using a "donut-RD" design as suggested by Barreca et al. (2016). In particular, our preferred specification drops individuals born within 7 days of January 1 in order to be symmetric around the cutoff (shown in light gray in the figure). We will also show our main results when dropping individuals born within 14 days of January 1.

A formal test for differences in density around the January 1 school cutoff is shown in column (1) of Appendix Table 1. This reveals no significant differences for either the

[28] In contrast to most other orthodox denominations, Christmas always remained on December 25 for the Romanian Orthodox. Consistent with this explanation, it appears the spike in observations occurs on January 2 in years when January 1 is a Sunday.

[29] Torun and Tumen (2016) document a similar pattern of heaping for January 1 in Turkey. Barreca et al. (2016) document some heaping at the beginning of each month in the California Vital Statistics records used by McCrary and Royer (2011), which we also observe in our data.

[30] Any remaining discontinuities in the density could be due to a change in the degree of heaping over time. Given our understanding of the institutional setting, as more children were born in clinics rather than at home, and as the state institutions created by the communist government expanded to rural areas, the correct reporting of the exact date of birth may have improved over time.

full sample in Panel A, or the 7-day and 14-day donut specifications in Panels B and C respectively. In columns (2)-(7), we examine whether the covariates available in the 1992 Census vary smoothly around the January 1 discontinuity by estimating equation (3) using these covariates as dependent variables. Among the covariates indicating gender, categories for the main ethnic groups in Romania, and an indicator for being born in Bucharest, only gender is statistically significant across the three specifications.[31]

6. Results

In presenting our results, we focus on the impact of the schooling expansion captured by δ, the coefficient on the interaction term of *AFTER*TREAT* in equation (3). This represents the effect of being born just after January 1 during the treatment years over and above the effect in control years that did not experience the schooling expansion. We also present the coefficient γ on *AFTER*, which represents the effect of being born just after the school entry cutoff during control years. As noted earlier, the effect of being born just after January 1 during the treatment years, coefficient α in equation (2), is equivalent to the sum of γ and δ.[32]

6.1 Effects on educational attainment

We begin by estimating, in Table 2, the impact of Romania's schooling expansion on years of completed schooling (column 1) and for specific educational categories (columns 2-5) based on the level of education recorded in the 1992 Census: completed primary, lower-secondary, upper-secondary, and higher education.[33] We report these

[31] To understand the extent to which this is a problem, we conduct a heterogeneity analysis by gender which reveals no significant results when we look at our main outcomes for males vs. females.
[32] The results from estimating equation (2) directly are available upon request.
[33] Appendix Table 3 uses the 1994-1996 LSMS household surveys to estimate the impact of the schooling expansion on reported years of schooling rather than an imputed measure based on completed

"first stage" results for our preferred bandwidth of 180 days around the January 1 cutoff. The estimates in Panel A, using the full sample, indicate that each successive cohort during the school expansion period 1945-1950 received an additional 1/5 of a year of schooling relative to later cohorts. Our preferred specification in Panel B, which excludes observations within 7 days of the January 1 cutoff, shows an increase of 1/9 of a year of schooling.

The increase in educational attainment is driven by an increase in the fraction of students who continue beyond the four years of primary school (column 2) to complete an additional three years of lower secondary education (column 3) with some students continuing even further onto upper-secondary schools (column 4).. The pattern of results is broadly similar across all three panels, including Panel C where we exclude observations within 14 days of the January 1 cutoff.[34]

Our significant effects on completed years of schooling are robust to alternative bandwidths. Panel A of Appendix Figure 6 plots the values of δ from equation (3) for bandwidths between 180 and 30. The range of these estimate is not altogether surprising given the large number of different specifications that we consider. Overall these estimates, while being quite precisely estimated, are smaller than those in Clark and Royer (2013) who show first stages between 2/5 and 1/2 of a year of schooling for the change in compulsory education in Britain, and those in Meghir et al. (2019) who show that the schooling reform in Sweden led to an increase of about 1/4 of a year of schooling.[35]

educational levels. However, one drawback with the LSMS data is that we cannot look at donut specifications. The result on years of schooling shows an increase of about 2/3 of a year of schooling, which is larger than in our specifications in Table 2.

[34] The coefficient on higher education in column (5) is positive and significant for the full sample in Panel A but, in contrast to all the other coefficients, it falls to zero in Panels B and C.

[35] When estimating our models using the optimal CCT bandwidths, we generate bandwidths that are smaller than 30 days with point estimates of 0.66 for the full sample and 0.34 for the 7-day donut.

We also present our "first stage" results graphically in Figure 3. Panels A, C and E plot average years of schooling by day of birth for individuals born six months before and after January 1st of each year; panels B, D and F plot the same data by week of birth, which makes it easier to discern the patterns. The graphs are normalized so that day 0 corresponds to January 1 and week 0 corresponds to the week of January 1 to January 7, and the fitted lines are based on linear spline regressions. Panels A to D show clear discontinuities after January 1 for the both the treatment and the control years. Lastly, Panels E and F use both treatment and control years in an attempt to estimate a version of equation (3) that differences out the impacts of the discontinuities in the control years from those in the treatment years. While these panels still show some outliers around the January 1 cutoff (shown in lighter gray), these are excluded from our 7- and 14- days donut specifications.[36]

6.2 Effects on socio-economic outcomes

While the main focus of this paper is on the effects of education on mortality and health later in life, it is useful to examine whether the expansion also had an impact on other relevant outcomes, such as labor market or household outcomes. These factors represent potential mechanisms for understanding the relationship between education and mortality (as discussed in Galama et al., 2020), and are interesting outcomes themselves.

Table 3 reports estimates for the impact of the schooling expansion on several socio-economic outcomes measured at the January 1992 Census. For labor market

[36] Note, we also examine the effects of Romania's schooling expansion on each education category recorded in the 1992 Census in Appendix Figure 3. While there is a clear reduction in the likelihood of completing primary school, the increases are spread throughout the higher levels of educational distribution. This is consistent with our understanding of Romania's schooling expansion, which required students to complete lower secondary education and opened up opportunities for further educational attainment.

outcomes, we consider the effects on employment and occupational skill. Column (1) of Panel A shows the estimates for the likelihood of being employed using the full sample, suggesting that each successive cohort during the schooling expansion was 0.4 percentage points more likely to be employed. The analogous estimates using the donut specifications in Panels B and C are smaller and insignificant. Similarly, column (2) reveals large increases in occupational skill that are significant in the full sample (Panel A) but not significant in the donut specifications (Panels B and C).

For household outcomes: Column (3) shows that Romania's schooling expansion had some significant effects on women's fertility. For the full sample in Panel A, we estimate that women who were affected by the schooling expansion had 0.03 fewer children. However, the estimated impacts become smaller and insignificant for the donut specifications in Panels B and C. There are also some significant effects on spousal schooling in the full sample, but these are also not robust when dropping observations close to the cutoff.[37]

We provide a graphical presentation of the impacts of the educational reform on employment, occupational skill, fertility, and spousal schooling in Panels A to D of Appendix Figure 4, which plot the difference between treatment and control discontinuities for these outcomes at the weekly level.[38]

Our estimated impacts of the schooling expansion on labor market and household outcomes indicate that the schooling expansion had some consequential impacts, but the

[37] This is also relevant for the possibility of intergenerational transmission of human capital since parental education has been shown to improve a broad range of child outcomes, including health and education (see e.g. Almond and Currie, 2011). Also note that we do not find different results when we look separately at our results by gender.

[38] We can also express the impacts on socio-economic outcomes as the effect of an additional year of schooling using a 2SLS framework where we instrument for schooling with the interaction of AFTER*TREAT. These are shown in columns (1) through (4) of Appendix Table 2, where most of the coefficients show significant effects in the full sample but insignificant effects in the 7-day and 14-day donut specifications.

effects are not robust across all samples. We consider four possible explanations for these findings. First, because we lack good data on income or wages, we have to rely on somewhat crude measures of labor market outcomes such as employment, in a setting where labor market participation was very high for both men and women. As a result, we have relatively little variation in these outcomes.[39]

Second, our cohorts graduated and entered a labor market characterized by a highly centralized wage-setting process, with small wage differentials and high employment. Similar to other communist leaders in the former Eastern Bloc, Ceausescu, promoted a policy of equalizing income and material wealth such that socio-economic conditions were similar across people.[40] Equality was one of the fundamental ideological tenets in socialist states and many individuals (especially at the lower end of the skill distribution) worked on collective agricultural farms and state industries. Therefore, it is not surprising that the effects of education on the probability of employment or the occupational skills are small in magnitude and therefore sensitive across different specifications.

Third, it is possible that treated cohorts gained few useful skills despite their additional years of schooling, similar to Pischke and Von Wachter (2008) for Germany. While we cannot measure the quality of schooling directly, there is no evidence of any major changes in the curriculum and educational standards across schools (Braham,

[39] We do use a set of household (LSMS) surveys collected during the mid-1990s to examine whether the impact of the school expansion affected income and find positive but insignificant effects (see Appendix Table 3). Note that, using the 1994-1994 and 1996 LSMS data, we find positive and significant impacts of the schooling expansion on employment. These estimates may be higher than in Table 2 because of the large fluctuations in Romania's labor market during early 1990s or because of specification issues in the LSMS where we only observe the month of birth.

[40] The transition from a centrally planned to a market system resulted in a major but gradual increase in the rates of return to education in the former Eastern Block: Munich et al. (2005) show that, in the Czech Republic, the rates of return to education reached the Western European levels in the period immediate after 1990.

1972). Moreover, as mentioned earlier, there were commensurate increases in the number of teachers, leaving the pupil-teacher ratios quite similar over time.

Fourth, the relatively weak impacts of the Romanian schooling expansion on labor market and household outcomes are in line with the broader literature on this topic. Galama et al's (2020) review notes that the estimates of the returns to schooling from educational expansions vary widely from null effects to large effects, depending on the country, the period when the reform took place, and the institutional and labor-market conditions at the time of the reform.[41]

6.3 Effects on mortality and health outcomes

This subsection examines whether the school expansion policy had an impact on mortality, our main outcome of interest in this paper. We focus on the mortality rate calculated from Vital Statistics data between 1994 and 2016, as described earlier. Column (1) of Table 4 reveals no evidence of a statistically significant effect of being born after the school cutoff of January 1 cutoff on mortality in any of the specifications. These estimates are remarkably stable across the three specifications. Given our preferred estimates using the 7- day donuts, we can rule out, with 95% confidence, that the schooling expansion reduced mortality by more than 0.4 percentage points between 1994-2016 when the average mortality rate was 25 percent.

The null effect on mortality is robust to alternative bandwidths. Panel B of Appendix Figure 6 plots the values of δ from equation (3) for bandwidths between 180 and 30. None of the estimates are statistically significant.[42] We also present a graphical

[41] Given the findings in Aaronson et al. (2017), we also examined the role of internal migration. However, we did not find significant effects of the schooling expansion on internal migration, measured as an indicator for whether the person lives in the locality of birth in 2011.

