CESifo WORKING PAPERS

7927
2019

October 2019

School Tracking and Mental Health

Petri Böckerman, Mika Haapanen, Christopher Jepsen, Alexandra Roulet

CESifo

Impressum:

CESifo Working Papers
ISSN 2364-1428 (electronic version)
Publisher and distributor: Munich Society for the Promotion of Economic Research - CESifo GmbH
The international platform of Ludwigs-Maximilians University's Center for Economic Studies and the ifo Institute
Poschingerstr. 5, 81679 Munich, Germany
Telephone +49 (0)89 2180-2740, Telefax +49 (0)89 2180-17845, email office@cesifo.de
Editor: Clemens Fuest
www.cesifo-group.org/wp

An electronic version of the paper may be downloaded
· from the SSRN website: www.SSRN.com
· from the RePEc website: www.RePEc.org
· from the CESifo website: www.CESifo-group.org/wp

CESifo *Working Paper No. 7927*
Category 5: Economics of Education

School Tracking and Mental Health

Abstract

We examine the effects of a comprehensive school reform on mental health. The reform postponed the tracking of students into vocational and academic schools from age 11 to age 16. The reform was implemented gradually across Finnish municipalities between 1972 and 1977. We use difference-in-differences variation and administrative data. Our results show that there is no discernible effect on mental health related hospitalizations on average even though the effect is precisely estimated. Heterogeneity analysis shows that, after the reform, females from highly-educated families were more likely to be hospitalized for depression.

JEL-Codes: I120, I260, I280.

Keywords: tracking age, comprehensive school, mental health, depression, hospitalization.

Petri Böckerman
University of Jyväskylä
Jyväskylä / Finland
petri.bockerman@labour.fi

Mika Haapanen
University of Jyväskylä
Jyväskylä / Finland
mika.p.haapanen@jyu.fi

Christopher Jepsen
University College Dublin
Dublin / Ireland
christopher.jepsen@ucd.ie

Alexandra Roulet
INSEAD
Fontainebleau / France
alexandra.roulet@insead.edu

October 23, 2019

1 Introduction

Education leads to monetary (Angrist and Krueger, 1991) and non-monetary (Oreopoulos and Salvanes, 2011) gains at the individual level. A crucial part of the non-monetary return provided by education is the potential positive effect on health. The positive correlation between education and health is well established (Cutler and Lleras-Muney, 2008). However, quasi-experimental evidence using natural policy experiments on the causal link between education and health outcomes remains inconclusive (Galama et al., 2018).

We advance the understanding of the education-health relationship by studying the effect of a change in school tracking age on mental health in adulthood.[1] So far the literature has focused on the effects of education on physical health. The lack of evidence on the effects on mental health outcomes is a salient gap, because mental health is an increasingly important domain of health, especially in the developed countries (Frank and McGuire, 2000; Layard, 2013). Depressive disorders are a leading and often underestimated cause of the global disease burden (Vigo et al., 2016). For example, depressive disorders account for 12% of total years lived with disability, and depression is the largest contributor to the disease burden attributable to non-fatal health outcomes (Ustun and Chatterji, 2001; Whiteford et al., 2013). Mental health problems also lead to substantial indirect costs such as absenteeism and productivity losses at work (Bubonya et al., 2017). Additionally, mental health problems increase the risk of poor physical health (Sareen et al., 2006).

Moreover, most of the literature studies the effects on health of one additional year of education whereas other relevant aspects of education, such as how long students are exposed to a common curriculum before being split between more

[1] In our context of school systems in Europe, tracking refers to the streaming of students between the academic and vocational educational tracks, whereas in the US literature tracking usually refers to ability grouping within schools (Hall, 2012).

academic tracks and more vocational ones, might also have an effect on health, especially on mental well-being. Indeed, school tracking fundamentally affects the set of peers to which students are exposed as well as the type of skills they acquire and the degree of competition they face in the classroom. Peer effects and exposure to competition are potential determinants of mental health and well-being. Understanding whether and how changes in school tracking affect mental health provides insights into the mechanisms through which education relates to health. Many European countries implemented comprehensive schooling reforms since the end of the Second World War to delay the age at which students are selected into different tracks (Brunello et al., 2007).[2] The primary policy motivation for these reforms was that early tracking systems were considered unfair to pupils from disadvantaged backgrounds (Jones et al., 2014).

We use difference-in-differences variation triggered by the Finnish comprehensive school reform, which was implemented gradually across Finnish municipalities over the period 1972–1977, to identify the effect of school tracking on mental health.[3] The phase-in of the reform offers plausibly exogenous variation in the tracking age and its occurrence in the 1970s allows us to identify long-run health effects. Key to our identification strategy is the fact that the gradual implementation was orthogonal to the incidence of mental health problems prior to the reform. Our evidence confirms that the municipalities that were treated first were not different in terms of mental health from those that were treated later. Moreover, the reform postponed the tracking of students into vocational and academic schools from age 11 to age 16 without affecting the length of compulsory education. Thus, the reform provides a unique opportunity to study the effect of increasing the school tracking age, holding fixed the number of years of compulsory education.

[2]Nevertheless, there are still significant differences in the tracking ages between the OECD countries (OECD, 2004, p. 262).

[3]Previous studies have used the reform to study non-mental health outcomes (Pekkarinen, 2008; Pekkarinen et al., 2009; Pekkala Kerr et al., 2013).

To identify the effects of this reform on mental health, we use administrative panel data for the Finnish population born in the 1960s. We have access to complete registers on all-cause mortality, suicides, and hospitalizations from the late 1960s to 2013. The registers include all hospital admissions related to mental health disorders in Finland. Using the gradual rollout of the comprehensive school reform across regions and over time, we estimate difference-in-differences models to identify the effects of the reform on these health outcomes by age 45.

The earlier empirical literature on the effects of school tracking on health outcomes is thin. Jones et al. (2014) and Meghir et al. (2018) find negligible effects on health. However, none of the earlier studies have focused on mental health, which is likely to be affected by the change in the composition of peers induced by the postponing of the tracking age.

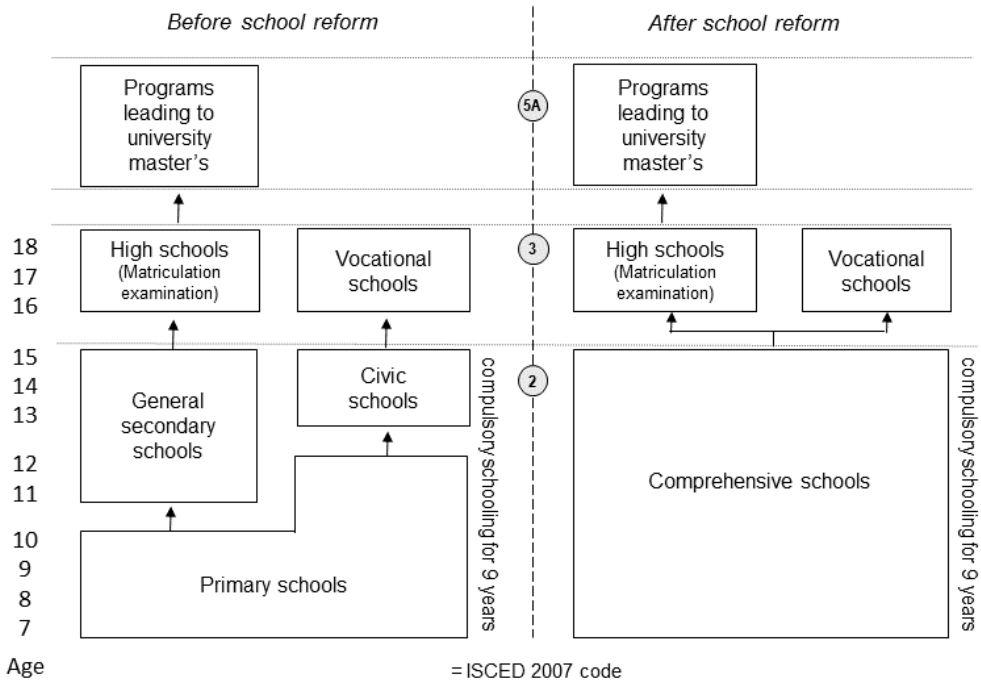
2 The Education Reform and Its Expected Effects on Mental Health

2.1 The Structure of the Finnish Education System Before and After the Reform

Finland had a selective two-track school system until the 1970s. The reform replaced the old two-track system with a uniform comprehensive school system (Somerkivi, 1982). Figure 1 describes the structure of the old and the new systems (cf. Sahlberg, 2014).

The reform postponed the tracking age from age 11 to age 16. Both before and after the reform school starts at age 7 and is compulsory until age 16. In the old system, pupils were taught together in the same class for only four years, from age

Figure 1: The Structure of the Finnish Education System Before and After the Comprehensive School Reform in the 1970s



7 to 11. Then the students were placed into academic or vocational tracks for the remaining five years. In contrast, in the new system, there is an almost uniform curriculum[4] for all nine years, until age 16. At its core, the reform significantly affected the composition of the peers to which pupils were exposed between age 11 and 16.[5]

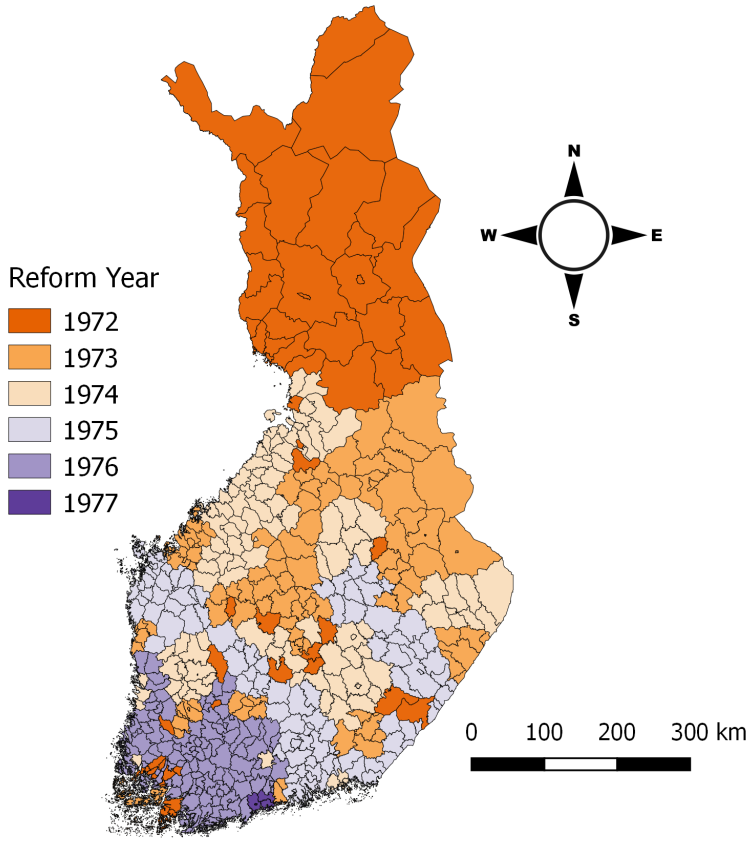
The comprehensive school reform was rolled out gradually across Finnish municipalities over the period 1972–1977 (Figure 2). The timing of the reform in the different municipalities was decided by the central government. The country was divided into geographically and functionally uniform implementation regions. The principle was that each province was divided into several implementation regions

[4]Ability groups in foreign languages and mathematics were used in the comprehensive school (grades 7 to 9) until 1985 (Sahlberg, 2014, p. 28).

[5]Schooling from age 7 to 11 (grades 1 to 4) remained unchanged after the reform as the teachers from the pre-reform system were assigned to the comprehensive schools, and curriculum in both systems were similar for grades 1 to 4 (Somerkivi, 1982, p. 28).

(Somerkivi, 1982, p. 27–28). There is considerable variation in exposure to the new comprehensive school system, both across birth cohorts and across municipalities (see Table 1 below). This variation provides a quasi-experimental research setting.

Figure 2: Implementation of the School Reform over the period 1972–1977



Previous research has shown that the reform decreased the intergenerational correlation of earnings between fathers and sons (Pekkarinen et al., 2009). The reform also slightly improved the verbal and mathematical test scores of boys coming from a low socio-economic background (Pekkala Kerr et al., 2013). Finally, the reform increased the gender difference in the probability of choosing an academic track and obtaining tertiary education (Pekkarinen, 2008). However, the potential effects of the reform on mental health outcomes have not been studied. In Section 6, we discuss the potential mediating role of education and income on mental health.

2.2 Expected Effects on Mental Health

A key feature of the Finnish reform is that it did not affect the length of compulsory schooling: the reform did not change the minimum school leaving age of 16. This feature allows us to focus on the effects of a change in the tracking age on mental health outcomes, holding fixed the number of years of compulsory education.

Interactions with peers in school during childhood and adolescence are important determinants of mental health outcomes in adulthood (WHO, 2014). In theory, the predicted effects of the reform on health outcomes, and particularly on mental health, are ambiguous. After the reform, students aged 11 to 16 have a common set of peers rather than peers from their specific track only. Low-achieving students, who would have been assigned to the vocational track, are exposed to higher-achieving peers in the post-reform system. This change is expected to improve education and related economic outcomes for low-achieving students, which may improve health outcomes in the long run. It is also possible that low-achieving students will adopt and imitate health behaviors of their higher-achieving peers. Conversely, higher-achieving students may be adversely affected by being exposed to lower-achieving peers.