[42] As a result it is not surprising that our main results on mortality are also small and statistically insignificant when using optimal CCT bandwidths (see Calonico, Cattaneo and Titiunik, 2014).

analysis of the mortality results in Figure 4, which is structured similarly to the one for years of schooling. The patterns in Panels A-F provide a visual interpretation of the regression estimates from Table 4. We do not see evidence for large discontinuities in the mortality rate between 1994 and 2016 and, if anything, they point against the finding that education reduces mortality. In addition to the impact of the schooling expansion on mortality, we examine its effect on other measures of health that may affect quality of life: the total number of days spend in hospital from the 1997-2017 based on in-patient registers, and self-reported health measured as described in Section 4 from the 2011 Romanian Census. We do not find an impact of the schooling expansion on the number of days spent in hospital, as shown in column 2 of Table 4. Nor are there significant impacts on the number of hospitalizations, or on the duration of hospitalizations by specific cause such as cancer and circulatory diseases (results available on request).

In column (3) of Table 4, we show that there are no significant impacts of the schooling expansion on self-reported health problems among individuals who survived until 2011.[43] These null effects are estimated with substantial precision in all specifications. Given the standard errors in the 7-day donut specification, we can rule out with 95% confidence that the schooling expansion reduced the fraction of people with health related problems by more than 0.5 percentage points, which corresponds to 0.019 standard deviation units. We also explored the specific dimensions of health problems associated with our variable of self-reported health (i.e. vision, hearing, impaired movement, memory, self-care and communication) and did not find any meaningful impacts for these specific categories.

[43] Note that our self-reported health index is explicitly linked with the ability to work and thus with labour market outcomes. However, less than 40 percent of individuals who reported an impairment (hearing, visually, or movement impaired, memory or concentration problems, self-care or difficulties in communication with their peers) also ranked the gravity of their impairment as "great" or "complete insufficiency", which could arguable impact the ability to be active in the labor market.

Finally, we examine the effect of the schooling expansion on specific causes of death. First, we focus on mortality from the two most common causes of death in Romania: cancer and circulatory diseases. The regression estimates for these causes of death are shown in columns (1) and (2) of Table 5. Second, we classify certain causes of death as preventable or treatable, similar to Meghir et al. (2018) and show them in columns (3) and (4). These results do not indicate a consistent effect of the schooling expansion on specific causes of mortality.[44] Similarly, none of the corresponding graphs in Appendix Figure 5 show clear discontinuities around the regression discontinuity cutoffs. Thus, we do not find any more evidence for the impact of the schooling expansion on specific causes of death than on the mortality rate as a whole.

We can express the impacts on mortality and health outcomes as the effect of an additional year of schooling using a two-sample 2SLS framework, where we instrument for schooling with the interaction of AFTER*TREAT following Inoue and Solon (2010). These are shown in columns (1) through (3) of Table 6. Our preferred estimates for the 7-day donut specifications suggest that we can rule out, with 95% confidence, that an additional year of schooling reduced mortality by more than 1.2 percentage points respectively. Similarly, we can rule out that an additional year of schooling reduced self-reported health problems by more than 0.19 of a standard deviation based on the 7-day donut specifications.

6.4 Robustness

We consider a number of robustness checks of our main results. These include the following: (i) we only consider cohorts born 1949-1952 that are relatively similar in age;

[44] There are some significant (positive) effects for mortality from circulatory diseases.

(ii) we drop the 1945 cohort due to the possibility of being affected in utero by WWII; (iii) we separately examine mortality between 1994-2005 and 2006-2016; (iv) we look at mortality rates for people aged 52-62 only; (v) using alternative polynomial functions of our running variable, instead of local linear regressions, and (vi) using the uncollapsed individual-level data. Our findings remain qualitatively similar in each of these alternative specifications. With the exception of the parametric specifications shown in Appendix Table 4 and the individual-level specifications shown in Appendix Table 5, these results are available upon request. We also considered a placebo test in which we looked for discontinuities around July 1 rather than January 1 in treatment and control years. There were no significant effects for years of schooling, socio-economic outcomes, or on our measures of health and mortality.

To address concerns about bias due to migration, we consider whether our school expansion directly affected the probability of external migration. The 2011 census contains information on all persons who migrated abroad for a period of at least 12 months (at the time of the census). Hence, the vast majority of the Romanian emigrants are covered i.e., all individuals working abroad who maintain their houses, identity cards or/and remain registered by the Romanian administrative bodies.[45] Using a similar strategy as before, column (1) of Appendix Table 6 confirms that the effects are similar for both the treatment and control years. Thus, there is no overall impact of the schooling expansion on the likelihood that individuals (who survived until 2011) have emigrated.

The migration results presented above, while reassuring, are not able to capture the possible effects of the schooling expansion on permanent migration. We address this possibility through an indirect test. Using information from the 1992 and 2011 census

[45] According to Statistics Romania, 95% of Romanian emigrants are temporary migrants, who keep their Romanian IDs, and whose death is reported in the Romanian Mortality Files.

samples, we calculate the number of people born in a given day who are in the 2011 census as a fraction of the number in the 1992 census. This ratio should capture a combination of both mortality and migration between 1992-2011. These results are presented in column (2) of Appendix Table 6, and confirm that there is no impact of the school expansion on this combined measure of mortality and migration.

6.5 Discussion

Our findings indicate that the Romanian schooling expansion led to substantial increases in educational attainment but did not improve health or reduce mortality. We estimate null effects on health and mortality measured up to the age of 71 that are reasonably precise. In this section, we attempt to understand and explore the mechanisms underlying these findings.

One potential channel through which education can affect mortality and health is by increasing income and thereby enabling individuals to purchase more health services or better health insurance. While we observe some positive impacts on employment (in the 1992 Census) and income (in the LSMS data), these effects are relatively small and sensitive to alternative specifications. As we discussed in section 6.2, this is likely because of the low labor market returns in Communist Romania, and could explain why, in our setting, the educational expansion did not translate into health and mortality effects. However, studies in other European countries also show that education expansions had little to no health returns, despite more flexible labor markets. One possible explanation is that, compared to the U.S., most European countries, including Romania, have had universal public healthcare so that income differences are potentially less relevant (see also Galama et al., 2020).

Even if more education would have led to better labor market outcomes, the impact of income on health would not necessarily be positive (through access to better health care). For example, more income may lead to an increased consumption of unhealthy goods, such as alcohol and smoking.[46] While using the LSMS survey we find positive and significant correlations between education and smoking, when we attempted to estimate our regression discontinuity specifications using this data, we find no significant effects of education on smoking behavior (see Appendix Table 3).[47]

It is also possible that schooling expansion could affect mortality and health through other channels. For example, we find some evidence that Romania's schooling expansion may have increased occupational skill, although these findings are also sensitive to alternative specifications. Whether this should have led to improved health is not completely clear.[48] Similarly, we also observe some positive impacts of the schooling expansion on non-labor outcomes. We show in some specifications that women with more education have fewer kids. This could have a positive impact on the mother's health if it leads to more resources (and more information, via e.g. employment) at the household level. However, more active mothers could be more stressed which may negatively impact their health.

Another important channel is skill formation in school. This is especially relevant when an educational expansion affects a lot of individuals, as in Romania, which may

[46] In Romania the alcohol consumption was not dependent on income because large amounts of spirits and wine were commonly produced and consumed in households; beverages are often used for family consumption and trade in small agricultural markets.
[47] Specifically, we use the 2001-2009 LSMS survey data that is nationally representative, and covers about 30,000 households each year and contains detailed socio-economic information on all household members. Note that the LSMS data does not have the day of birth, but only the month and year and therefore we cannot show the donuts specifications.
[48] Moreover, skilled occupations may imply better peers, better working conditions, and higher social status which might have a positive impact on health. At the same time, some more skilled occupations may be associated with more stress than certain less skilled occupations or it is possible that some relatively skilled manufacturing jobs may have worse working conditions than jobs in the informal sector such as agriculture; these might have a negative impact on health.

result in lower quality or in a different style of instruction (see Galama et al., 2020). Measuring school quality directly is extremely difficult in our setting. However, in Communist Romania the curriculum and syllabuses were centrally prepared and distributed across schools. Moreover, the curriculum of the general schools was standard and suffered no major changes for our cohorts (Braham, 1972). Lastly, as we discussed in Section 3, during the period we study here, the number of employed teachers also increased a lot, especially for the general schools, leaving the pupil-teacher ratios quite similar across years.

Finally, access to more education could lead to improvements in health and longevity via channels that are not linked directly with standard socio-economic outcomes such as monetary resources or fertility; for example non-cognitive skills, peers, social norms, social status and increased informal connections could all impact health. Given the data we have available, it is very difficult for us to test for these channels.

Thus, while our analysis does not yield any strong conclusions about the role of particular mechanisms in explaining our results, we believe that the null effects of more schooling on health and mortality may be due to both the limited effects of the schooling expansion on labor market outcomes and the free access to health care in Romania, consistent with the arguments that this relationship depends on country-specific institutional and political conditions (see Galama et al., 2020).

7. Conclusion

This paper analyzes a schooling expansion in Romania, which aimed to ensure that all students received at least 7 years of compulsory schooling. The schooling expansion affected five consecutive cohorts born between 1945 and 1950 and opened up opportunities for further educational attainment. We use a "difference in regression discontinuity" (D-RD) design to estimate impacts by comparing the impacts of the schooling expansion in successive cohorts of affected individuals with later cohorts who were entered school after the expansion was completed. We find that beginning school in a (one year) later cohort increases educational attainment by approximately 1/9 to 1/5 of a year of schooling. We do not find any consistent significant impacts of the schooling reform on mortality, hospitalizations, or self-reported health. Moreover, we can rule out that the schooling expansion reduced mortality by more than 0.4 percentage points between 1994 and 2016, or that it reduced self-reported health problems as measured in 2011 by more than 0.019 standard deviation units for the full sample of individuals in the affected cohorts.

Whether education causally affects health and mortality is an important question for both developed and developing countries alike. However, most of the previous work has focused on the United States and Western Europe. The findings in this literature are mixed and we lack strong evidence about whether education significantly improves health or decreases mortality. We extend the literature by estimating causal impacts for a population that is substantially poorer and that experienced changes at a lower margin of educational attainment. Our findings indicate the absence of a causal effect of education health and mortality, even in this setting. While we attempted to examine the underlying mechanisms for these findings, more work needs to be done to better understand why we do not observe a strong relationship between education and health across a variety of different settings.

References

Aaronson D., B. Mazumder, S.G. Sanders, and E. Taylor (2017) "Estimating the Effect of School Quality on Mortality in the Presence of Migration: Evidence from the Jim Crow South". SSRN working paper.

Albouy, V. and l. Lequien (2009) "Does Compulsory education lower mortality?" Journal of Health Economics, 28(1): 155-168.

Almond D. and J. Currie (2011) "Human capital development before age five. In: Ashenfelter O, Card D (eds) Handbook of labor economics, chap 15.

Andrén, D and Andrén, T (2015) "Gender and occupational wage gaps in Romania: from planned equality to market inequality? IZA Journal of European Labor Studies, 4 (10)

Arendt, J.N. (2005) "Does education cause better health? A panel data analysis using school reforms for identification." Economics of Education Review, 24(2):149 –160.

Barreca, A., J. Lindo, and G. Waddel (2016) "Heaping- Induced Bias in Regression-Discontinuity Designs" Economic Inquiry, 54(1): 268-293.

Bedard, K and E. Dhuey (2006) "The Persistence of Early Childhood Maturity: International Evidence of Long-Run Age Effects" The Quarterly Journal of Economics, 121(4): 1437-1472.

Black, D. A., Hsu, Y. C. & Taylor, L. J. (2015). "The effect of early-life education on later-life mortality". *Journal of Health Economics*, 44, 1-9.

Black, S.E., P.J. Devereux, and K.G. Salvanes (2011) "Too young to leave the nest? The effects of school starting age". *Review of Economics and Statistics* 93(2):455–467

Braham, R. L. (1963) Education in the Rumanian People's Republic. Washington, D.C : U.S. Dept. of Health, Education, and Welfare, Office of Education.

Braham, R.L., (1972). Education in Romania: A Decade of Change. US Government Printing Press.

Buckles, K., A. Hagemann, O. Malamud, M. Morrill, and A. Wozniak (2016) "The effect of college education on mortality" *Journal of Health Economics,* 50: 99-114.

Cascio, E.U. and D. W. Schanzenbach (2016) "First in the Class? Age and the Education Production Function" *Education Finance and Policy* 11(3): 225-250.

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

Chou, S., J. Liu, M. Grossman and T. Joyce, 2010. "Parental Education and Child Health: Evidence from a Natural Experiment in Taiwan," *American Economic Journal: Applied Economics*, 2(1), 33-61.

Clark, D. and H. Royer (2013) “The effect of education on adult mortality and health: Evidence from Britain” *The American Economic Review*, *103*(6), 2087-2120.

Davies, N. M., Dickson, M., Smith, G. D., Van den Berg, G. & Windmeijer, F. (2016) “The causal effects of education on health, mortality, cognition, well-being, and income in the UK Biobank”. bioRxiv 074815. Preprint.

De Walque, D. (2007) “Does education affect smoking behaviors? Evidence using the Vietnam draft as an instrument for college education” *Journal of Health Economics* 26 (5), 877–895.

Dursun B., Cesur R., and N. Mocan (2018) “The Impact of Education on Health Outcomes and Behaviors in a Middle-Income, Low-Education Country” *Economics Human Biology* 27 (31):94-114.

Earle, J. S., and C. Pauna, “Incidence and Duration of Unemployment in Romania,” *European Economic Review* 40 (1996), 829–837.

Filipescu, V. and Oprea O. (1972) Invatamantul obligatoriu in Romania si in alte tari. Editura Didactica si Pedagogica, Bucuresti

Galama, T.J., A. Lleras-Muney, H. van Kippersluis (2018) “The Effect of Education on Health and Mortality: A Review of Experimental and Quasi-Experimental Evidence”. *NBER Working Paper* No. 24225

Giurescu, C, I Ivanov and N. Mihaileanu, editors (1971) Istoria învăţământului din România : compendiu. Editura Didactică şi Pedagogică, Bucuresti

Grembi, V., T. Nannicini, and U. Troiano (2016) “Do Fiscal Rules Matter?” *American Economic Journal: Applied Economics* 8(3): 1-30

Grimard, F. and D. Parent (2007) “Education and smoking: Were Vietnam war draft avoiders also more likely to avoid smoking?” *Journal of Health Economics,* 26 (5), 896–926.

Grossman, M. (2006) "Education and Nonmarket Outcomes" in the *Handbook of the Economics of Education*, Elsevier. Vol. 1:577-633

Hahn, J., Todd, P. and van der Klaauw (2001) “Identification and Estimation of Treatment Effects with a Regression-Discontinuity Design” *Econometrica*, 69 (1): 201-209.

Imbens, G. and K. Kalyanaraman (2012) “Optimal Bandwidth Choice for the Regression Discontinuity Estimator” Review of Economic Studies 79, 933–959.

Inoue A. and G. Solon (2010) “Two-Sample Instrumental Variables Estimators” The Review of Economics and Statistics, 92(3): 557-561

Kan, K. (2016) “The Impact of Education on Mortality: Evidence from a Compulsory Education Reform,” mimeo.

Kim, J. (2016) Female education and its impact on fertility. IZA World of Labor: 228 doi: 10.15185/izawol.228

Lee, D.S., and D. Card (2008) “Regression discontinuity inference with specification error” Journal of Econometrics, 142, 655-674.

Lleras-Muney, A. (2005) “The Relationship Between Education and Adult Mortality in the United States” *Review of Economic Studies,* 72, 189-221.

Mazumder, B. (2008) “Does Education Improve Health: A Reexamination of the Evidence from Compulsory Schooling Laws” *Economic Perspectives*, 33(2), 2-16.

Mazumder, B. (2012) “The effects of education on health and mortality” *Nordic Economic Policy Review* No. 1, 261-302.

McCrary, J., and H. Royer. 2011. "The Effect of Female Education on Fertility and Infant Health: Evidence from School Entry Policies Using Exact Date of Birth." *American Economic Review*, 101(1): 158–95.

Meghir, C., M. Palme and E. Simeonova (2018) “Education, Health and Mortality: Evidence from a Social Experiment” *American Economic Journal: Applied Economics* 10(2): 234-96

Munich, D., J. Svejnar and K. Terrell (2005) “Returns to Human Capital under the Communist Wage Grid and during the Transition to a Market Economy” The Review of Economics and Statistics 87(1): 100-123.

Oreopoulos, P. (2006) “Estimating Average and Local Average Treatment Effects of Education when Compulsory School Laws Really Matter” *American Economic Review* 96(1), 152-175.

van Kippersluis, H. O. O’Donnell and E. van Doorslaer (2011) “Long-Run Returns to Education: Does Schooling Lead to an Extended Old Age?” *Journal of Human Resources* 46(4): 695–721.

Table 1: Summary Statistics

	Individual-level			Collapsed to day-of-birth level		
	Mean	S.D.	Obs	Mean	S.D.	Obs
Panel A: 1992 Census data						
Female	0.501	0.500	2,607,116	0.503	0.023	3,240
Age	42.155	2.564	2,607,116	42.517	2.630	3,240
Ethnicity						
Romanian	0.896	0.305	2,607,116	0.894	0.019	3,240
Hungarian	0.075	0.263	2,607,116	0.077	0.016	3,240
Roma	0.012	0.110	2,607,116	0.012	0.005	3,240
Other	0.017	0.128	2,607,116	0.017	0.005	3,240
Years of schooling	9.974	3.643	2,598,587	9.905	0.614	3,240
Employed	0.846	0.361	2,607,116	0.844	0.023	3,240
Occupational skill	2.363	0.758	2,271,173	2.361	0.053	3,240
Number of children	2.286	1.607	1,304,914	2.288	0.106	3,240
Spouse's years of schooling	9.928	3.587	2,226,729	9.872	0.531	3,240
Panel B: Vital Statistics Mortality (1994-2016)						
Overall mortality rate	-	-	-	0.249	0.046	3,240
Mortality rate by category						
Cancer	-	-	-	0.074	0.015	3,240
Circulatory	-	-	-	0.100	0.028	3,240
Preventable	-	-	-	0.057	0.010	3,240
Treatable	-	-	-	0.038	0.010	3,240
Panel C: National Inpatient Registers (1997-2017)						
Time hospitalized (days)	-	-	-	24.764	2.797	3,240
Panel D: 2011 Census data						
Self-reported health index	0.076	0.265	2,028,307	0.077	0.013	3,240

Notes: Individual observations for cohorts born from mid-1944 to mid-1953 collapsed to the day-of-birth level from the 1992 Romanian Census (100%), 2011 Romanian Census (100%), In-patient registry data and Romania VSM files. Female, ethnicity categories, and employment status are coded as indicator variables so their means can be interpreted as fractions. Years of schooling is a summary measure imputed from levels of completed education. Number of children is only available for women and includes those with 0 children. Occupational skill (1-4) is based on International Labour Office (ILO) categories. Self-reported health is an indicator for any health-related problem that affects daily life at work, school, or at home. Mortality is the fraction of deaths between ages 42-71, computed by summing number of deaths in each day of birth and dividing by the corresponding number of people alive in 1992. Time hospitalized is based on in-patient days spent between ages 54-72.

Table 2: Effects of Educational Expansion on Schooling Outcomes

Outcome	Years of schooling (1)	Primary (2)	Lower secondary (3)	Upper secondary (4)	Higher education (5)
Panel A: Full sample					
AFTER	0.008	-0.000	-0.011***	0.017***	-0.006***
	[0.021]	[0.002]	[0.003]	[0.003]	[0.001]
AFTER*TREAT	0.192***	-0.021***	0.006*	0.010***	0.006***
	[0.033]	[0.003]	[0.003]	[0.004]	[0.002]
Sample size	3,240	3,240	3,240	3,240	3,240
R-squared	0.899	0.951	0.361	0.900	0.341
Panel B: 7-day donut sample					
AFTER	-0.038*	0.003**	-0.008***	0.011***	-0.006***
	[0.019]	[0.001]	[0.003]	[0.003]	[0.002]
AFTER*TREAT	0.116***	-0.017***	0.007**	0.009**	0.000
	[0.028]	[0.002]	[0.003]	[0.004]	[0.002]
Sample size	3,114	3,114	3,114	3,114	3,114
R-squared	0.918	0.959	0.368	0.906	0.368
Panel C: 14-day donut sample					
AFTER	-0.060***	0.005***	-0.008***	0.008***	-0.006***
	[0.020]	[0.001]	[0.003]	[0.003]	[0.002]
AFTER*TREAT	0.088***	-0.015***	0.008**	0.007*	-0.000
	[0.029]	[0.002]	[0.004]	[0.004]	[0.002]
Sample size	2,988	2,988	2,988	2,988	2,988
R-squared	0.921	0.960	0.379	0.909	0.370

Notes: Individual observations for cohorts born from mid-1944 to mid-1953 in the 1992 Census collapsed to the day-of-birth level. Outcome variables are defined as in Table 1. AFTER is an indicator that equals 1 for individuals born after January 1. TREAT is an indicator that equals 1 for cohorts who experienced the education expansion during 1945-1950. Local linear regressions use triangular kernels and include saturated controls for day-of birth interacted with AFTER and TREAT. All specifications use a bandwidth of 180 days. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively.