The comprehensive school reform also affected the ability ranking in the classroom. Vocational-track students who were high in the ability rank in the pre-reform system are now on average ranked lower in a comprehensive school classroom. Cicala et al. (2018) have shown that students' academic achievements and disruptive behaviors depend on their ordinal rank among their peers. Moreover, Elsner and Isphording (2018) provide evidence that a student's ordinal ability rank in a high-school cohort is an important determinant of engaging in risky behaviors (such as smoking, drinking, and proneness to physical fights). This evidence suggests that pupils who would have been tracked into vocational schooling absent the reform

could have lower health, and in particular poor mental health outcomes, after the reform. It goes against the positive effect of being exposed to higher-achieving peers mentioned above.

Moreover, the fact that students in the post-reform system follow a common curriculum from age 11 to 16, rather than a track-specific curriculum, also means that, after the reform, students in a given classroom are less homogeneous. It is much more difficult for teachers to tailor their pedagogical approaches to a more heterogeneous mix of students and this can result in negative education and health outcomes (Betts, 2011). Students' achievement may be better when they are surrounded by peers with similar characteristics. Indeed, the Finnish experience suggests that mixed-ability groups led to learning difficulties; disruptive behaviors increased following the reform, and the number of pupils in special education more than doubled between school years 1974–75 and 1979–80 (Somerkivi, 1982, p. 40).

Finally, one of the main reason in favor of delaying tracking is that the likelihood of a student being placed in the "wrong" track is reduced and the anxiety associated with tracking lessens because the amount of pre-tracking information about students' abilities is higher at the time of the tracking decision (Brunello et al., 2007). This would suggest that mental health could improve as a result of the reform. However, Pekkarinen (2008) argues that for boys, this benefit is offset by the fact that tracking now occurs during puberty, in contrast to girls, for whom it occurs after puberty.

In short, previous work does not point towards clear predictions about the sign of the reform's effect on mental health since different mechanisms are pushing in different directions but suggests that the effect will likely differ by gender and by academic ability.

3 Data

To evaluate the long-run effect of the reform on mental health, we link three data sets: i) census data covering the total population of Finland, ii) data on the causes of death from the comprehensive death certificates, and iii) complete hospital admissions data.

3.1 Census Data

We use the population register data on individuals who are permanent residents in Finland. The data originate from the Longitudinal Population Census Files from Statistics Finland. Demographic and labor market information are available for the years 1975 and 1985 and after that annually over the period 1987–2014. The municipality of residence is recorded annually from 1971 onwards. The data have almost complete household and parental links. The date and municipality of birth are also recorded. Furthermore, the data contain detailed information about degrees that were completed from 1970 onwards.

The core data include the universe of individuals who were born in Finland between 1962 and 1966, following Pekkala Kerr et al. (2013, p. 586). We start with birth cohort 1962 and end with birth cohort 1966 to increase the homogeneity of the cohorts under study.[6] This sample restriction also allows us to follow all birth cohorts to the age of 45. As shown in Table 1, there are between 72,248 and 74,248 persons in each birth cohort. Hence, we have approximately 366,000 individuals in total. Annual information about the municipality of residence, together with the birth date, determines whether a pupil attended the tracked or comprehensive school system. The reform was effective for students who were at most 11 years old, i.e., entering 5th grade, at the end of the year in which the reform was imple-

[6] In addition, the quality of health data is weaker before 1972 (Sund, 2012, p. 507).

mented in their region of residence. For instance, for people who, in the year in which they turned 11 years old, lived in a region in which the reform was implemented in 1975 had post-reform schooling if they were born in 1964, 1965 or 1966 but had pre-reform schooling if they were born in 1962 or 1963.[7]

Table 1: Number of Observations and Implementation of the Reform

Birth cohort	Implementation year in the municipality						Total
	1972	1973	1974	1975	1976	1977	
1962	7,460	10,739	14,361	14,965	15,465	9,348	72,338
1963	7,402	10,656	15,507	15,157	15,934	9,592	74,248
1964	7,112	10,309	15,086	15,710	16,013	9,865	74,095
1965	6,646	9,754	14,608	15,548	16,315	9,913	72,784
1966	6,524	9,638	14,425	15,146	16,726	10,390	72,849
Total	35,144	51,096	73,987	76,526	80,453	49,108	366,314

Note: For each birth cohort, people affected by the reform were those who lived in municipalities where the reform was implemented in the years that correspond to the cells shaded in gray. For instance, for people born in 1962, the treatment group consists of people who at age 11, i.e., in 1973, lived in a municipality in which the reform was implemented in 1972 or 1973.

We exclude from the estimation sample foreign-born individuals (most of whom immigrated to Finland after the reform) and those living in the Åland Islands (8% of the original sample), retaining individuals living in 465 different municipalities. We also exclude a small number of individuals who migrated at ages 11–15 between municipalities with a different year of implementation of the reform (less than 3% of the original sample) because the reform indicator cannot be assigned unambiguously for these people (i.e. they may or may not have migrated between the implementation regions). Finally, we exclude a very small number of emigrants from the original data.

Since measures of academic ability prior to tracking are not available in the data, we investigate heterogeneity by parents' education.[8] The variable takes three values:

[7]We have also investigated the potential role of endogenous selection of municipality of residence. Thus, we checked the robustness of our results to assigning the individuals to treatment based on their municipality of *birth* and their date of birth, as in Meghir et al. (2018).

[8]The correlation between parents' education and children's academic ability is strong when

i) an individual's parents are classified as low educated (53%) if neither of them completed post-compulsory education, i.e., if they both have a maximum of nine years of schooling; ii) an individual's parents are classified as mid educated (27%) if either or both of the parents completed a vocational degree but none studied further; iii) an individual's parents are classified as highly educated (20%) if at least one of them completed high school or a higher level of (tertiary) education (see Figure 1 for a reminder of the pre-reform system).

3.2 Mortality and Mental Health Disorders

To relate our findings to previous research on education-health nexus (Clark and Royer, 2013; Meghir et al., 2018), we first examine mortality outcomes. We use data about the year and causes of death from the comprehensive death certificates (until 2013). All diagnoses of the causes of death pass a routine validation conducted by Statistics Finland, and unclear cases are judged by a panel (Lahti and Penttilä, 2001). All-cause mortality is measured by a dummy variable for dying by age 45. We picked age 45 such that all birth cohorts can be followed over the same window: the youngest cohort born in 1966 is 45 years old in 2011.

We use suicides as the second mortality variable (also a dummy for occurrence before age 45). Suicides are defined by the codes X60 to X84 and Y87.0 in the International Classification of Diseases (ICD-10), which is the standard diagnostic tool for clinical purposes. Suicide is a relevant outcome for three reasons. First, approximately 25% of all deaths by age 45 are suicides. Second, suicides are closely related to mental health problems. For example, approximately 90% of suicides are associated with psychiatric disorders (Pirkola et al., 2009; Henriksson et al., 1993).

the ability is measured by high school completion (see Table A1 in Appendix B). Note also that, contrary to the parental sub-samples, the full sample includes a small number of individuals whose parental information is missing (less than 5%). However, the results remain intact if we exclude these individuals from the sample.

Third, the suicide mortality of young Finns is among the highest in the world (Lahti et al., 2011).

We then study the effects of the reform on mental health related hospitalizations using inpatient data.[9] We focus on *serious* mental health outcomes for two reasons. First, their treatment costs are particularly high in the universal health care system. Second, severe mental illnesses cause substantial indirect costs in terms of absenteeism, weak long-run labor market attachment and early disability pensions (Hakulinen et al., 2019). Therefore, there is no doubt that society cares about these outcomes.

We use register-based measures that are free from the potential measurement error inherent to self-reported mental health symptoms (Ritter et al., 2001). Information about mental health disorders is extracted from the Hospital Discharge Register (HDR) compiled by the National Institute for Health and Welfare over the period 1969–2013. The data include dates of admission to the hospital, dates of discharge, and the primary reason for hospitalization. Diagnosis codes are from the 8th, 9th, and 10th revisions of the ICD. Spells due to mental health disorders correspond to a diagnosis code starting with the letter F in the ICD-10 classification and to 290–319 in ICD-8 or 9. Validation studies have shown that the Hospital Discharge Register data are of high quality from 1972 onwards (Sund, 2012).

In the baseline model, we use a dummy variable indicating whether the individual had any mental health related hospitalization between ages 16 and 45.[10] Approximately 10 percent of men and 6 percent of women had mental health problems that resulted in a hospital entry (Table 2). Conditional on having a mental health related hospitalization, the average time spent in the hospital between ages 16 and

[9] Finnish outpatient data are available only for the most recent years and the data are not nationally representative.

[10] Most mental disorders emerge before the age of 25 (Pedersen et al., 2014).

45 is 151 days for men and 193 days for women.[11] We then also consider separately whether the individual had any spell starting at age 16–25, 26–35, or 36–45. Again, we record new hospitalization spells until age 45 such that all birth cohorts can be followed during the same age window.

To get a more comprehensive picture, we also examine the effect of the reform on different types of mental health disorders (Santavirta et al., 2015; Suvisaari et al., 2009): i) schizophrenia, a mental disorder characterized by hallucinations, delusions and cognitive deficits; ii) other psychoses that are not related to emotions or moods (non-affective psychosis); iii) bipolar disorders, which are examples of affective psychosis involving emotional and mood abnormalities (and manic episodes); iv) depressive disorders, which can include repeated episodes of severe depression or chronic milder depression (dysthymia); and v) severe anxiety, stress, and neurotic disorders, which can interfere with daily activities such as job performance, school work, and social relationships. The final category, vi) substance-use disorders, includes all psychiatric hospitalizations related to alcohol or substance abuse or addiction. Appendix A contains details about the codes used to define these categories.

Table 2 provides the summary statistics for the main outcomes of interest, broken down by gender, and Table A1 in Appendix B provides these same statistics, also broken down by socio-economic background, in addition to some summary statistics on additional outcomes.

[11] The average duration of one spell, conditional on being hospitalized between age 16 and 45, is 33 days for men and 38 days for women (the corresponding medians are 7 and 11 days).

Table 2: Summary Statistics, by Gender

	Males		Females	
Variable	Mean	Std. dev.	Mean	Std. dev.
Has a high school degree	0.291	0.454	0.498	0.500
Years of schooling	12.777	2.723	13.489	2.635
Mortality				
Death by age 45	0.046	0.209	0.018	0.131
Suicide by age 45	0.0129	0.1128	0.0030	0.0549
Hospitalizations due to mental health disorder				
At age 6–10	0.006	0.080	0.004	0.063
At age 11–15	0.006	0.079	0.005	0.070
At age 16–25	0.044	0.205	0.017	0.127
At age 26–35	0.044	0.206	0.028	0.166
At age 36–45	0.050	0.219	0.034	0.181
At age 16–45	0.100	0.300	0.056	0.231
Hospitalizations at ages 16-45 due to				
Schizophrenia	0.012	0.109	0.008	0.092
Other non-affective psychosis	0.018	0.132	0.015	0.123
Bipolar disorder	0.005	0.070	0.005	0.069
Depressive disorder	0.017	0.131	0.019	0.136
Anxiety, stress, neurotic disorder	0.016	0.127	0.007	0.084
Substance-use disorder	0.042	0.200	0.013	0.115
Mental health related hospitalization days at ages 16-45				
Unconditional	15.017	172.39	10.861	133.25
Conditional on being hospitalized	150.66	527.00	192.59	529.01
Parental education				
Low-educated parents	0.530	0.499	0.529	0.499
Mid-educated parents	0.266	0.442	0.268	0.443
Highly-educated parents	0.204	0.403	0.203	0.402
Observations	186,777		179,537	

Note: Low-educated parents means neither parent completed post-compulsory schooling; mid-educated parents means at least one parent completed a vocational degree but not more; highly-educated parents means at least one parent completed a higher education degree (including high school). The sum of the last two education dummies is equal to the parental education variable used in Pekkala Kerr et al. (2013).

4 Empirical Approach

To identify the average long-run effects of the comprehensive school reform, we estimate difference-in-differences models with the following structure:

$$y_{ijt} = \alpha + \eta_j + \tau_t + \beta \times REFORM_{jt} + \gamma' X_i + \epsilon_{ijt} \quad (1)$$

where y_{ijt} is the health outcome of individual i, who was born in year t and schooled in municipality j when entering 5th grade. η_j and τ_t are the municipality and the birth cohort fixed effects. There are permanent regional differences in the outcomes that we need to control for. Similarly, birth cohorts may also be exposed to different shocks in childhood and adolescence that have impacts on mental health in adulthood. $REFORM_{jt}$ is a dummy that varies across municipalities and cohorts and equals one if individual i was exposed to the reform, i.e., experienced comprehensive school until age 16. Thus, β is the policy parameter of interest in the models. The baseline specification does not include control variables X_i except a constant. We check the sensitivity of our results to the inclusion of controls.

Because we identify the estimates using a difference-in-difference framework, the timing of the reform implementation needs to be unrelated to underlying cohort trends in mental health disorders across municipalities. Figure A1 provides evidence for the lack of relationship between all six types of individual mental health disorders studied and the timing of the reform. We also show that there is no evidence of a systematic relationship between the timing of the reform implementation and baseline pre-treatment municipality characteristics that may affect mental health disorders (Figure A2). Moreover, we have estimated dynamic responses using lead and lag year dummies around the reform, omitting the year prior to the reform, as in Pekkala Kerr et al. (2013). Time zero represents the first birth

cohort in the municipality affected by the reform. This model allows us to separate the pre-existing trends from policy responses over time. The results reported in Figures A3–A5 of Appendix B suggest that our main findings are not affected by pre-reform trends.[12]

The estimated models identify average treatment effects for the treated (ATT). We report estimates from linear probability models because they facilitate the interpretation of the estimated coefficients and are less sensitive to distributional assumptions (Wooldridge, 2001).

We estimate the empirical specifications separately by gender, because there is substantial variation in all outcomes by gender. For example, suicide mortality is much higher among males. We also estimate the models by parental level of education (three mutually exclusive categories) because earlier research suggests that the effects of the reform may differ significantly by socioeconomic background (Pekkala Kerr et al., 2013). Additionally, there may be socioeconomic differences in the utilization of hospital care even though Finland has a universal health care system (Gerdtham, 1997; Bijwaard et al., 2018). Possible regional differences in the utilization of hospital care are captured by the municipality fixed effects that are included in all models.

Throughout the paper, standard errors are clustered at the municipal level. In addition to the conventional clustered standard errors, we also report significance levels based on adjusted standard errors that account for testing multiple hypotheses. We apply the step-down approach of Romano and Wolf (2005), which takes advantage of the dependence structure of individual tests.

We provide robustness checks for our baseline results. First, we control for region-

[12]Furthermore, Figure A6 in Appendix B shows that there is no evidence for significant effects at ages 6–10 on mental health related hospitalizations. However, this test is relatively weak because it is possible that the replacement of the old system (primary schools) with comprehensive school system also affected 7–10-year-old children even though the curriculum remained largely unchanged (Somerkivi, 1982, p. 28).

specific linear time trends, which makes identification of the effects much less reliant on the common trend assumption. Second, we use mental health related hospitalizations at ages 6–10 as an additional control to account for the possible relationship between prior mental health disorders and treatment status. Third, we estimate models with and without the Helsinki metropolitan area because some private schools were operating in the Helsinki region after the reform (Pekkarinen et al., 2009).[13] Fourth, we re-estimate the baseline models using a sub-sample of individuals who were 11 years old at most three years prior to or at most three years after the implementation of the reform (in their municipality of residence). The idea is that with this restriction we rely only on individuals who participated in compulsory schooling closest to the implementation of the reform (i.e. diagonal elements in Table 1) to increase the homogeneity of the treated and untreated birth cohorts (see also Table A4). Fifth, we augment the baseline model by expanding the sample to include also individuals born in 1960–1961, as in Pekkarinen et al. (2009). In comparison, our preferred sample focuses on individuals born in 1962–1966, the same years as in Pekkala Kerr et al. (2013). These robustness checks are reported in the appendix for all-cause mortality (Table A5), suicides (Table A6), and mental health outcomes (Table A7), and they are briefly commented in the results section.

[13]We also estimate the models with six regional dummies instead of the full set of 465 municipal dummies (cf. Figure 2). We do this because previous studies using the Finnish comprehensive school reform only had access to the aggregated regional classification (i.e., six regional dummies) due to data limitations. The results are robust to this (see Appendix Table A2). The results are also robust to using additional birth month dummies or birth month by birth year dummies (Table A3), as estimated in Lager et al. (2016).

5 Results

5.1 All-cause Mortality and Suicides

To resolve the overall health effects of the tracking reform, Table 3 reports the estimates in which all-cause mortality is used as the outcome variable. We find no evidence that the comprehensive school reform affected mortality by age 45. The coefficients are close to zero and precisely estimated.[14] For males, we can rule out increases of 0.17 percentage points or decreases of 0.5 percentage points at the 5% risk level, relative to the mean outcome of 4.6%. For females, we can rule out increases of 0.4 percentage points or decreases of 0.15 percentage points, relative to the mean outcome of 1.8%.[15] Finally, despite the universal health care system in Finland, the baseline differences in mortality by parental education are substantial – males from low-educated families have a three times higher mortality than those from highly-educated families – but the treatment effect is insignificant across groups. Table A5 in Appendix B provides evidence for the robustness of these results (e.g., adding controls and excluding some observations). Our conclusions are also robust to the use of Cox proportional hazards models (Table A10).

However, all-cause mortality is only a coarse measure of health status at the individual level, and the aggregate measure may disguise substantial responses among specific causes of mortality. For this reason, we redo the analysis using the incidence of suicide by age 45 as the dependent variable, since we are mostly interested in severe mental health outcomes (Table 4). Echoing the aggregate results for all-

[14] During our research project we became aware of a working paper based on the Finnish comprehensive school reform that uses all-cause mortality as the outcome (Ravesteijn et al., 2017). For females, our findings are consistent with the results of Ravesteijn et al. (2017). For males, they find occasionally negative or positive effects, depending on the specification, but their results are much less precise due to the use of a 11% random sample. In our paper, we focus on the impact of the reform on mental health, because the set of peers to which students are exposed to is a potentially important determinant of mental health and well-being in adulthood.

[15] The aggregate results for both genders together are reported in Table A9. None of the estimated effects for mortality are significant at the standard 5% level.

Table 3: Effect of the Reform on All-cause Mortality by Age 45

		Parental education		
	Full sample	Low	Mid	High
		Males		
Treatment effect	-0.0016	-0.0034	0.0022	-0.0014
	(0.0017)	(0.0022)	(0.0018)	(0.0017)
R squared	0.0032	0.0064	0.0099	0.0118
Mean outcome	0.0458	0.0339	0.0132	0.0102
Observations	186,777	94,037	47,224	36,282
		Females		
Treatment effect	0.0007	-0.0005	0.0009	-0.0008
	(0.0011)	(0.0013)	(0.0011)	(0.0012)
R squared	0.0025	0.0051	0.0098	0.0112
Mean outcome	0.0176	0.0122	0.0050	0.0047
Observations	179,537	90,881	46,092	34,828

Note: The full sample also includes individuals for whom information about parents (such as parental education) is missing. All models include cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. None of the coefficients are significant at 10% level.

Table 4: Effect of the Reform on Suicides by Age 45

		Parental education		
	Full sample	Low	Mid	High
		Males		
Treatment effect	-0.0002	-0.0016	0.0013	0.0003
	(0.0008)	(0.0011)	(0.0010)	(0.0010)
R squared	0.0032	0.0054	0.0077	0.0132
Mean outcome	0.0129	0.0099	0.0034	0.0023
Observations	186,777	94,037	47,224	36,282
		Females		
Treatment effect	-0.0003	-0.0006	-0.0001	0.0002
	(0.0005)	(0.0005)	(0.0004)	(0.0007)
R squared	0.0025	0.0044	0.0095	0.0148
Mean outcome	0.0030	0.0021	0.0007	0.0009
Observations	179,537	90,881	46,092	34,828

Note: The full sample also includes individuals for whom information about parents (such as parental education) is missing. All models include cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. None of the coefficients are significant at 10% level.

cause mortality, we find no significant effects. For instance, for males, we can rule out increases of 0.14 percentage points or decreases of 0.18 percentage points, relative to the mean outcome of 1.3%. Table A6 shows that the results for the full sample are robust to the variations of the model.

5.2 Mental Health Disorders

Table 5 reports the results that use as the outcome variable the incidence of mental health related hospitalizations over the age range 16–45. The average of the outcome variable is 10% and 6% for males and females, respectively. These figures indicate that these incidents are not rare in our data.

Table 5: Effect of the Reform on Mental Health Related Hospitalizations at Ages 16–45

		Parental education		
	Full sample	Low	Mid	High
		Males		
Treatment effect	0.0027	0.0074**	0.0012	-0.0058
	(0.0025)	(0.0036)	(0.0045)	(0.0043)
R squared	0.0054	0.0092	0.0136	0.0128
Mean outcome	0.0997	0.1080	0.0847	0.0735
Observations	186,777	94,037	47,224	36,282
		Females		
Treatment effect	0.0029	0.0022	-0.0007	*0.0095***
	(0.0019)	(0.0025)	(0.0034)	(0.0041)
R squared	0.0042	0.0071	0.0106	0.0133
Mean outcome	0.0564	0.0593	0.0498	0.0521
Observations	179,537	90,881	46,092	34,828

Note: The full sample also includes individuals for whom information about parents (such as parental education) is missing. All models include cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. ** = significant at 5% (all two-sided tests). Coefficients in italics survive a Romano-Wolf (2005) correction for multiple hypotheses testing on 10% significance level (see Table A8).

Our results show that the reform led to an increase in mental health related hospi-

talizations, but only for some subgroups. The first finding is that for men coming from low-educated families individuals exposed to late tracking have a 0.74 percentage points higher probability of severe mental health disorders that results in hospitalization compared to individuals educated in the pre-reform system, i.e. engaged in early tracking. The size of the effect represents a 7% increase relative to a mean outcome of 10.8%. The effect is robust to using an outcome measure that captures also the intensive margin and not just the extensive one, i.e. the number of days spent in the hospital for mental health reasons between age 16 and 45 (see Table A11 in Appendix B). However, the effect for males with low-educated parents is no longer statistically significant after adjusting the standard errors for a Romano-Wolf (2005) correction for multiple hypotheses.

The second finding is that for women from highly-educated families the probability of severe mental health disorders increased by 0.95 percentage points as a consequence of the postponement of the tracking age from 11 to 16. We stress this finding for several reasons. First, the quantitative magnitude of the effect is rather large, given the low baseline probability of mental health disorders for women coming from highly-educated families (5.2%). Second, the effect remains significant (at 10% level) after adjusting the standard errors for a Romano-Wolf (2005) correction for multiple hypotheses (Table A8). Third, the effect remains intact, regardless of whether we control for the incidence of prior mental health disorders and/or parents' mental health status during the pre-reform period (Table A7). Fourth, the finding for females is also robust to several other sensitivity checks, e.g. accounting for the full set of region-specific linear time trends (Table A7). Fifth, the estimated dynamic response models show that the finding is not driven by pre-reform trends (Figure A5). Therefore, it is unlikely that the result is affected by unobserved regional characteristics that are potentially correlated with the implementation of the reform. Sixth, the finding remains intact if individuals are classified into treatment

based on their municipality of birth instead of municipality at age 11 (Table A12).

The timing of the reform together with longitudinal data enable us to examine the effects over the life cycle. Hence, we report the results for relevant age categories in Table 6. We find that mental health disorders are significantly more likely at ages 36 to 45 for women from highly-educated families who have been exposed to the post-reform school system compared to those who were educated in the pre-reform system. However, for this affected group, we do not find significant effects during the school years. Additionally, for men coming from low-educated families, the positive effect of the comprehensive school reform on the probability to be hospitalized for mental health reasons peaks at ages 26 to 35, but there is also a significant effect at later ages, between the ages of 36 and 45.

Table 7 explores whether any specific disorder drives the estimated treatment effects. We observe that for women coming from highly-educated families, the probability of depressive disorders has increased significantly. Conversely, for men who come from low-educated families, the overall increase in hospitalizations is driven by the increase in alcohol-related mental disorders. Alcohol abuse may be caused by self-medication related to perceived stress (Enoch, 2011). The pattern that we observe is plausible, because alcohol-related mental disorders and depressive disorders are the most prevalent mental health problems in our data for men and women, respectively (Table 2).

6 Potential Mechanisms

The reform could have affected mental health through various channels discussed in Section 2. On the positive side, in addition to exposure to higher-achieving peers and access to a larger set of cognitive skills for students who would have started vocational training at age 11 absent the reform, the key theoretical argument in favor

Table 6: Effect of the Reform on Mental Health Related Hospitalizations over the Life Cycle

	Full sample	Parental education		
		Low	Mid	High
		Males		
a) Age 11–15	-0.0002	0.0011	-0.0016	-0.0016
	(0.0008)	(0.0009)	(0.0012)	(0.0013)
b) Age 16–25	0.0024	0.0041	-0.0007	0.0027
	(0.0021)	(0.0031)	(0.0040)	(0.0028)
c) Age 26–35	0.0020	*0.0062****	0.0003	-0.0057
	(0.0016)	(0.0022)	(0.0031)	(0.0035)
d) Age 36–45	0.0018	0.0046*	-0.0018	-0.0032
	(0.0016)	(0.0023)	(0.0031)	(0.0032)
Observations	186,777	94,037	47,224	36,282
		Females		
a) Age 11–15	-0.0002	-0.0009	*0.0024***	-0.0017
	(0.0006)	(0.0008)	(0.0010)	(0.0014)
b) Age 16–25	0.0005	0.0007	-0.0012	0.0009
	(0.0011)	(0.0015)	(0.0021)	(0.0023)
c) Age 26–35	0.0020	0.0011	0.0034	0.0026
	(0.0012)	(0.0016)	(0.0024)	(0.0029)
d) Age 36–45	0.0012	0.0018	-0.0040	0.0058**
	(0.0014)	(0.0021)	(0.0025)	(0.0027)
Observations	179,537	90,881	46,092	34,828

Note: The table reports the treatment effect of the reform. Each row corresponds to a different outcome, and each column to a different sample. See Table A1 for the mean values of the outcome variables by parental education and Figure A6 for graphical illustration. The full sample also includes individuals for whom information about parents (such as parental education) is missing. Individuals who died before the observation period were removed from the estimation sample. All models include cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. * = significant at 10%; ** = significant at 5%; *** = significant at 1% (all two-sided tests). Coefficients in italics survive a Romano-Wolf (2005) correction for multiple hypotheses testing on 10% significance level (see Table A8).