Table 3: Effects of Educational Expansion on Socio-Economic Outcomes

Outcome	Employment (1)	Occupational skill (2)	Fertility (3)	Spouse's schooling (4)
Panel A: Full sample				
AFTER	0.004**	-0.013***	0.019*	-0.010
	[0.002]	[0.004]	[0.011]	[0.021]
AFTER*TREAT	0.004*	0.022***	-0.033**	0.061*
	[0.003]	[0.006]	[0.015]	[0.031]
Sample size	3,240	3,240	3,240	3,240
R-squared	0.576	0.546	0.291	0.877
Panel B: 7-day donut sample				
AFTER	0.002	-0.016***	0.015	-0.038**
	[0.002]	[0.004]	[0.012]	[0.019]
AFTER*TREAT	0.002	0.007	-0.015	-0.001
	[0.003]	[0.006]	[0.015]	[0.026]
Sample size	3,114	3,114	3,114	3,114
R-squared	0.587	0.585	0.292	0.896
Panel C: 14-day donut sample				
AFTER	-0.001	-0.015***	0.007	-0.015
	[0.002]	[0.005]	[0.012]	[0.021]
AFTER*TREAT	0.001	0.002	0.009	-0.060**
	[0.003]	[0.006]	[0.016]	[0.029]
Sample size	2,988	2,988	2,988	2,988
R-squared	0.595	0.588	0.312	0.897

Notes: Individual observations for cohorts born from mid-1944 to mid-1953 in the 1992 Census collapsed to the day-of-birth level. Outcome variables are defined as in Table 1. AFTER is an indicator that equals 1 for individuals born after January 1. TREAT is an indicator that equals 1 for cohorts who experienced the education expansion during 1945-1950. Local linear regressions use triangular kernels and include saturated controls for day-of birth interacted with AFTER and TREAT. All specifications use a bandwidth of 180 days. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively.

Table 4: Effects of Educational Expansion on Mortality and Health Outcomes

Outcome	Mortality	Hospitalizations	Self-reported Health
	(1)	(2)	(3)
Panel A: Full sample			
AFTER	0.002	0.737**	-0.000
	[0.002]	[0.352]	[0.002]
AFTER*TREAT	0.004	-0.304	-0.001
	[0.003]	[0.451]	[0.002]
Sample size	3,240	3,240	3,240
R-squared	0.771	0.065	0.310
Panel B: 7-day donut sample			
AFTER	0.005*	1.216***	0.000
	[0.002]	[0.374]	[0.001]
AFTER*TREAT	0.004	-0.510	-0.001
	[0.004]	[0.477]	[0.002]
Sample size	3,114	3,114	3,114
R-squared	0.779	0.070	0.322
Panel C: 14-day donut sample			
AFTER	0.006**	1.338***	0.002
	[0.003]	[0.387]	[0.002]
AFTER*TREAT	0.002	-0.655	-0.002
	[0.004]	[0.511]	[0.002]
Sample size	2,988	2,988	2,988
R-squared	0.780	0.067	0.315

Notes: Individual observations for cohorts born from mid-1944 to mid-1953 in the 1992 Census collapsed to the day-of-birth level. Outcome variables are defined as in Table 1. AFTER is an indicator that equals 1 for individuals born after January 1. TREAT is an indicator that equals 1 for cohorts who experienced the education expansion during 1945-1950. Local linear regressions use triangular kernels and include saturated controls for day-of birth interacted with AFTER and TREAT. All specifications use a bandwidth of 180 days. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively.

Table 5: Effects of Educational Expansion on Mortality by Cause of Death

Outcome	Cancer	Circulatory	Preventable	Treatable
	(1)	(2)	(3)	(4)
Panel A: Full Sample				
AFTER	0.002*	0.002	0.001	0.000
	[0.001]	[0.001]	[0.001]	[0.001]
AFTER*TREAT	-0.001	0.004*	-0.001	0.001
	[0.002]	[0.002]	[0.002]	[0.001]
Sample size	3,240	3,240	3,240	3,240
R-squared	0.483	0.777	0.176	0.437
Panel B: 7-day donut sample				
AFTER	0.003**	0.002	0.001	0.000
	[0.001]	[0.002]	[0.001]	[0.001]
AFTER*TREAT	-0.002	0.006***	-0.001	0.001
	[0.002]	[0.002]	[0.002]	[0.001]
Sample size	3,114	3,114	3,114	3,114
R-squared	0.489	0.781	0.181	0.437
Panel C: 14-day donut sample				
AFTER	0.004***	0.003**	0.001	0.000
	[0.001]	[0.002]	[0.001]	[0.001]
AFTER*TREAT	-0.003	0.004*	-0.001	0.002
	[0.002]	[0.002]	[0.002]	[0.001]
Sample size	2,988	2,988	2,988	2,988
R-squared	0.494	0.782	0.178	0.438

Notes: Individual observations for cohorts born from mid-1944 to mid-1953 in the 1992 Census collapsed to the day-of-birth level. Outcome variables are defined as in Table 1. AFTER is an indicator that equals 1 for individuals born after January 1. TREAT is an indicator that equals 1 for cohorts who experienced the education expansion during 1945-1950. Local linear regressions use triangular kernels and include saturated controls for day-of birth interacted with AFTER and TREAT. All specifications use a bandwidth of 180 days. We use the ICD 10 codes for defining cancer (Chapter C) and circulatory diseases (Chapter I). Treatable causes of death (cf. ICD 10) include: Tuberculosis (A15-A19. B90), Malignant neoplasm of cervix uteri (C53); Chronic rheumatic heart disease (I05-I09); All respiratory diseases (J00-J99); Asthma (J45, J46); Appendicitis (K35-K38); Abdominal hernia (K40-K46); Hypertensive and cerebrovascular disease (I10-I15, I60-I69); Chollelthiasis and cholecystitis (K80-K81). Finally, the preventable causes of death include: Lung cancer (C33-C34), Cirrhosis of liver (K70, K74.3-K74.6), and External causes of death (V, W, X, Y). Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively.

Table 6: 2SLS estimates of Years of Schooling on Mortality and Health Outcomes

Outcome	Mortality (1)	Hospitalizations (2)	Self-reported Health (3)
Panel A: Full Sample			
Years of schooling	0.033* [0.020]	-2.274 [2.344]	-0.011 [0.010]
Sample size	3,240	3,240	3,240
Panel B: 7-day donut sample			
Years of schooling	0.048 [0.030]	-4.336 [3.635]	-0.020 [0.015]
Sample size	3,114	3,114	3,114
Panel c: 14-day donut sample			
Years of schooling	0.045 [0.038]	-5.796 [4.665]	-0.027 [0.020]
Sample size	2,988	2,988	2,988

Notes: Individual observations for cohorts born from mid-1944 to mid-1953 in the 1992 Census collapsed to the day-of-birth level. Outcome variables are defined as in Table 1. 2SLS regressions instrument for years of schooling with the interaction of AFTER*TREAT, control for AFTER, and include saturated controls for day-of birth interacted with AFTER and TREAT. Columns (1)-(8) report standard 2SLS estimates while columns (9)-(12) use report estimates from two-sample 2SLS. All specifications use a bandwidth of 180 days. Heteroskedacticity-robust standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively.

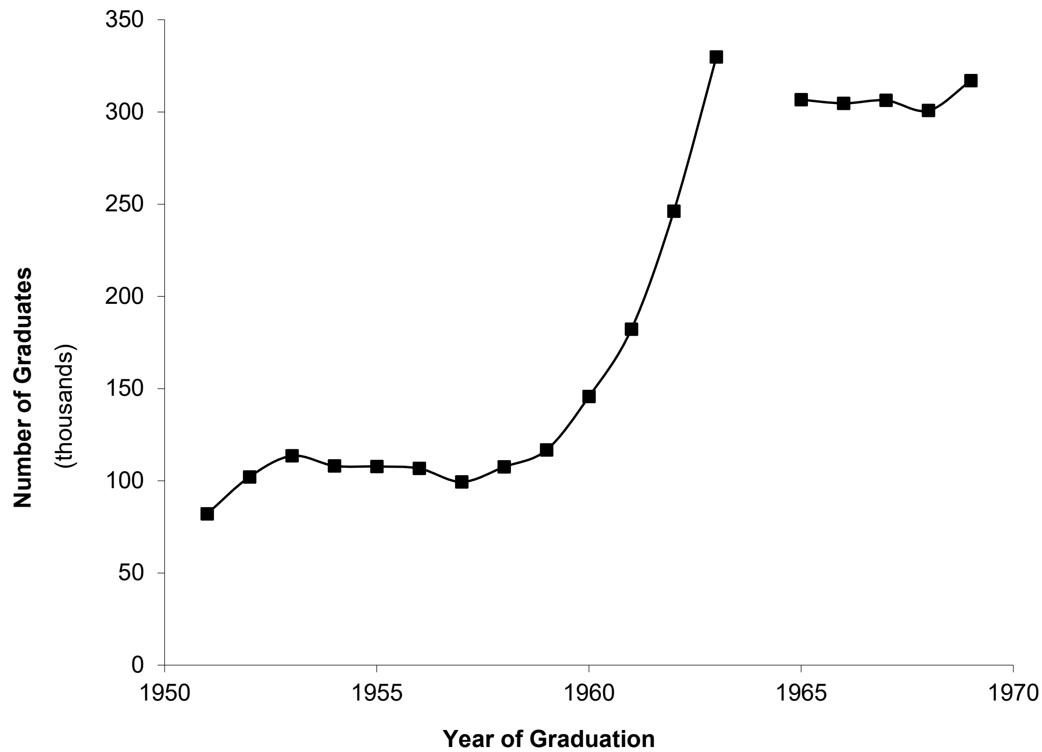


Figure 1: Graduates from Lower Secondary (Gymnasium) Schools by Year of Graduation. Notes: Figure 1 plots the number of students graduating from 7 year gymnasium between 1951 and 1969. Source: Romanian Statistical Yearbook