Table 7: Effect of the Reform on Specific Mental Health Related Hospitalizations at Ages 16–45

	Full sample	Parental education		
		Low	Mid	High
		Males		
a) Schizophrenia	0.0005	0.0018	-0.0013	0.0014
	(0.0008)	(0.0013)	(0.0017)	(0.0018)
b) Other non-affective psychosis	-0.0005	0.0013	-0.0007	-0.0028
	(0.0009)	(0.0015)	(0.0022)	(0.0020)
c) Bipolar disorder	0.0007	0.0011	0.0015	-0.0012
	(0.0005)	(0.0008)	(0.0010)	(0.0013)
d) Depressive disorder	0.0013	0.0017	0.0007	-0.0033
	(0.0010)	(0.0015)	(0.0019)	(0.0021)
e) Anxiety, stress, neurotic disorder	0.0010	0.0026	-0.0015	-0.0013
	(0.0013)	(0.0018)	(0.0021)	(0.0018)
f) Substance-use disorder	*0.0038***	0.0048**	0.0015	0.0021
	(0.0016)	(0.0023)	(0.0030)	(0.0025)
Observations	186,777	94,037	47,224	36,282
		Females		
a) Schizophrenia	0.0004	0.0013	0.0001	-0.0017
	(0.0008)	(0.0010)	(0.0013)	(0.0020)
b) Other non-affective psychosis	0.0006	0.0015	-0.0016	-0.0000
	(0.0010)	(0.0013)	(0.0020)	(0.0021)
c) Bipolar disorder	0.0005	-0.0000	-0.0007	0.0026*
	(0.0005)	(0.0007)	(0.0011)	(0.0014)
d) Depressive disorder	0.0015	0.0009	0.0004	*0.0062****
	(0.0009)	(0.0015)	(0.0020)	(0.0022)
e) Anxiety, stress, neurotic disorder	0.0001	0.0003	0.0006	-0.0010
	(0.0008)	(0.0011)	(0.0012)	(0.0016)
f) Substance-use disorder	0.0009	0.0021	-0.0013	0.0024
	(0.0009)	(0.0014)	(0.0018)	(0.0015)
Observations	179,537	90,881	46,092	34,828

Note: The table reports the treatment effect of the reform on various outcomes. Each row corresponds to a different outcome, and each column corresponds to a different sample. See Table A1 for the mean values of the outcome variables by parental education. The full sample also includes individuals for whom information about parents (such as parental education) is missing. All models include cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. * = significant at 10%; ** = significant at 5%; *** = significant at 1% (all two-sided tests). Coefficients in italics survive a Romano-Wolf (2005) correction for multiple hypotheses testing on 10% significance level (see Table A8).

of postponing tracking is to allow the system to have more relevant and accurate information about abilities and comparative advantages when matching students to a certain type of education (Brunello et al., 2007). Increasing the tracking age should thus improve the efficiency of the match and subsequent labor-market outcomes, which could translate into better mental health later in adulthood. By contrast, we do not find any improvement in mental health as a result of the comprehensive school reform.

In fact, we do not find that the reform led to an improvement in education or labor market outcomes; see Tables A14 and A15. For males coming from low-educated families, we even find some negative effects of the reform on the number of years of schooling and economic outcomes. These results suggest that the theoretical mechanisms related to the efficiency of matching students to suitable education are not at play here.

To pin down the potential mechanisms further, we follow Acharya et al. (2016) and estimate the average controlled direct effects of the reform as described in Appendix C. The negative effects on mental health for females that we find remain intact even when controlling for the education and income mediators, as can be seen from Table 8. Thus, for females the observed effects on mental health cannot be explained by the education/income channels.

Instead, we conjecture that peer effects is a potential mechanism driving the observed adverse effects for females from highly-educated families. These females, prior to the reform, would most likely have been tracked into the selective academic curriculum from age 11 to 16 and exposed only to high-ability peers. After the reform, they are exposed to comprehensive school and to peers who, absent the reform, would have pursued the vocational track. Presumably, this new set of peers, on average, is less well-behaved and of lower academic ability. This may result in a higher probability of being exposed to disruptive behaviors. Prior work

Table 8: The Estimated Reform Effect, Controlling for Education and Income Mediators Later in Life

Outcome / Sample	Baseline estimates	Mediation analysis	
		Education	Educ. & Income
Outcome	*Panel (A): Males, Low-educated parents*		
Death by age 45	-0.0034	-0.0040*	-0.0042**
	(0.0022)	(0.0022)	(0.0021)
Suicide by age 45	-0.0016	-0.0018	-0.0017*
	(0.0011)	(0.0012)	(0.0010)
MHD at age 16–45	0.0074**	0.0061*	0.0060
	(0.0036)	(0.0036)	(0.0037)
MHD at age 26–35	*0.0062***	*0.0056***	0.0049*
	(0.0022)	(0.0022)	(0.0022)
MHD at age 36–45	0.0046*	0.0038	0.0032
	(0.0023)	(0.0024)	(0.0024)
Substance-use disorder	0.0048**	0.0041*	0.0033
at age 16-45	(0.0023)	(0.0024)	(0.0023)
Outcome	*Panel (B): Females, Highly-educated parents*		
Death by age 45	-0.0008	-0.0008	-0.0008
	(0.0012)	(0.0013)	(0.0012)
Suicide by age 45	0.0002	0.0003	0.0005
	(0.0007)	(0.0007)	(0.0007)
MHD at age 16–45	*0.0095***	*0.0096***	*0.0097***
	(0.0041)	(0.0041)	(0.0042)
MHD at age 26–35	0.0026	0.0026	0.0025
	(0.0029)	(0.0029)	(0.0029)
MHD at age 36–45	0.0058**	0.0057**	0.0061**
	(0.0027)	(0.0028)	(0.0029)
Depression at age 16–45	*0.0062****	*0.0063****	*0.0065****
	(0.0022)	(0.0022)	(0.0022)

Note: We report the results only on samples (outcomes) that show significant effects in Tables 3–7. MHD = Mental health disorder requiring hospitalization spell. Columns (2) and (3) present controlled direct effects based on Acharya et al. (2016). Education controls include the years of schooling and a dummy for having a high school degree. Income control is the log of taxable income at ages 26–35. Each cell reports the estimated effect of the comprehensive school reform from separate models. All models include cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. In columns (2) and (3), standard errors have been bootstrapped using 1,000 replications and clustered at the municipal level. * = significant at 10%; ** = significant at 5%; *** = significant at 1% (all two-sided tests). Coefficients in italics survive a Romano-Wolf (2005) correction for multiple hypotheses testing on 10% significance level (see Table A13).

has shown the negative effects of being bullied at school on mental health in adulthood (Sigurdson et al., 2014). And in particular girls, rather than boys, who are victimized by bullying, even infrequently, are more likely to suffer from depression symptoms in adulthood (Brunstein Klomek et al., 2007, p. 43).

An alternative interpretation for the observed adverse effects relies on the more intensive competition induced by the reform. Indeed, the reform implied that children from disadvantaged background now had a better chance to compete for places in the higher education system: it equalized opportunities by socioeconomic status and decreased the intergenerational correlation of earnings (Pekkarinen et al., 2009). After the reform, children from privileged backgrounds faced more effective competition from the most talented children from disadvantaged backgrounds. This change may have affected females more as females tend to respond less favorably to more intense competition than males (Niederle and Vesterlund, 2011).

Regarding the weaker and less robust effects on males from disadvantaged background, we find that after controlling for education and income effects, the reform had negligible impact on males with low-educated parents at age 16–45 (see the results from the mediator analysis in Table 8).

7 Conclusion

We contribute to the small literature on the effects of school tracking regimes in Europe. Our results are based on a comprehensive school reform that was rolled out gradually across Finnish municipalities over the period 1972–1977. The reform resulted in children from different socioeconomic backgrounds and potentially different academic abilities held in the same classes for five extra years. Although the generalization of our estimates to current policy settings is not straightforward, the

long-run health effects of school reforms can only be identified for birth cohorts treated many decades ago. As outcome variables, we focus on severe mental disorders, which cause substantial costs to the health care system and lead to lasting negative outcomes at the individual level such as poor labor market attachment (Ettner et al., 1997).

We find no significant effect on mental health on average, even though the average zero effect is precisely estimated. Heterogeneity analysis shows that postponing the age at which students are tracked had an adverse long-run effect on mental health outcomes for females from highly-educated families who were more likely to be hospitalized for depression after the reform. Thus, increasing the tracking age may come at the cost of negative mental health effects for some groups. We further find that the finding for females is not accounted for by changes in education or income induced by the reform. Instead, we propose that a possible explanation for our finding are peer effects for the affected females. Exploring this mechanism is a promising avenue for future research.

Acknowledgments

We thank comments from Artturi Björk, Hannu Karhunen, Sari Kerr, Kristian Koerselman, Jani-Petri Laamanen, Krista Riukula, Kerstin Schneider, Tuomo Suhonen, and Annual Conference of the Royal Economic Society (2019), EALE (2018), IIPF (2018) and Irish Economic Association (2018) as well as seminar participants at the Queen's University Belfast, ESRI, Labour Institute for Economic Research, University of Tampere, University of Jyväskylä, and University of Queensland for their useful comments. This study is part of a project supported by the Strategic Research Council at the Academy of Finland (project nos. 293120 and 314208), Alfred Kordelin Foundation, OP Group Research Foundation, and Yrjö Jahnsson Founda-

tion (no. 6621).

References

Acharya, A., M. Blackwell, and M. Sen (2016): "Explaining causal findings without bias: Detecting and assessing direct effects," *American Political Science Review*, 110, 512–529.

Angrist, J. D. and A. B. Krueger (1991): "Does compulsory school attendance affect schooling and earnings?" *The Quarterly Journal of Economics*, 106, 979–1014.

Betts, J. (2011): "The economics of tracking in education," in *Handbook of Economics of Education*, ed. by E. Hanushek, S. Machin, and L. Woessmann, Elsevier, vol. 3, 341–381.

Bijwaard, G. E., M. Myrskylä, and P. Tynelius (2018): "The impact of mental problems on mortality and how it is moderated by education," IZA Discussion Paper No. 11591.

Brunello, G., M. Giannini, and K. Ariga (2007): "The optimal timing of school tracking," in *Schools and the Equal Opportunity Problem*, ed. by P. Peterson and L. Woessmann, MIT Press, 129–156.

Brunstein Klomek, A., F. Marrocco, M. Kleinman, I. S. Schonfeld, and M. S. Gould (2007): "Bullying, depression, and suicidality in adolescents," *Journal of the American Academy of Child & Adolescent Psychiatry*, 46, 40–49.

Bubonya, M., D. Cobb-Clark, and M. Wooden (2017): "Mental health and productivity at work: Does what you do matter?" *Labour Economics*, 46, 150–165.

Cicala, S., R. G. Fryer, and J. L. Spenkuch (2018): "Self-selection and comparative advantage in social interactions," *Journal of the European Economic Association*, forthcoming.

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

Cutler, D. and A. Lleras-Muney (2008): "Education and health: Evaluating theories and evidence," in *Making Americans Healthier: Social and Economic Policy as Health Policy*, ed. by J. House, R. Schoeni, G. Kaplan, and H. Pollack, New York: Russell Sage Foundation, 29–60.

Elsner, B. and I. E. Isphording (2018): "Rank, sex, drugs and crime," *Journal of Human Resources*, forthcoming.

Enoch, M.-A. (2011): "The role of early life stress as a predictor for alcohol and drug dependence," *Psychopharmacology*, 214, 17–31.

Ettner, S. L., R. G. Frank, and R. C. Kessler (1997): "The impact of psychiatric disorders on labor market outcomes," *Industrial and Labor Relations Review*, 51, 64–81.

Frank, R. G. and T. G. McGuire (2000): "Economics and mental health," in *Handbook of Health Economics*, ed. by A. J. Culyer and J. P. Newhouse, Elsevier, vol. 1, chap. 16, 893–954, 1 ed.

Galama, T. J., A. Lleras-Muney, and H. v. Kippersluis (2018): "The effect of education on health and mortality: a review of experimental and quasi-experimental evidence," Working Paper 24225, National Bureau of Economic Research.

Gerdtham, U.-G. (1997): "Equity in health care utilization: Further tests based on hurdle models and Swedish micro data," *Health Economics*, 6, 303–319.

Hakulinen, C., M. Elovainio, M. Arffman, S. Lumme, S. Pirkola, I. Keskimki, K. Manderbacka, and P. Böckerman (2019): "Mental disorders and long-term labour market outcomes: nationwide cohort study of 2 055 720 individuals," *Acta Psychiatrica Scandinavica*, 140, 371–381.

Hall, C. (2012): "The effects of reducing tracking in upper secondary school: Evidence from a large-scale pilot scheme," *Journal of Human Resources*, 47, 237–269.

Henriksson, M., H. Aro, M. Marttunen, M. Heikkinen, E. Isometsä, K. Kuoppasalmi, and J. Lönnqvist (1993): "Mental disorders and comorbidity in suicide," *American Journal of Psychiatry*, 150, 935–940.

Jones, A. M., J. E. Roemer, and P. R. Dias (2014): "Equalising opportunities in health through educational policy," *Social Choice and Welfare*, 43, 521–545.

Lager, A., D. Seblova, D. Falkstedt, and M. Lövdén (2016): "Cognitive and emotional outcomes after prolonged education: A quasi-experiment on 320 182 Swedish boys," *International Journal of Epidemiology*, 46, 303–311.

Lahti, A., P. Räsänen, K. Riala, S. Keränen, and H. Hakko (2011): "Youth suicide trends in Finland, 1969-2008," *Journal of Child Psychology and Psychiatry*, 52, 984–991.

Lahti, R. and A. Penttilä (2001): "The validity of death certificates: Routine validation of death certification and its effects on mortality statistics," *Forensic Science International*, 115, 15–32.

Layard, R. (2013): "Mental health: The new frontier for labour economics," *IZA Journal of Labor Policy*, 2, 1–16.

Meghir, C., M. Palme, and E. Simeonova (2018): "Education and mortality: Evidence from a social experiment," *American Economic Journal: Applied Economics*, 10, 234–256.

Niederle, M. and L. Vesterlund (2011): "Gender and competition," *Annual Review of Economics*, 3, 601–630.

OECD (2004): *Learning for Tomorrow's World – First Results from PISA 2003*, Paris.

Oreopoulos, P. and K. G. Salvanes (2011): "Priceless: The nonpecuniary benefits of schooling," *Journal of Economic Perspectives*, 25, 159–84.

Pedersen, C. B., O. Mors, A. Bertelsen, B. L. Waltoft, E. Agerbo, J. J. McGrath, P. B. Mortensen, and W. W. Eaton (2014): "A comprehensive nationwide study

of the incidence rate and lifetime risk for treated mental disorders," *JAMA Psychiatry*, 71, 573–581.

Pekkala Kerr, S., T. Pekkarinen, and R. Uusitalo (2013): "School tracking and development of cognitive skills," *Journal of Labor Economics*, 31, 577–602.

Pekkarinen, T. (2008): "Gender differences in educational attainment: Evidence on the role of tracking from a Finnish quasi-experiment," *Scandinavian Journal of Economics*, 110, 807–825.

Pekkarinen, T., R. Uusitalo, and S. Pekkala Kerr (2009): "School tracking and intergenerational income mobility: Evidence from the Finnish comprehensive school reform," *Journal of Public Economics*, 93, 965–973.

Pirkola, S., R. Sund, E. Sailas, and K. Wahlbeck (2009): "Community mental-health services and suicide rate in Finland: A nationwide small-area analysis," *The Lancet*, 373, 147–153.

Ravesteijn, B., H. van Kippersluis, M. Avendano, P. Martikainen, H. Vessari, and E. van Doorslaer (2017): "The impact of later tracking on mortality by parental income in Finland," Working Paper 2017-030, Tinbergen Institute.

Ritter, P. L., A. L. Stewart, H. Kaymaz, D. S. Sobel, D. A. Block, and K. R. Lorig (2001): "Self-reports of health care utilization compared to provider records," *Journal of Clinical Epidemiology*, 54, 136–141.

Romano, J. P. and M. Wolf (2005): "Stepwise Multiple Testing as Formalized Data Snooping," *Econometrica*, 73, 1237–1282.

Sahlberg, P. (2014): *Finnish Lessons 2.0: What Can the World Learn from Educational Change in Finland?*, New York: Teachers College Press, Columbia University, 2 ed.

Santavirta, T., N. Santavirta, T. S. Betancourt, and S. E. Gilman (2015): "Long term mental health outcomes of Finnish children evacuated to Swedish families

during the second world war and their non-evacuated siblings: cohort study," *BMJ*, 350, g7753.

Sareen, J., F. Jacobi, B. J. Cox, S.-L. Belik, I. Clara, and M. B. Stein (2006): "Disability and Poor Quality of Life Associated With Comorbid Anxiety Disorders and Physical Conditions," *Archives of Internal Medicine*, 166, 2109–2116.

Sigurdson, J., J. Wallander, and A. Sund (2014): "Is involvement in school bullying associated with general health and psychosocial adjustment outcomes in adulthood?" *Child Abuse & Neglect*, 38, 1607–1617.

Somerkivi, U. (1982): *Peruskoulu. Synty, Kehittyminen ja Tulevaisuus [Comprehensive School: Origins, Development and Future]*, Vantaa: Kunnallispaino.

Sund, R. (2012): "Quality of the Finnish Hospital Discharge Register: A systematic review," *Scandinavian Journal of Public Health*, 40, 505–515.

Suvisaari, J., T. Aalto-Setälä, A. Tuulio-Henriksson, et al. (2009): "Mental disorders in young adulthood," *Psychological Medicine*, 39, 287–299.

Ustun, T. and S. Chatterji (2001): "Global burden of depressive disorders and future projections," in *Depression: Social and Economic Timebomb*, ed. by A. Dawson and A. Tylee, BMJ Books, 31–43.

Vigo, D., G. Thornicroft, and R. Atun (2016): "Estimating the true global burden of mental illness," *The Lancet Psychiatry*, 3, 171–178, doi: 10.1016/S2215-0366(15)00505-2; 24.

Whiteford, H., L. Degenhardt, J. Rehm, et al. (2013): "Global burden of disease attributable to mental and substance use disorders: Findings from the Global Burden of Disease Study 2010," *The Lancet*, 382, 1575–1586.

WHO (2014): *Social Determinants of Mental Health*, Geneva: World Health Organization.

Wooldridge, J. M. (2001): *Econometric Analysis of Cross Section and Panel Data*, MIT Press.

A Appendix: ICD-8, ICD-9, and ICD-10 Diagnosis Codes

i) Schizophrenia:

ICD-8 and ICD-9: 295.0-295.3, 295.5, 295.6, 295.8, 295.9;

ICD-10: F20

ii) Other non-affective psychosis:

ICD-8 and ICD-9: 295.4, 295.7, 297, 298, 299;

ICD-10: F22–F25, F28, F29

iii) Bipolar disorders:

ICD-8: 2961, 2963;

ICD-9: 2962-2967;

ICD-10: F30–F31

iv) Depressive disorders:

ICD-8: 2960, 2962;

ICD-9: 2961, 3004A;

ICD-10: F32–F33, F341

v) Anxiety, stress, neurotic disorders:

ICD-8 and ICD-9: 3000, 3002, 3003;

ICD-10: F40–F42, F430–F431

vi) Substance-use disorders:

ICD-8: 291, 303–304;

ICD-9: 291–292, 303–305;

ICD-10: F10–F19.

B Appendix: Tables and Figures

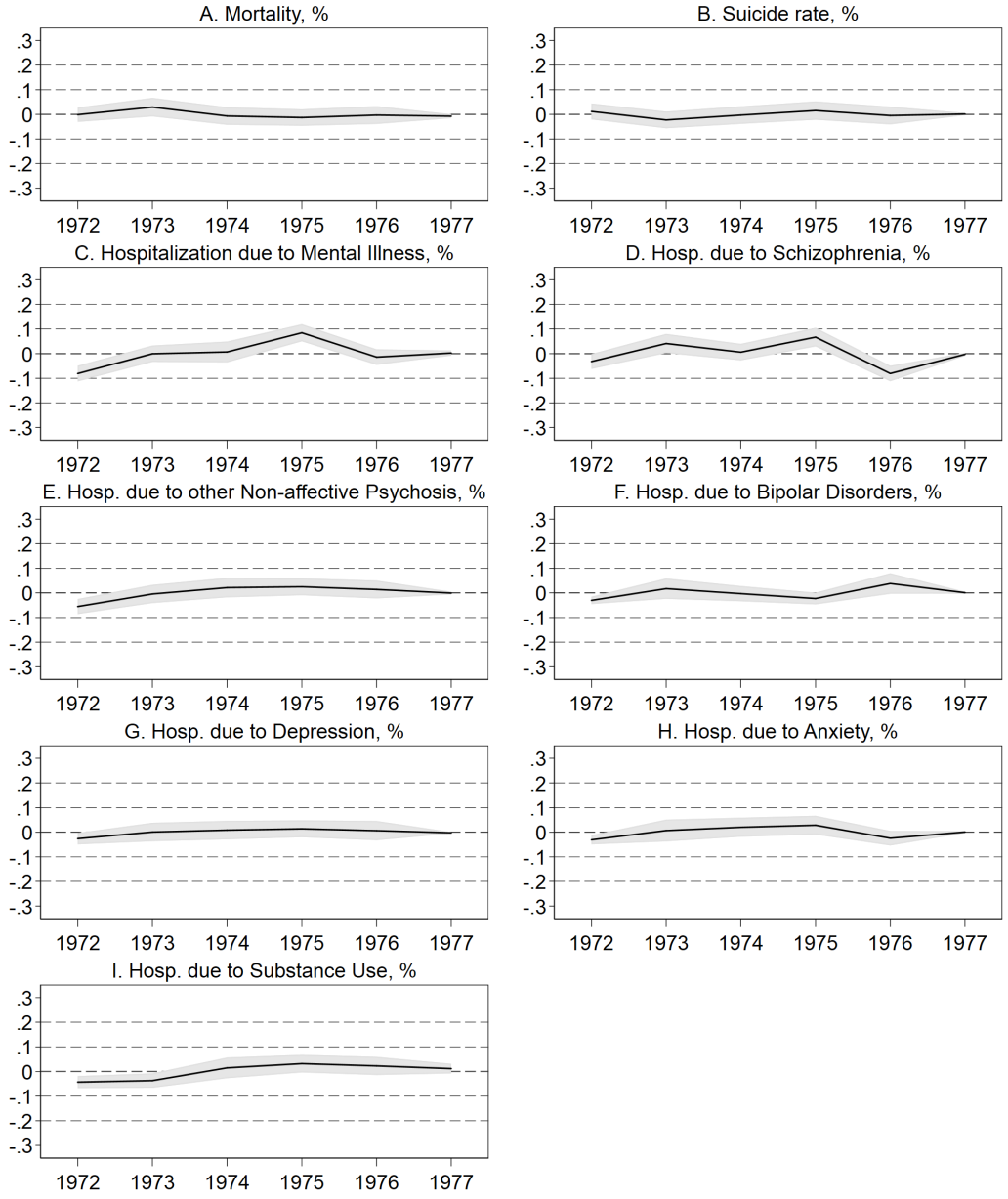
Table A1: Mean Values by Sex and Parental Background

Variable	Males			Females		
	Low-educated parents	Mid-educated parents	Highly-educated parents	Low-educated parents	Mid-educated parents	Highly-educated parents
Mortality						
Death by age 45 or under	0.034	0.013	0.010	0.012	0.005	0.005
Cause of death is suicide by age 45	0.0099	0.0034	0.0023	0.0020	0.0007	0.0009
Hospitalization due to mental health disorder						
At ages 6–10	0.0066	0.0061	0.0046	0.0043	0.0038	0.0027
At ages 11–15	0.0068	0.0057	0.0034	0.0052	0.0041	0.0041
At ages 16–25	0.047	0.038	0.032	0.017	0.014	0.016
At ages 26–35	0.047	0.037	0.032	0.029	0.025	0.027
At ages 36–45	0.056	0.042	0.036	0.036	0.031	0.030
At ages 16–45	0.108	0.085	0.074	0.059	0.050	0.052
Hospitalizations at ages 16–45 due to						
Schizophrenia	0.012	0.010	0.012	0.008	0.008	0.009
Other non-affective psychosis	0.018	0.015	0.016	0.015	0.014	0.016
Bipolar disorder	0.005	0.005	0.005	0.005	0.004	0.006
Depressive disorder	0.019	0.015	0.014	0.020	0.017	0.018
Anxiety, stress, neurotic disorder	0.017	0.014	0.013	0.007	0.006	0.006
Substance-use disorder	0.047	0.033	0.023	0.015	0.010	0.009
Mental health related hospitalization days at ages 16–45						
Unconditional	15.474	12.477	13.709	10.946	9.259	11.784
Conditional on being hospitalized	143.63	147.29	186.49	184.60	186.01	226.15