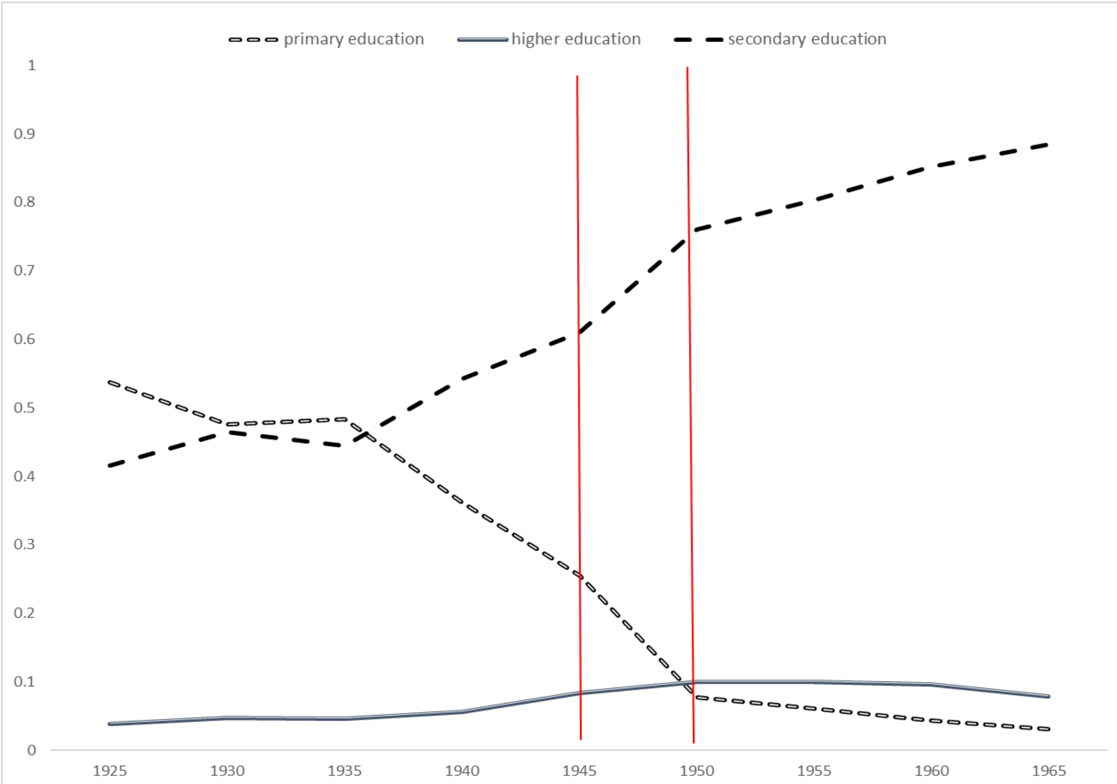


Figure 2: Educational attainment in Romania by year of birth
Notes: Figure 2 plots the highest educational attainment by year of birth for cohorts of individuals. Source: 1992 Romanian Census (complete count).

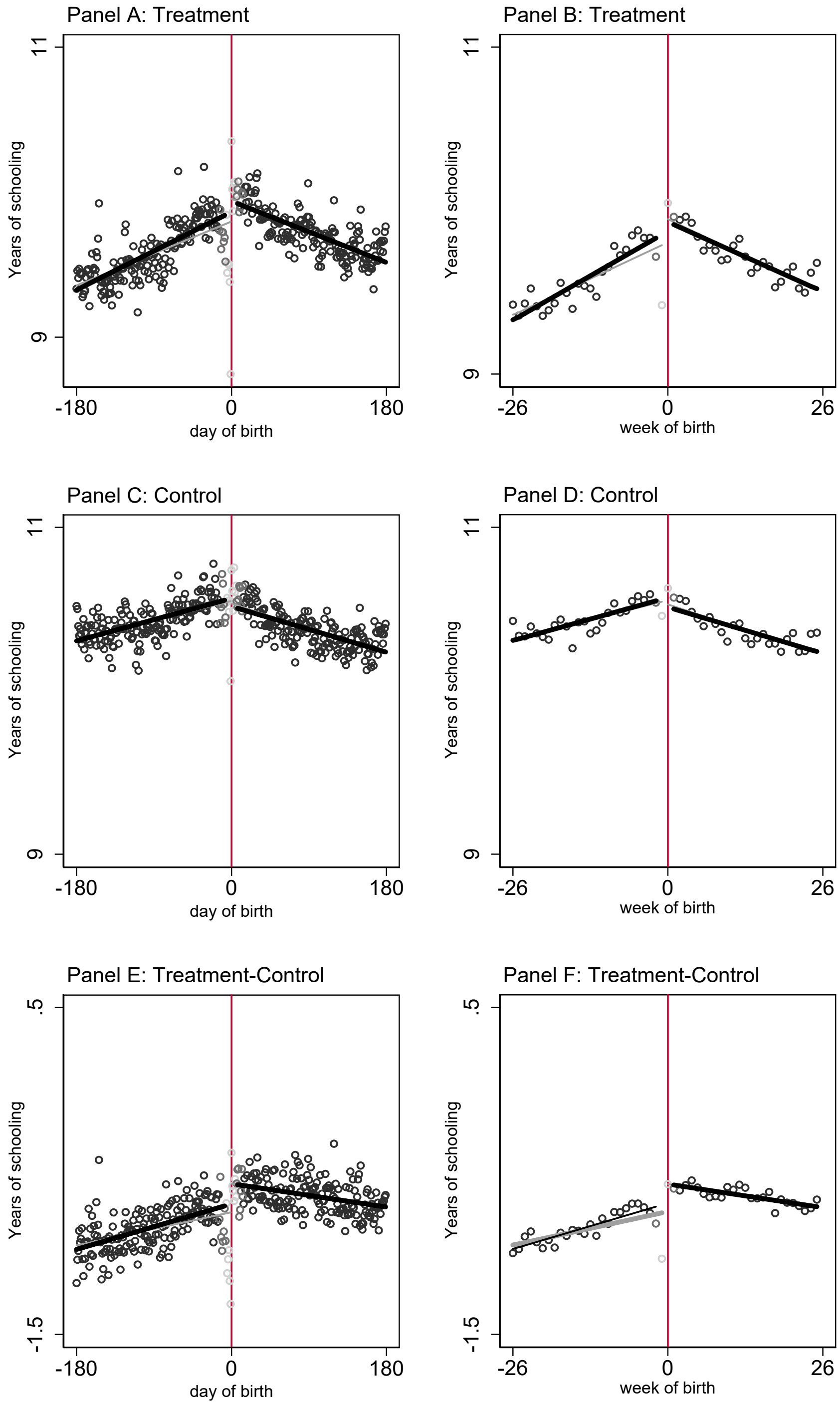


Figure 3: Regression discontinuity (RD) plots of Years of Schooling

Notes: Panels A and B are restricted to individuals born in the treatment years (1944-1950). Panels C and D are restricted to individuals born in the control years (1950-1953). Panels E and F are restricted to individuals born in both treatment and control years (1944-1953). The open circles indicate the mean of the outcome by day of birth (panels A, C and E) or week of birth (panels B, D and F). The solid lines are based on linear spline regressions. The light gray lines and circles show our preferred 7-day donut specification (when we drop individuals born within 7 days of January 1 in order to be symmetric around the cutoff). Source: 1992 Romanian Census (complete count).

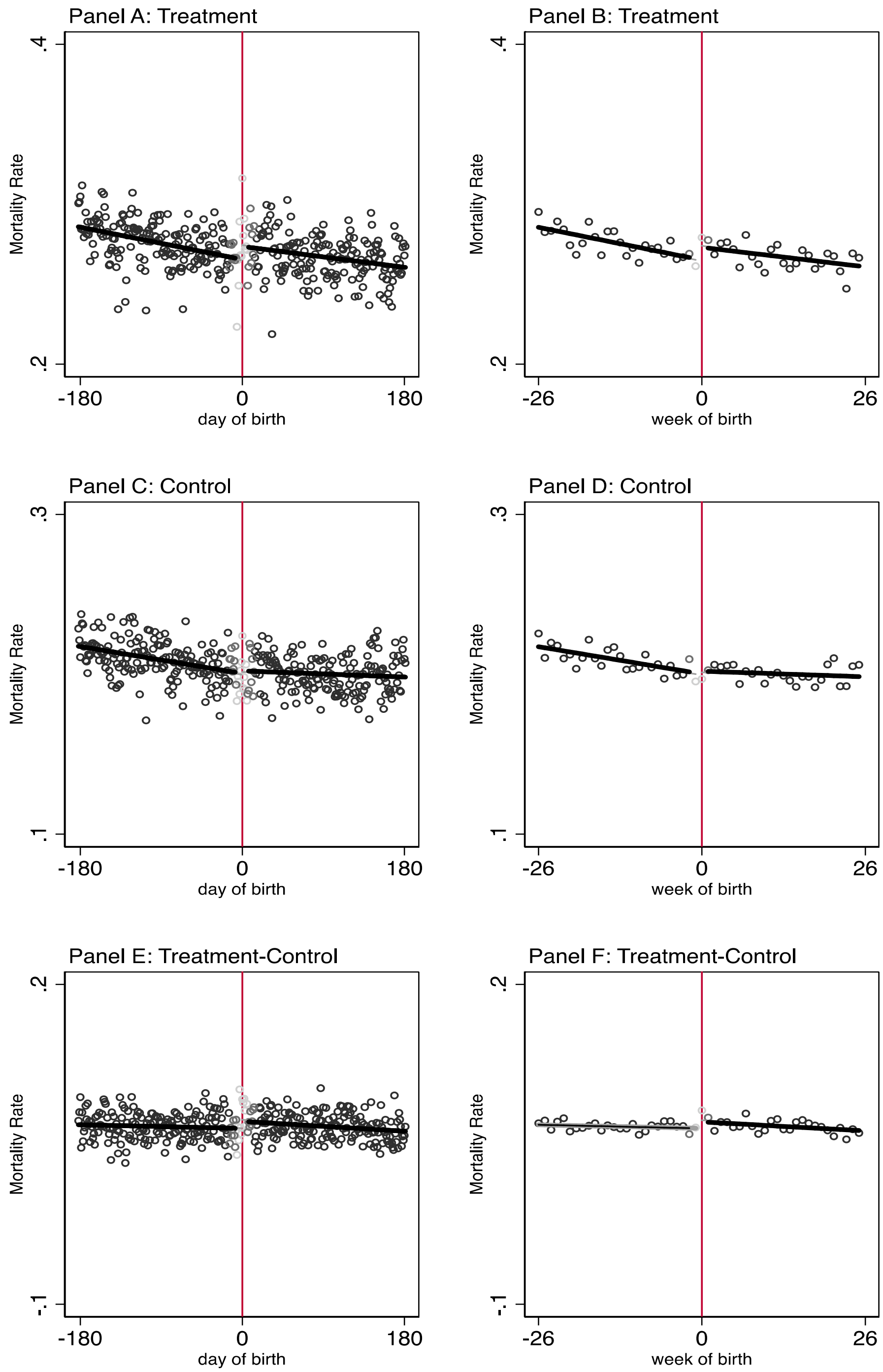


Figure 4: Regression discontinuity (RD) plots of the Mortality Rate

Notes: Panels A and B are restricted to individuals born in the treatment years (1944-1950). Panels C and D are restricted to individuals born in the control years (1950-1953). Panels E and F are restricted to individuals born in both treatment and control years (1944-1953). The open circles indicate the mean of the outcome by day of birth (panels A, C and E) or week of birth (panels B, D and F). The solid lines are based on linear spline regressions. The light gray lines and circles show our preferred 7-day donut specification (when we drop individuals born within 7 days of January 1 in order to be symmetric around the cutoff). Source: 1994-2011 Vital Statistics Mortality files and 1992 Romanian Census (complete count).