Table A1: Continued

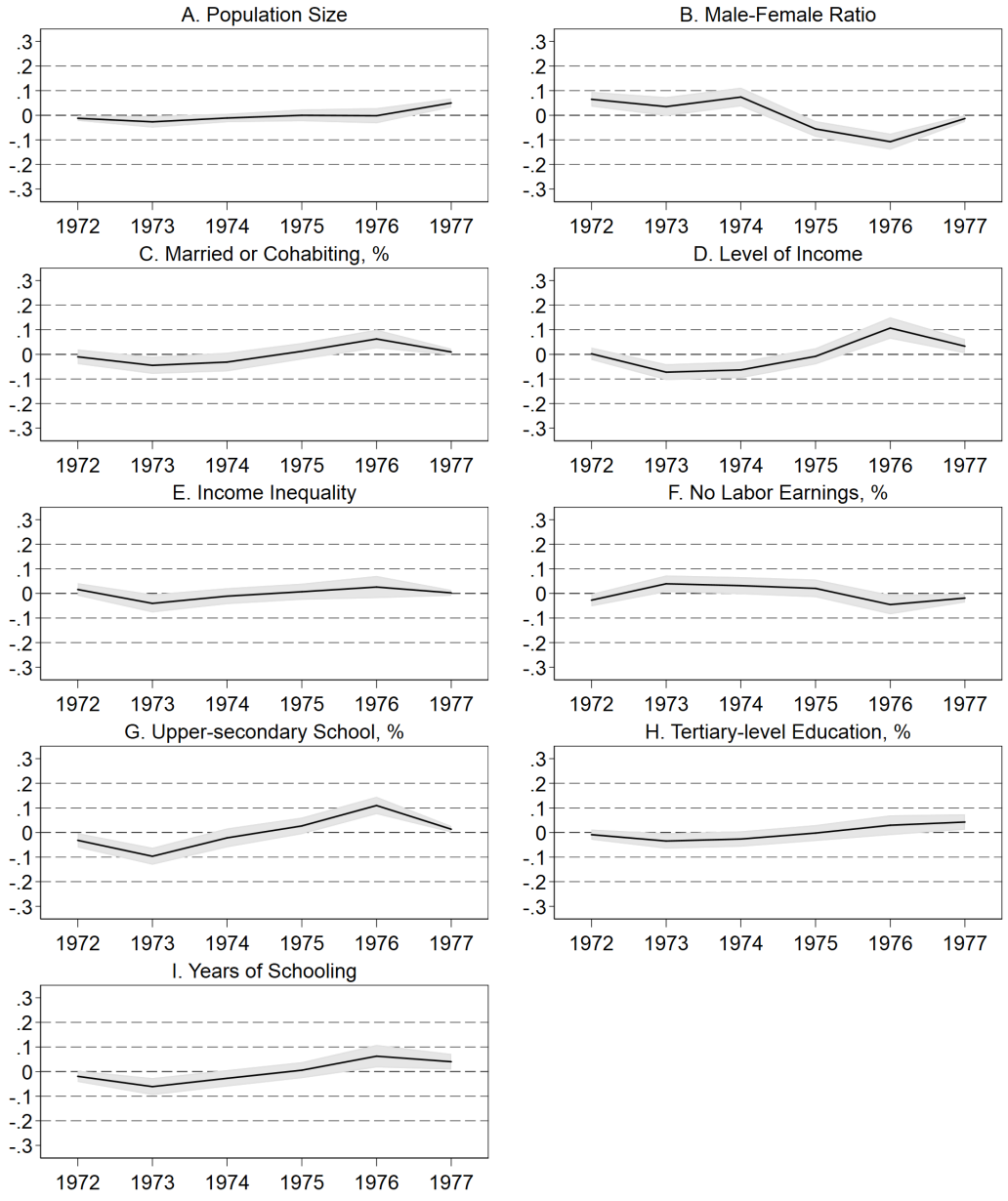
Variable	Males			Females		
	Low-educated parents	Mid-educated parents	Highly-educated parents	Low-educated parents	Mid-educated parents	Highly-educated parents
Educational outcomes						
Years of schooling	12.183	12.790	14.577	12.978	13.532	14.977
Has a high school (HS) degree	0.177	0.265	0.636	0.383	0.505	0.790
HS exam score in native language (percentile)	0.431	0.429	0.503	0.457	0.472	0.568
HS exam score in advanced math (percentile)	0.500	0.511	0.571	0.447	0.455	0.528
HS exam score in basic math (percentile)	0.485	0.493	0.535	0.472	0.487	0.556
Has vocational secondary education (highest)	0.536	0.502	0.232	0.454	0.394	0.194
Has completed vocational college degree (highest)	0.116	0.157	0.192	0.244	0.275	0.264
Has completed university master's degree (highest)	0.055	0.090	0.301	0.086	0.131	0.337
Labor market outcomes						
Average income at ages 26–45 (deflated to 2012)	29.269	32.164	41.657	22.073	23.745	28.691
Average income at ages 26–35 (deflated to 2012)	24.124	25.762	30.460	18.434	19.446	22.480
Average income at ages 36–45 (deflated to 2012)	34.622	38.652	55.018	25.776	28.068	35.012
Employment rate at ages 26–45	0.765	0.806	0.815	0.732	0.760	0.764
Employment rate at ages 26–35	0.739	0.775	0.776	0.668	0.701	0.704
Employment rate at ages 36–45	0.799	0.841	0.858	0.790	0.821	0.824
Parental characteristics						
Parents' average taxable income in 1975 and 1985 (deflated to 2012)	15.716	18.639	31.137	15.711	18.586	30.942
Number of individuals	94,037	47,224	36,282	90,881	46,092	34,828

Figure A1: Baseline Municipality Health Characteristics and Implementation Year of the Reform



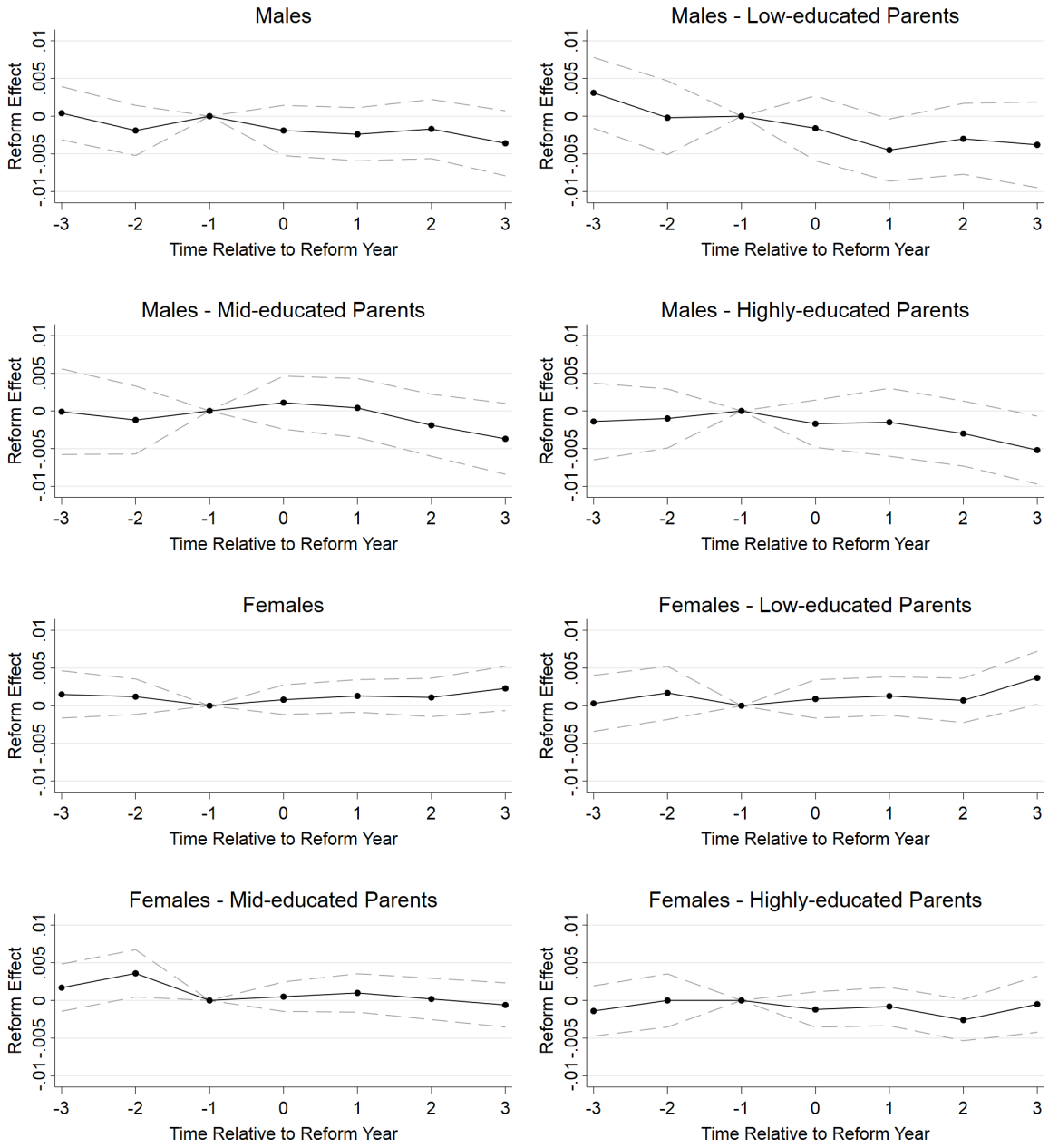
Note: We report results from separate regressions: $R_{jt} = (YEAR_t \times X_j)'\zeta_t + \tau_t + \epsilon_{jt}$, where the dependent variable R_{jt} is an indicator of the timing of the reform (1 if the reform was implemented in year t in municipality j). Explanatory variables contain year fixed effects τ_t and the year dummies $YEAR_t$ interacted with the outcome X_j indicated on the subfigure's title. The outcomes are measured in 1971. The figures plot the coefficients of the interaction terms ζ_t together with 95% confidence intervals (based on robust standard errors clustered at the municipal level). Estimated coefficients have been divided by the standard deviation of the corresponding variable.

Figure A2: Baseline Municipality Characteristics and Implementation Year of the Reform



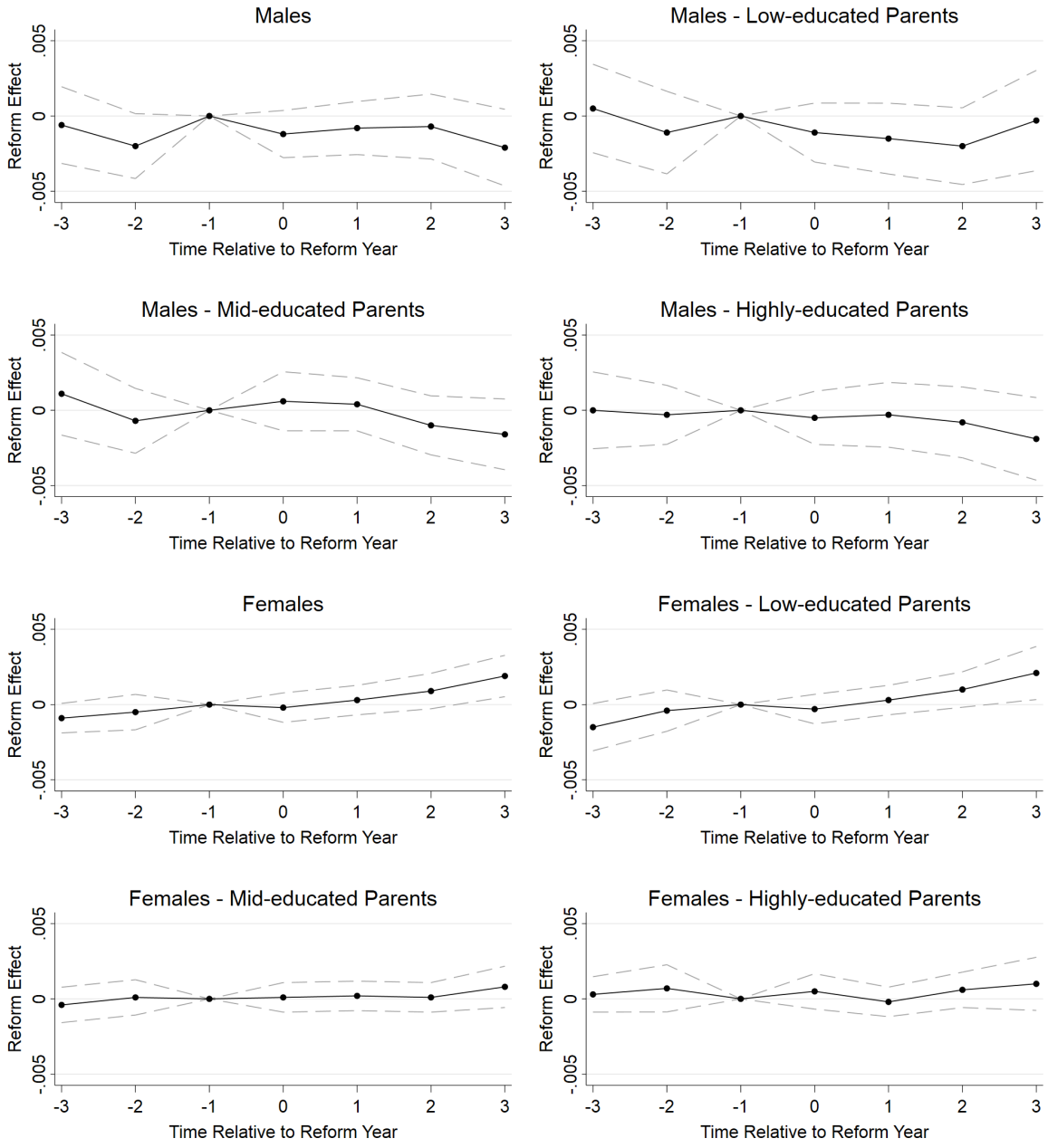
Note: See notes to Figure A1. The municipal characteristics are measured in 1971 or the nearest available year.

Figure A3: Reform Effect on All-cause Mortality at Age 16–45: Leads and Lags Around the Year Before the Reform



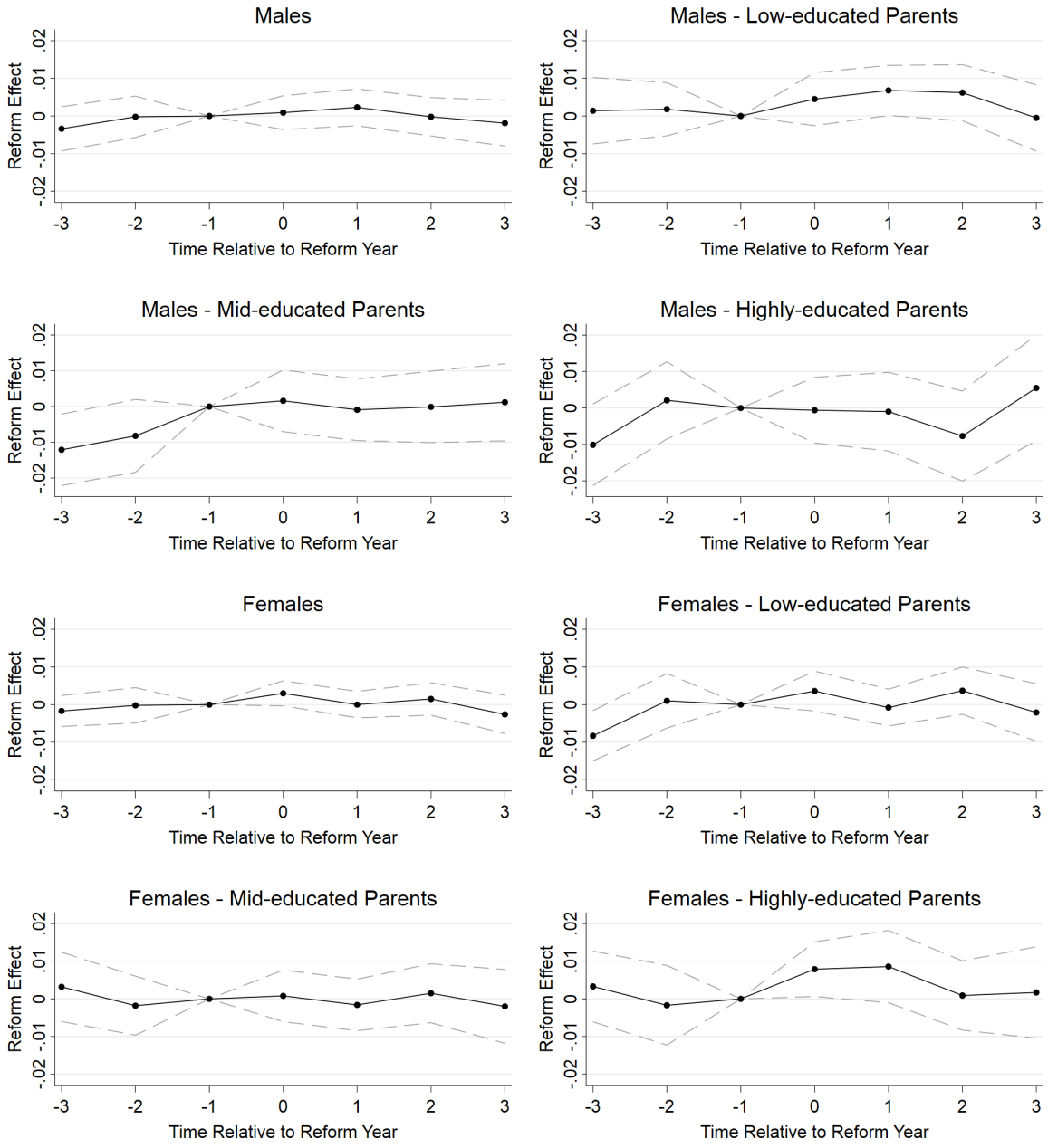
Note: Figures are based on baseline regression models, where the reform dummy is replaced with year dummies for the leads and lags around the reform year. The plotted points are the estimates on the lead and lag dummies. The omitted category is the year before the reform (-1). The estimated effects are reported together with their 95% confidence intervals. Standard errors are clustered at the municipal level.

Figure A4: Reform Effect on Suicide at Age 16–45: Leads and Lags Around the Year Before the Reform



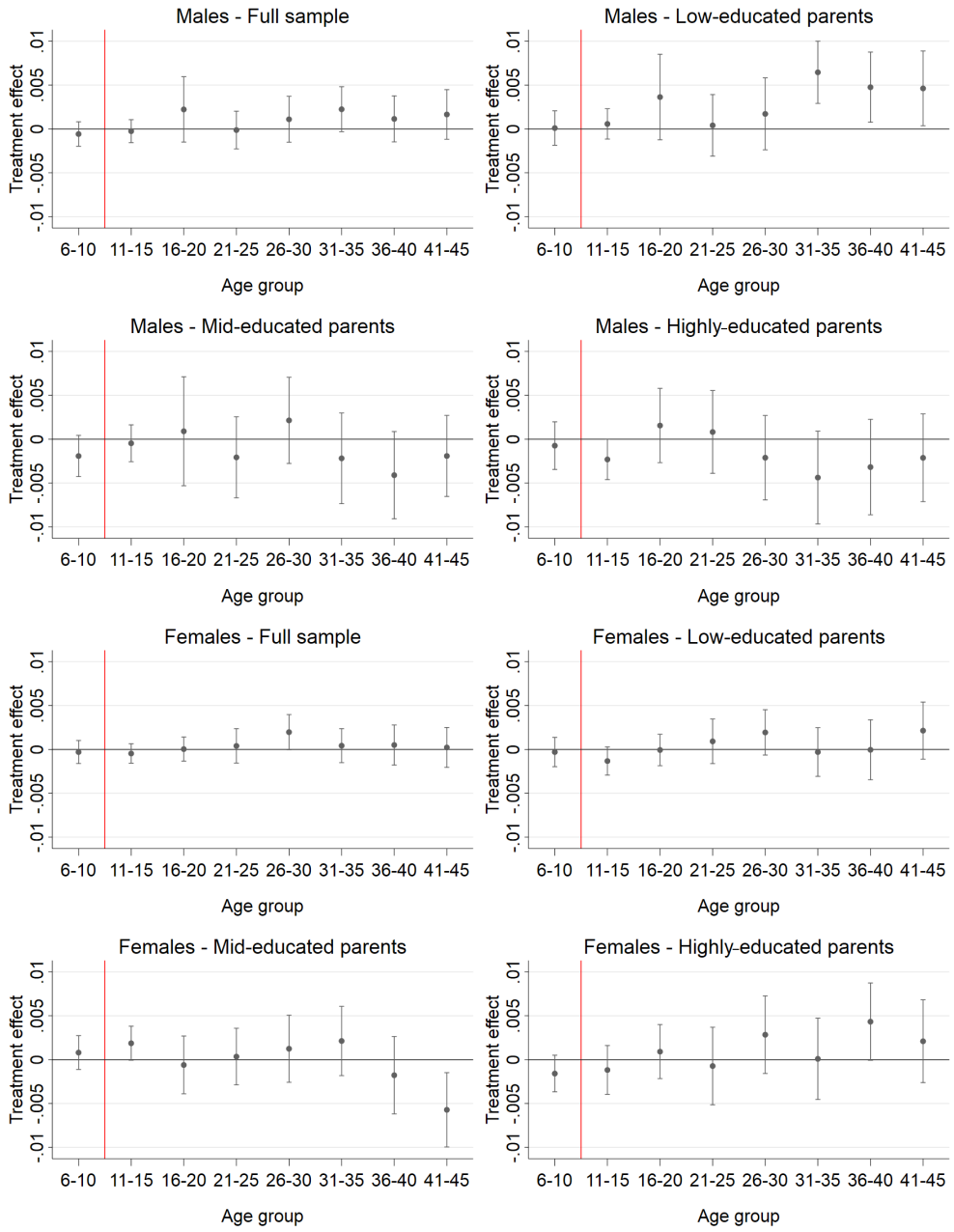
Note: Figures are based on baseline regression models, where the reform dummy is replaced with year dummies for the leads and lags around the reform year. The plotted points are the estimates on the lead and lag dummies. The omitted category is the year before the reform (-1). The estimated effects are reported together with their 95% confidence intervals. Standard errors are clustered at the municipal level.

Figure A5: Reform Effect on Mental Health Related Hospitalizations at Age 16–45: Leads and Lags Around the Year Before the Reform



Note: Figures are based on baseline regression models, where the reform dummy is replaced with year dummies for the leads and lags around the reform year. The plotted points are the estimates on the lead and lag dummies. The omitted category is the year before the reform (-1). The estimated effects are reported together with their 95% confidence intervals. Standard errors are clustered at the municipal level.

Figure A6: Age-specific Treatment Effects on Mental Health Related Hospitalizations by Sex and Parental Background



Note: Each age-specific treatment effect of comprehensive schooling is estimated from separate models. Dependent variable is a dummy variable indicating whether or not an individual was hospitalized due to mental health disorders during the specific age group (e.g. 6–10). The left-hand side of the vertical line denotes pre-treatment period. Treatment effects are reported together with their 95% confidence intervals (based on robust standard errors clustered at the municipal level).

Table A2: Estimated Reform Effect After Controlling for the Six Implementation Regions

Outcome	Full sample	Parental education		
		Low	Mid	High
		Males		
a) Death by age 45	-0.0012	-0.0028	0.0030*	-0.0011
	(0.0017)	(0.0022)	(0.0018)	(0.0017)
b) Suicide by age 45	-0.0001	-0.0015	0.0015	0.0005
	(0.0008)	(0.0011)	(0.0010)	(0.0009)
c) Mental health related	0.0029	0.0079**	0.0011	-0.0052
hospitalizations at age 16–45	(0.0025)	(0.0036)	(0.0045)	(0.0042)
		Females		
a) Death by age 45	0.0007	-0.0004	0.0008	-0.0007
	(0.0011)	(0.0013)	(0.0011)	(0.0012)
b) Suicide by age 45	-0.0002	-0.0005	-0.0001	0.0004
	(0.0005)	(0.0005)	(0.0004)	(0.0007)
c) Mental health related	0.0031*	0.0024	-0.0004	0.0101**
hospitalizations at age 16–45	(0.0019)	(0.0025)	(0.0034)	(0.0040)

Note: Each cell reports the estimated effect of the comprehensive school reform from separate models. In this robustness check, we have replaced the full municipal fixed effects with six implementation region dummies. All models include cohort fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. * = significant at 10%; ** = significant at 5% (all two-sided tests).

Table A3: Estimated Reform Effect after Controlling for Month of Birth

Outcome	Full sample	Parental education		
		Low	Mid	High
Panel A: Controlling for birth month				
		Males		
a) Death by age 45	-0.0016	-0.0034	0.0022	-0.0014
	(0.0017)	(0.0022)	(0.0018)	(0.0017)
b) Suicide by age 45	-0.0002	-0.0016	0.0013	0.0003
	(0.0008)	(0.0011)	(0.0010)	(0.0010)
c) Mental health related	0.0027	0.0074**	0.0012	-0.0058
hospitalizations at age 16–45	(0.0025)	(0.0036)	(0.0045)	(0.0043)
		Females		
a) Death by age 45	0.0007	-0.0006	0.0009	-0.0008
	(0.0011)	(0.0013)	(0.0011)	(0.0012)
b) Suicide by age 45	-0.0003	-0.0006	-0.0001	0.0002
	(0.0005)	(0.0005)	(0.0004)	(0.0007)
c) Mental health related	0.0029	0.0023	-0.0007	0.0096**
hospitalizations at age 16–45	(0.0019)	(0.0025)	(0.0034)	(0.0041)
Panel B: Controlling for birth month and its interaction with birth year				
		Males		
a) Death by age 45	-0.0016	-0.0034	0.0022	-0.0014
	(0.0017)	(0.0022)	(0.0018)	(0.0017)
b) Suicide by age 45	-0.0002	-0.0016	0.0013	0.0003
	(0.0008)	(0.0011)	(0.0010)	(0.0010)
c) Mental health related	0.0028	0.0075**	0.0012	-0.0060
hospitalizations at age 16–45	(0.0025)	(0.0036)	(0.0045)	(0.0043)
		Females		
a) Death by age 45	0.0007	-0.0006	0.0008	-0.0009
	(0.0011)	(0.0013)	(0.0011)	(0.0012)
b) Suicide by age 45	-0.0003	-0.0006	-0.0001	0.0002
	(0.0005)	(0.0005)	(0.0004)	(0.0007)
c) Mental health related	0.0029	0.0024	-0.0006	0.0096**
hospitalizations at age 16–45	(0.0019)	(0.0025)	(0.0034)	(0.0041)

Note: Each cell reports the estimated effect of the comprehensive school reform from separate models. All models include cohort as well as municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. * = significant at 10%; ** = significant at 5% (all two-sided tests).

Table A4: Timing of Schooling Relative to the Year of the Reform Implementation in the Municipality of Residence

Birth cohort	Implementation year in the municipality					
	1972	1973	1974	1975	1976	1977
1960	-1	-2	-3	-4	-5	-6
1961	0	-1	-2	-3	-4	-5
1962	1	0	-1	-2	-3	-4
1963	2	1	0	-1	-2	-3
1964	3	2	1	0	-1	-2
1965	4	3	2	1	0	-1
1966	5	4	3	2	1	0

Note: For each birth cohort, people affected by the reform were those who lived in municipalities where the reform was implemented in the years that correspond to the cells shaded in gray. People who were born in 1965 and lived in a region where the reform was implemented in 1974, entered the 5th grade two years after the first post-reform schooling cohort in the region.