Appendix Table 1: Specification Checks

Outcome	Density	Female	Ethnic Romanian	Ethnic Hungarian	Ethnic Roma	Ethnic Other	Born in Bucharest
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Panel A: Full Sample							
AFTER	0.105***	-0.024***	0.018***	-0.016***	-0.001	-0.001**	-0.012***
	[0.012]	[0.004]	[0.002]	[0.002]	[0.001]	[0.001]	[0.001]
AFTER*TREAT	0.010	-0.018***	-0.004	0.005**	-0.000	-0.001	0.002
	[0.016]	[0.005]	[0.003]	[0.002]	[0.001]	[0.001]	[0.002]
Sample size	3,240	3,240	3,240	3,240	3,240	3,240	3,240
R-squared	0.229	0.203	0.341	0.392	0.087	0.068	0.185
Panel B: 7-day donut sample							
AFTER	0.071***	-0.010***	0.011***	-0.012***	0.002***	-0.001	-0.008***
	[0.005]	[0.003]	[0.002]	[0.002]	[0.001]	[0.001]	[0.001]
AFTER*TREAT	-0.003	-0.010***	-0.003	0.003	0.000	-0.000	0.001
	[0.007]	[0.004]	[0.003]	[0.002]	[0.001]	[0.001]	[0.002]
Sample size	3,114	3,114	3,114	3,114	3,114	3,114	3,114
R-squared	0.289	0.096	0.308	0.368	0.158	0.046	0.152
Panel B: 14-day donut sample							
AFTER	0.072***	-0.001	0.008***	-0.010***	0.003***	-0.000	-0.006***
	[0.006]	[0.003]	[0.002]	[0.002]	[0.001]	[0.001]	[0.001]
AFTER*TREAT	-0.004	-0.014***	-0.002	0.002	0.001	-0.001	0.000
	[0.008]	[0.004]	[0.003]	[0.002]	[0.001]	[0.001]	[0.002]
Sample size	2,988	2,988	2,988	2,988	2,988	2,988	2,988
R-squared	0.229	0.053	0.270	0.332	0.177	0.038	0.126

Notes: Individual observations for cohorts born from mid-1944 to mid-1953 in the 1992 Census collapsed to the day-of-birth level. Outcome variables are defined as in Table 1. AFTER is an indicator that equals 1 for individuals born after January 1. TREAT is an indicator that equals 1 for cohorts who experienced the education expansion during 1945-1950. Local linear regressions use triangular kernels and include saturated controls for day-of birth interacted with AFTER and TREAT. All specifications use a bandwidth of 180 days. Heteroskedacticity-robust standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively.

Appendix Table 2: 2SLS estimates of Years of Schooling on Socio-Economic Outcomes

Outcome	Employment (1)	Occupational Skill (2)	Fertility (3)	Spouse's schooling (4)
Panel A: Full Sample				
Years of schooling	0.023*	0.113***	-0.170**	0.315**
	[0.012]	[0.023]	[0.071]	[0.124]
Sample size	3,240	3,240	3,240	3,240
R-squared	0.639	0.784	0.370	0.926
Panel B: 7-day donut sample				
Years of schooling	0.016	0.062	-0.125	-0.011
	[0.021]	[0.041]	[0.123]	[0.229]
Sample size	3,114	3,114	3,114	3,114
R-squared	0.625	0.730	0.363	0.895
Panel c: 14-day donut sample				
Years of schooling	0.017	0.022	0.101	-0.686
	[0.029]	[0.065]	[0.193]	[0.505]
Sample size	2,988	2,988	2,988	2,988
R-squared	0.632	0.650	0.199	0.744

Notes: Individual observations for cohorts born from mid-1944 to mid-1953 in the 1992 Census collapsed to the day-of-birth level. Outcome variables are defined as in Table 1. 2SLS regressions instrument for years of schooling with the interaction of AFTER*TREAT, control for AFTER, and include saturated controls for day-of birth interacted with AFTER and TREAT. Columns (1)-(8) report standard 2SLS estimates while columns (9)-(12) use report estimates from two-sample 2SLS. All specifications use a bandwidth of 180 days. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively.

Appendix Table 3: Outcomes from LSMS household surveys

Outcome	Years of schooling	Employment	Income	Smoking	Chronic conditions
	(1)	(2)	(3)	(4)	(5)
Panel A: Full Sample					
AFTER	0.0609	-0.0188	0.0105	0.0031	0.0032
	[0.2240]	[0.0130]	[0.0438]	[0.0058]	[0.0047]
AFTER*TREAT	0.6913**	0.0415***	0.0195	0.0086	0.0083
	[0.2997]	[0.0044]	[0.0229]	[0.0089]	[0.0074]
Sample size	10,315	30,340	30,133	84,032	84,032
R-squared	0.034	0.020	0.009	0.006	0.001

Notes: Individual observations for cohorts born from mid-1944 to mid-1953 in the 1992 Census collapsed to the day-of-birth level. Outcome variables are defined as in Table 1. AFTER is an indicator that equals 1 for individuals born after January 1. TREAT is an indicator that equals 1 for cohorts who experienced the education expansion during 1945-1950. Regressions include saturated linear controls for month-of birth interacted with AFTER and TREAT. All specifications use a bandwidth of 90 days. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively.

Appendix Table 4: Parametric RD Estimates

Outcome		Years of schooling			Mortality			Density	
Polynomial function	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic	Linear	Quadratic	Cubic
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)
AFTER	-0.050***	-0.018	0.032	0.005*	0.008**	0.004	0.059***	0.091***	0.074***
	[0.018]	[0.029]	[0.046]	[0.002]	[0.004]	[0.006]	[0.005]	[0.007]	[0.010]
AFTER*TREAT	0.123***	0.104**	0.177***	0.004	0.003	0.011	-0.005	-0.001	-0.015
	[0.025]	[0.042]	[0.065]	[0.004]	[0.006]	[0.009]	[0.007]	[0.009]	[0.013]
Sample size	3,114	3,114	3,114	3,114	3,114	3,114	3,114	3,114	3,114
R-squared	0.920	0.922	0.923	0.779	0.779	0.780	0.167	0.246	0.258

Notes: Individual observations for cohorts born from mid-1944 to mid-1950 in the 1992 Census collapsed to the day-of-birth level. Outcome variables are defined as in Table 1. AFTER is an indicator that equals 1 for individuals born after January 1. AFTER is an indicator that equals 1 for individuals born after January 1. TREAT is an indicator that equals 1 for cohorts who experienced the education expansion during 1945-1950. Local, quadratic, and cubic specifications include saturated controls for day-of birth interacted with AFTER and TREAT. All specifications use a bandwidth of 180 days. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively.

Appendix Table 5: Uncollapsed RD Estimates

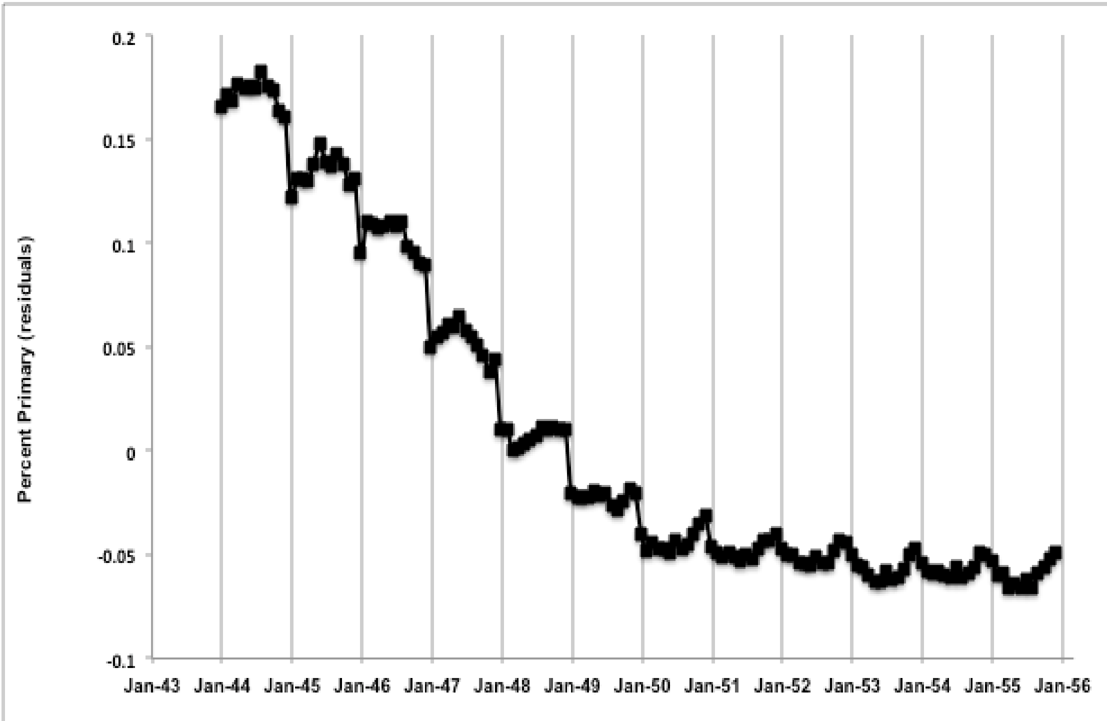
Outcome	Years schooling (1)	Primary (2)	L. Secondary (3)	U. Secondary (4)	Higher ed. (5)	Employment (6)	Occ skill (7)	Fertility (8)	Spouse's school (9)
AFTER	-0.038**	0.004***	-0.009***	0.011***	-0.006***	0.001	-0.015***	0.014	-0.033*
	[0.017]	[0.001]	[0.002]	[0.003]	[0.002]	[0.002]	[0.004]	[0.011]	[0.019]
AFTER*TREAT	0.102***	-0.016***	0.008***	0.007**	0.000	0.001	0.006	-0.012	-0.013
	[0.023]	[0.002]	[0.003]	[0.003]	[0.002]	[0.002]	[0.005]	[0.014]	[0.024]
Sample size	2,466,846	2,466,846	2,466,846	2,466,846	2,466,846	2,474,835	2,155,293	1,240,796	2,114,104
R-squared	0.022	0.044	0.001	0.014	0.001	0.002	0.002	0.001	0.017

Notes: Individual observations for cohorts born from mid-1944 to mid-1950 in the 1992 Census. Outcome variables are defined as in Table 1. AFTER is an indicator that equals 1 for individuals born after January 1. Local linear regressions use triangular kernels and include controls for day-of birth and day-of-birth interacted with AFTER. All specifications use a bandwidth of 180 days. Heteroskedasticity-robust standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively.

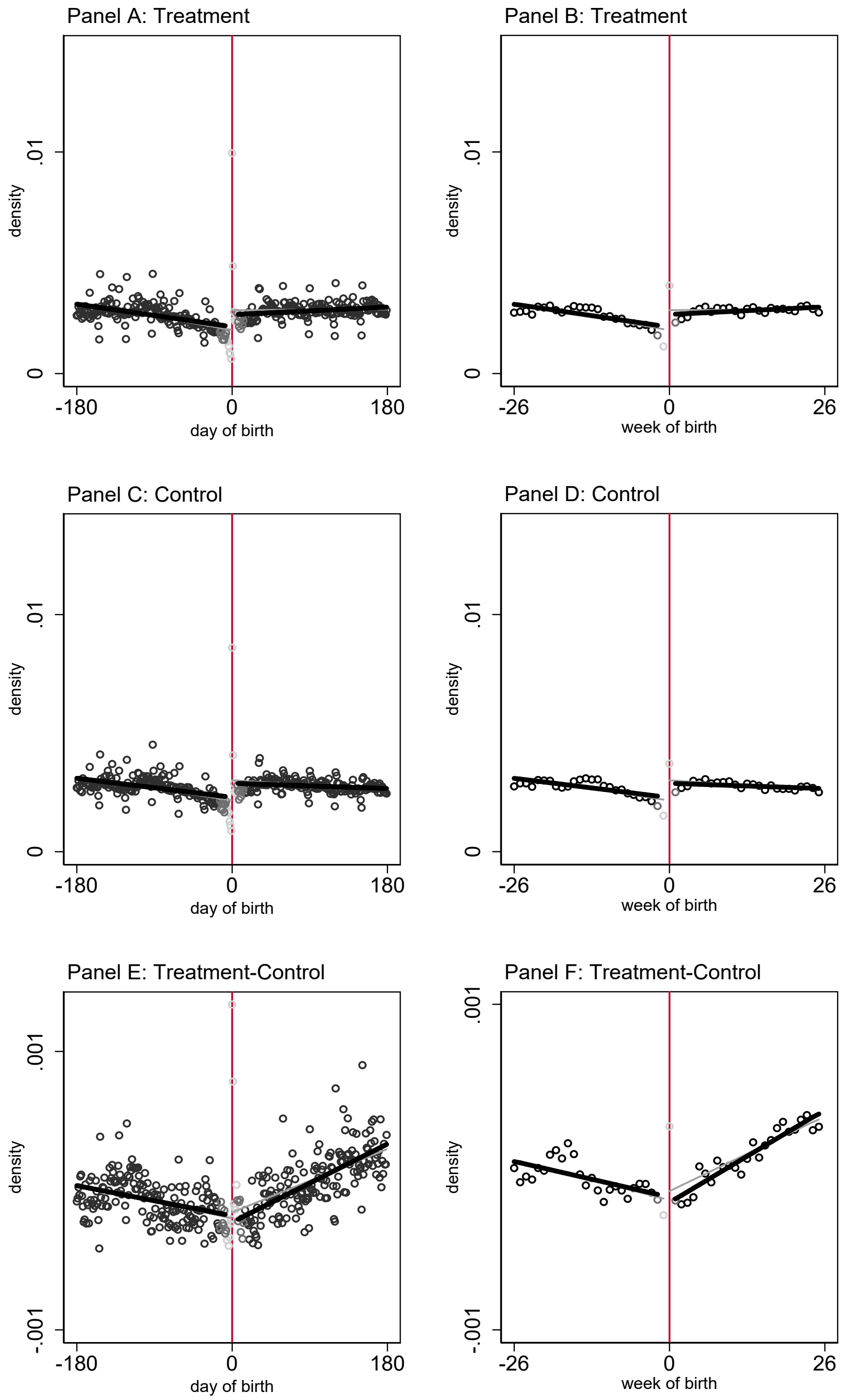
Appendix Table 6: Confounders: Migration and Attrition checks

Outcome	International Migration	Attrition
	(1)	(2)
Panel A: Full Sample		
AFTER	0.011***	-0.007
	[0.003]	[0.006]
AFTER*TREAT	-0.001	-0.006
	[0.003]	[0.008]
Sample size	3,240	3,240
R-squared	0.563	0.218
Panel B: 7-day donut sample		
AFTER	0.009***	-0.007
	[0.003]	[0.006]
AFTER*TREAT	-0.000	-0.004
	[0.004]	[0.008]
Sample size	3,114	3,114
R-squared	0.576	0.227
Panel C: 14-day donut sample		
AFTER	0.008***	-0.004
	[0.003]	[0.007]
AFTER*TREAT	0.002	-0.007
	[0.004]	[0.009]
Sample size	2,988	2,988
R-squared	0.581	0.224

Notes: Individual observations for cohorts born from mid-1944 to mid-1953 in the 1992 Census collapsed to the day-of-birth level. Outcome variables are defined as in Table 1. AFTER is an indicator that equals 1 for individuals born after January 1. TREAT is an indicator that equals 1 for cohorts who experienced the education expansion during 1945-1950. Local linear regressions use triangular kernels and include saturated controls for day-of birth interacted with AFTER and TREAT. All specifications use a bandwidth of 180 days. Heteroskedacticity-robust standard errors are in parentheses. ***, **, and * indicate statistical significance at the 1, 5, and 10 percent level respectively.

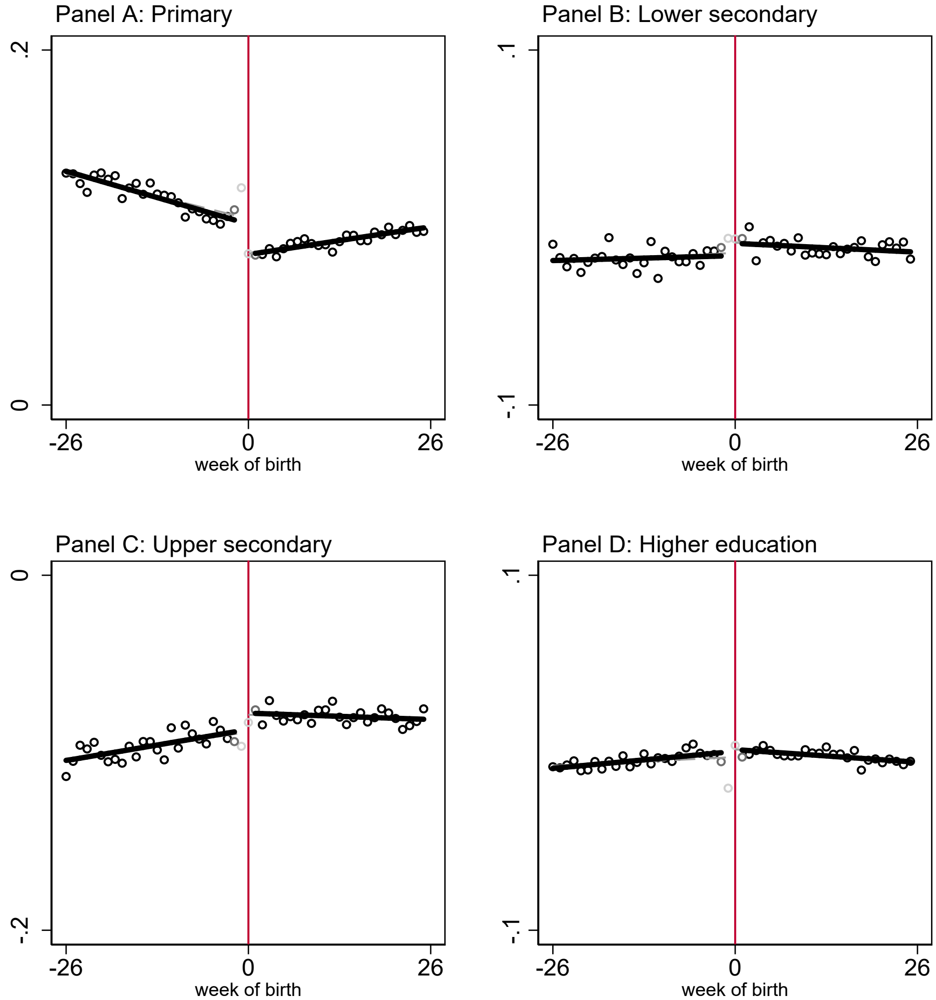


Appendix Figure 1: Primary school completion (residualized) for cohorts born 1943-1955
Notes*:* This figure plots the percent of individuals born between 1943 and 1955 who completed only primary education by their month of birth, which are based on residuals. Source: 1992 Romanian Census (complete count).

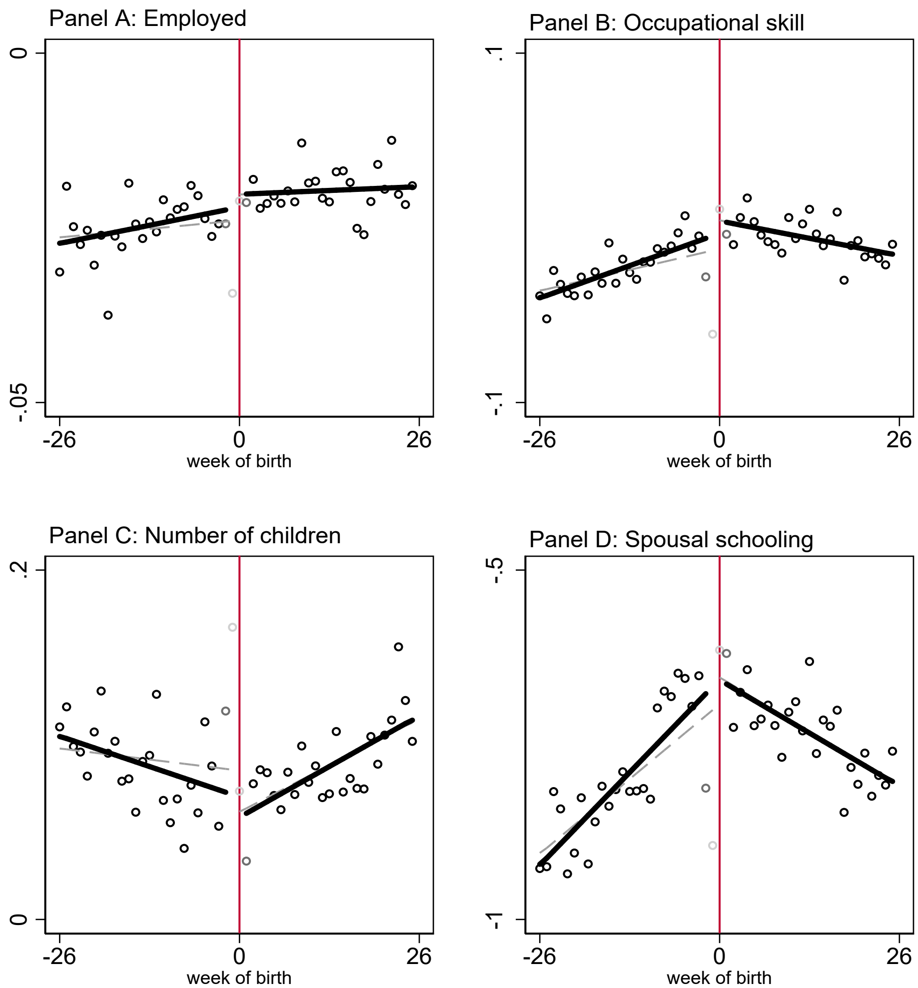


Appendix Figure 2: Regression discontinuity (RD) plots of the density

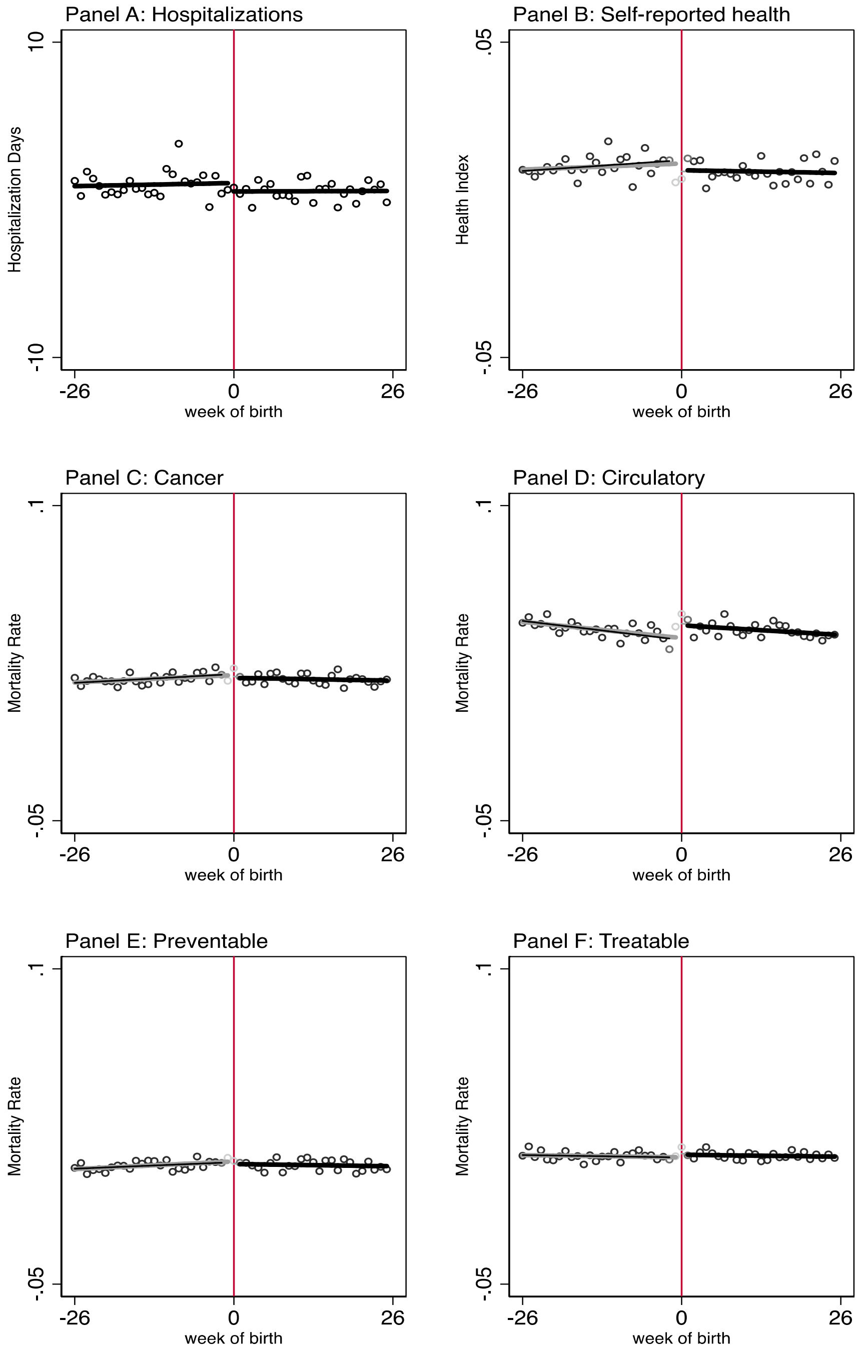
Notes: Panels A and B are restricted to individuals born in the treatment years (1944-1950). Panels C and D are restricted to individuals born in the control years (1950-1953). Panels E and F are restricted to individuals born in both treatment and control years (1944-1953). The open circles indicate the mean of the outcome by day of birth (panels A, C and E) or week of birth (panels B, D and F).The solid lines are based on linear spline regressions. The light gray lines and circles show our preferred 7-day donut specification (when we drop individuals born within 7 days of January 1 in order to be symmetric around the cutoff). Source: 1992 Romanian Census (complete count).



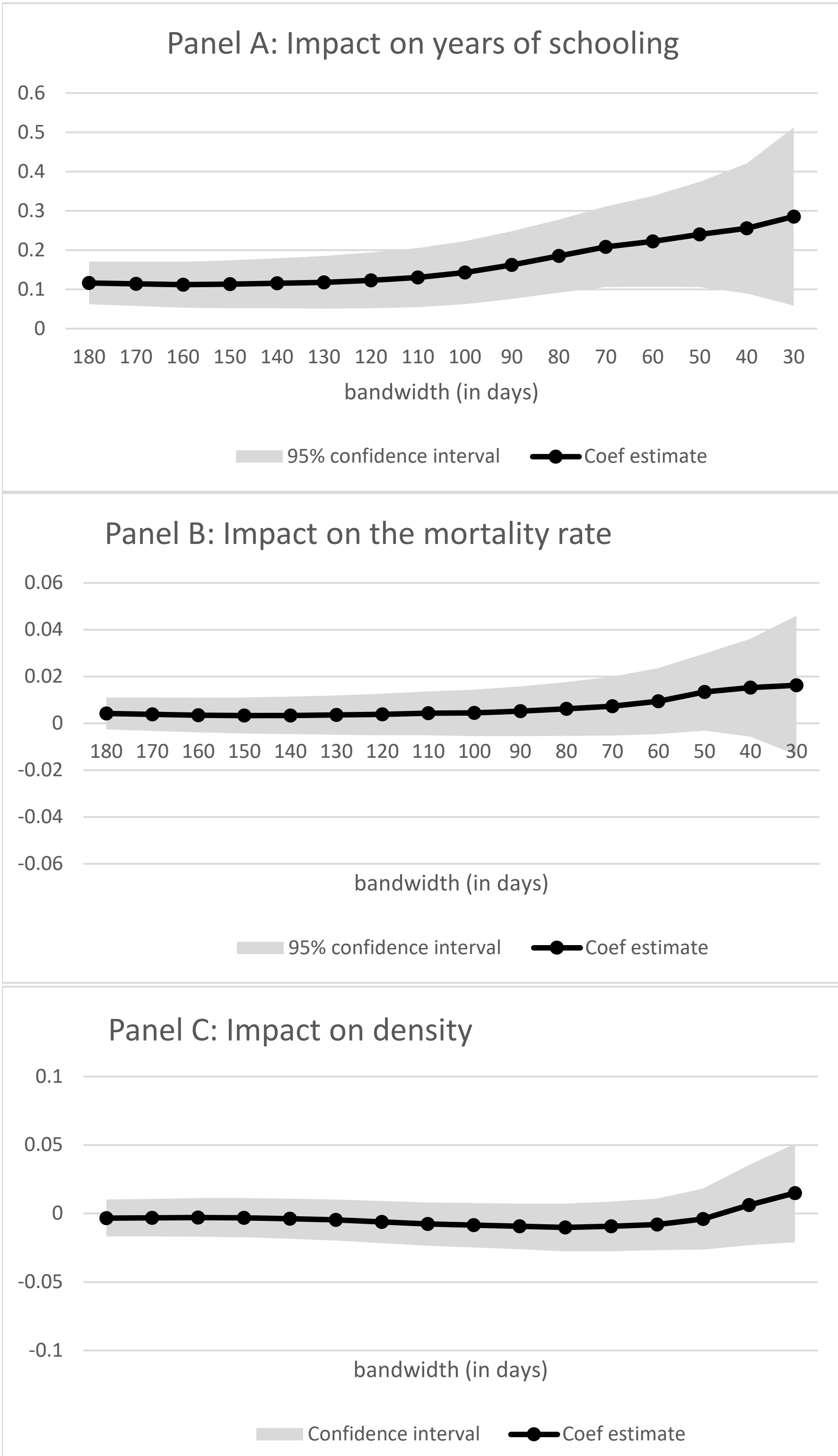
Appendix Figure 3: Regression discontinuity (RD) plots for specific schooling categories
Notes: All Panels are restricted to individuals born in both treatment and control years (1944-1953). The open circles indicate the mean of the outcome by week of birth. The solid lines are based on linear spline regressions. The light gray lines and circles show our preferred 7-day donut specification (when we drop individuals born within 7 days of January 1 in order to be symmetric around the cutoff). Source: 1992 Romanian Census (complete count).



Appendix Figure 4: Regression discontinuity (RD) plots of Socio-Economic Outcomes
Notes: All Panels are restricted to individuals born in both treatment and control years (1944-1953). The open circles indicate the mean of the outcome by week of birth. The solid lines are based on linear spline regressions. The light gray lines and circles show our preferred 7-day donut specification (when we drop individuals born within 7 days of January 1 in order to be symmetric around the cutoff). Source: 1992 Romanian Census (complete count).



Appendix Figure 5: Regression discontinuity (RD) plots of Health and Mortality by Cause
Notes: All Panels are restricted to individuals born in both treatment and control years (1944-1953). The open circles indicate the mean of the outcome by week of birth. The solid lines are based on linear spline regressions. The light gray lines and circles show our preferred 7-day donut specification (when we drop individuals born within 7 days of January 1 in order to be symmetric around the cutoff). Source: 2011 Romanian Census (complete count) and 1994-2011 Vital Statistics Mortality files



Appendix Figure 6: RD estimates for alternative bandwidths

Notes: This figure plots RD estimates for three outcomes (years of schooling, mortality rate and density) over a broad range of alternative bandwidths between 180 and 30 days using the 7-day donut sample. Source: 1992 Romanian Census (complete count), Vital Statistics Mortality files from 1994 to 2011.