Table A5: Robustness Checks of the Estimated Reform Effect on All-cause Mortality

		Parental education		
	Full sample	Low	Mid	High
		Males		
Baseline results	-0.0016	-0.0034	0.0022	-0.0014
	(0.0017)	(0.0022)	(0.0018)	(0.0017)
a) Controlling for prior mental	-0.0016	-0.0034	0.0023	-0.0014
health disorders	(0.0017)	(0.0022)	(0.0018)	(0.0017)
b) Controlling for prior mental	-0.0016	-0.0034	0.0022	-0.0014
health and parents' mental health	(0.0017)	(0.0022)	(0.0018)	(0.0017)
c) Controlling for prior mental	-0.0014	-0.0033	0.0023	-0.0014
health disorders and its interaction	(0.0017)	(0.0022)	(0.0018)	(0.0017)
with treatment status				
d) Controlling for region-specific	-0.0022	-0.0007	0.0008	-0.0019
linear time trends	(0.0020)	(0.0026)	(0.0023)	(0.0020)
e) Excluding Helsinki	-0.0016	-0.0042	0.0037*	-0.0018
metropolitan area	(0.0020)	(0.0026)	(0.0020)	(0.0021)
f) Excluding if $t < -3$ or $t > 3$	-0.0010	-0.0014	0.0035*	0.0001
	(0.0019)	(0.0024)	(0.0019)	(0.0017)
g) Extending birth cohorts to 1960–66,	0.0000	-0.0011	0.0043**	0.0015
excluding if $t < -3$ or $t > 3$	(0.0018)	(0.0023)	(0.0019)	(0.0018)
		Females		
Baseline results	0.0007	-0.0005	0.0009	-0.0008
	(0.0011)	(0.0013)	(0.0011)	(0.0012)
a) Controlling for prior mental	0.0007	-0.0005	0.0009	-0.0008
health disorders	(0.0011)	(0.0013)	(0.0011)	(0.0012)
b) Controlling for prior mental	0.0007	-0.0005	0.0009	-0.0008
health and parents' mental health	(0.0011)	(0.0013)	(0.0011)	(0.0012)
c) Controlling for prior mental	0.0007	-0.0007	0.0010	-0.0009
health disorders and its interaction	(0.0011)	(0.0013)	(0.0011)	(0.0012)
with treatment status				
d) Controlling for region-specific	0.0020*	0.0015	0.0010	-0.0002
linear time trends	(0.0012)	(0.0017)	(0.0013)	(0.0014)
e) Excluding Helsinki	0.0003	-0.0020	0.0013	-0.0006
metropolitan area	(0.0013)	(0.0015)	(0.0013)	(0.0016)
f) Excluding if $t < -3$ or $t > 3$	0.0002	-0.0005	0.0005	-0.0013
	(0.0011)	(0.0015)	(0.0012)	(0.0014)
g) Extending birth cohorts to 1960–66,	0.0001	-0.0001	0.0003	-0.0010
excluding if $t < -3$ or $t > 3$	(0.0011)	(0.0013)	(0.0012)	(0.0013)

Note: Baseline results are those reported in Table 3. All specifications include cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. t is timing of the reform in the number of years relative to the reform year in the municipality of residence (see Table A4). * = significant at 10%; ** = significant at 5% (all two-sided tests).

Table A6: Robustness Checks of the Estimated Reform Effect on Suicides

	Full sample	Parental education		
		Low	Mid	High
		Males		
Baseline results	-0.0002	-0.0016	0.0013	0.0003
	(0.0008)	(0.0011)	(0.0010)	(0.0010)
a) Controlling for prior mental	-0.0002	-0.0016	0.0013	0.0003
health disorders	(0.0008)	(0.0011)	(0.0010)	(0.0010)
b) Controlling for prior mental	-0.0002	-0.0016	0.0013	0.0003
health and parents' mental health	(0.0008)	(0.0011)	(0.0010)	(0.0010)
c) Controlling for prior mental	-0.0002	-0.0016	0.0013	0.0003
health disorders and its interaction	(0.0008)	(0.0011)	(0.0010)	(0.0010)
with treatment status				
d) Controlling for region-specific	-0.0010	-0.0008	0.0008	-0.0002
linear time trends	(0.0009)	(0.0013)	(0.0012)	(0.0012)
e) Excluding Helsinki	-0.0008	-0.0023*	0.0019*	-0.0011
metropolitan area	(0.0010)	(0.0013)	(0.0011)	(0.0012)
f) Excluding if $t < -3$ or $t > 3$	-0.0003	-0.0011	0.0019*	0.0001
	(0.0009)	(0.0012)	(0.0011)	(0.0010)
g) Extending birth cohorts to 1960–66,	0.0002	-0.0010	0.0020**	0.0005
excluding if $t < -3$ or $t > 3$	(0.0009)	(0.0011)	(0.0010)	(0.0010)
		Females		
Baseline results	-0.0003	-0.0006	-0.0001	0.0002
	(0.0005)	(0.0005)	(0.0004)	(0.0007)
a) Controlling for prior mental	-0.0003	-0.0006	-0.0001	0.0002
health disorders	(0.0005)	(0.0005)	(0.0004)	(0.0007)
b) Controlling for prior mental	-0.0003	-0.0006	-0.0001	0.0002
health and parents' mental health	(0.0005)	(0.0005)	(0.0004)	(0.0007)
c) Controlling for prior mental	-0.0003	-0.0006	-0.0001	0.0002
health disorders and its interaction	(0.0005)	(0.0005)	(0.0004)	(0.0007)
with treatment status				
d) Controlling for region-specific	-0.0006	-0.0006	-0.0003	0.0002
linear time trends	(0.0006)	(0.0006)	(0.0005)	(0.0009)
e) Excluding Helsinki	-0.0006	-0.0010	-0.0002	0.0007
metropolitan area	(0.0005)	(0.0006)	(0.0005)	(0.0009)
f) Excluding if $t < -3$ or $t > 3$	-0.0010*	-0.0012**	-0.0002	-0.0001
	(0.0005)	(0.0006)	(0.0005)	(0.0009)
g) Extending birth cohorts to 1960–66,	-0.0007	-0.0010**	-0.0002	0.0004
excluding if $t < -3$ or $t > 3$	(0.0005)	(0.0005)	(0.0005)	(0.0008)

Note: Baseline results are those reported in Table 4. All specifications include cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. t is the number of years relative to the reform year in the municipality of residence (see Table A4). * = significant at 10%; ** = significant at 5% (all two-sided tests).

Table A7: Robustness Checks of the Estimated Reform Effect on Mental Health Related Hospitalizations at Age 16–45

		Parental education		
	Full sample	Low	Mid	High
		Males		
Baseline results	0.0027	0.0074**	0.0012	-0.0058
	(0.0025)	(0.0036)	(0.0045)	(0.0043)
a) Controlling for prior mental	0.0028	0.0075**	0.0014	-0.0057
health disorders	(0.0025)	(0.0036)	(0.0045)	(0.0043)
b) Controlling for prior mental	0.0029	0.0078**	0.0012	-0.0056
health and parents' mental health	(0.0025)	(0.0036)	(0.0046)	(0.0043)
c) Controlling for prior mental	0.0029	0.0076**	0.0018	-0.0057
health disorders and its interaction	(0.0025)	(0.0036)	(0.0046)	(0.0042)
with treatment status				
d) Controlling for region-specific	-0.0012	0.0030	-0.0011	-0.0043
linear time trends	(0.0030)	(0.0045)	(0.0053)	(0.0048)
e) Excluding Helsinki	0.0032	0.0051	0.0049	-0.0078
metropolitan area	(0.0026)	(0.0038)	(0.0051)	(0.0057)
f) Excluding if $t < -3$ or $t > 3$	0.0024	0.0068*	0.0022	-0.0037
	(0.0028)	(0.0041)	(0.0050)	(0.0045)
g) Extending birth cohorts to 1960–66,	0.0018	0.0039	0.0046	-0.0017
excluding if $t < -3$ or $t > 3$	(0.0024)	(0.0036)	(0.0046)	(0.0043)
		Females		
Baseline results	0.0029	0.0022	-0.0007	0.0095**
	(0.0019)	(0.0025)	(0.0034)	(0.0041)
a) Controlling for prior mental	0.0030	0.0023	-0.0007	0.0097**
health disorders	(0.0019)	(0.0025)	(0.0034)	(0.0041)
b) Controlling for prior mental	0.0029	0.0022	-0.0007	0.0097**
health and parents' mental health	(0.0019)	(0.0025)	(0.0034)	(0.0041)
c) Controlling for prior mental	0.0030	0.0022	-0.0009	0.0100**
health disorders and its interaction	(0.0018)	(0.0025)	(0.0034)	(0.0041)
with treatment status				
d) Controlling for region-specific	0.0023	0.0002	-0.0006	0.0114**
linear time trends	(0.0024)	(0.0031)	(0.0042)	(0.0047)
e) Excluding Helsinki	0.0027	0.0027	-0.0018	0.0099*
metropolitan area	(0.0021)	(0.0028)	(0.0040)	(0.0054)
f) Excluding if $t < -3$ or $t > 3$	0.0039*	0.0028	0.0014	0.0120***
	(0.0021)	(0.0029)	(0.0038)	(0.0042)
g) Extending birth cohorts to 1960–66,	0.0039**	0.0035	0.0013	0.0092**
excluding if $t < -3$ or $t > 3$	(0.0019)	(0.0026)	(0.0037)	(0.0039)

Note: Baseline results are those reported in Table 5. All specifications include cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. t is the number of years relative to the reform year in the municipality of residence (see Table A4). * = significant at 10%; ** = significant at 5%; *** = significant at 1% (all two-sided tests).

Table A8: Significance Levels for the Main Results after Adjusting for Multiple Hypotheses Testing

Outcome	Sample	Reform effect	Conventional p-value	Romano-Wolf p-value
Panel A: Main outcomes				
a) Mental health related hospitalizations at age 16–45	Males, low-educated parents	0.0074	0.039	0.146
b) Mental health related hospitalizations at age 16–45	Females, highly-educated parents	0.0095	0.020	0.083
Panel B: Timing of hospitalization				
a) Age 26–35	Males, low-educated parents	0.0062	0.005	0.022
b) Age 36–45	Males, low-educated parents	0.0046	0.052	0.183
c) Age 11–15	Females, mid-educated parents	0.0024	0.014	0.048
d) Age 36–45	Females, highly-educated parents	0.0058	0.039	0.144
Panel C: Type of mental disorder				
a) Substance-use disorder	Males, full sample	0.0038	0.017	0.062
b) Substance-use disorder	Males, low-educated parents	0.0048	0.038	0.106
c) Bipolar disorder	Females, highly-educated parents	0.0026	0.058	0.205
d) Depressive disorder	Females, highly-educated parents	0.0062	0.005	0.022

Note: We only report estimated reform effects that are significant in Tables 5–7. Conventional and Romano-Wolf step-down adjusted p-values are based on standard errors that are clustered at the municipal level. The adjusted p-values are robust to multiple hypotheses testing (jointly for the full sample and three sub-samples by parental background). The adjusted p-values have been calculated using 2000 bootstrap replications.

Table A9: Estimated Reform Effect in the Combined Sample of Males and Females

Outcome	Full sample	Parental education		
		Low	Mid	High
Panel A: Main outcomes				
a) Death by age 45	-0.0005	-0.0019	0.0018*	-0.0009
	(0.0010)	(0.0013)	(0.0010)	(0.0011)
b) Suicide by age 45	-0.0002	-0.0011*	0.0007	0.0004
	(0.0005)	(0.0006)	(0.0005)	(0.0006)
c) Mental health related	0.0028*	0.0051**	0.0002	0.0023
hospitalizations at age 16–45	(0.0015)	(0.0022)	(0.0026)	(0.0030)
Panel B: Timing of hospitalization				
a) Age 6–10	-0.0006	-0.0005	-0.0007	-0.0013
	(0.0004)	(0.0007)	(0.0007)	(0.0008)
b) Age 11–15	0.0002	0.0001	0.0004	-0.0016***
	(0.0004)	(0.0006)	(0.0008)	(0.0006)
c) Age 16–25	0.0015	0.0025	-0.0013	0.0021
	(0.0011)	(0.0016)	(0.0020)	(0.0018)
d) Age 26–35	0.0021**	0.0038***	0.0018	-0.0012
	(0.0010)	(0.0013)	(0.0020)	(0.0023)
e) Age 36–45	0.0015	0.0033**	-0.0026	0.0016
	(0.0011)	(0.0015)	(0.0020)	(0.0021)
Panel C: Type of mental disorder				
a) Schizophrenia	0.0004	0.0015*	-0.0006	0.0002
	(0.0006)	(0.0008)	(0.0011)	(0.0014)
b) Other non-affective	0.0001	0.0014	-0.0011	-0.0010
	(0.0007)	(0.0010)	(0.0014)	(0.0014)
c) Bipolar disorder	0.0006*	0.0006	0.0005	0.0007
	(0.0004)	(0.0005)	(0.0007)	(0.0010)
d) Depressive disorder	0.0014**	0.0014	0.0007	0.0013
	(0.0007)	(0.0010)	(0.0014)	(0.0014)
e) Anxiety, stress, neurotic	0.0006	0.0015	-0.0005	-0.0010
	(0.0008)	(0.0010)	(0.0012)	(0.0011)
f) Substance-use disorder	0.0025***	0.0036***	0.0002	0.0025*
	(0.0009)	(0.0013)	(0.0017)	(0.0015)
Observations	366,314	184,918	93,316	71,110

Note: Each cell reports the estimated effect of the comprehensive school reform from separate models. All models include gender dummy as well as cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. * = significant at 10%; ** = significant at 5%; *** = significant at 1% (all two-sided tests).

Table A10: Cox Proportional Hazards Model Results for All-cause Mortality

		Parental education		
	Full sample	Low	Mid	High
		Males		
Treatment effect	1.007	1.000	1.049	0.969
	[0.944, 1.070]	[0.909, 1.090]	[0.861, 1.237]	[0.751, 1.187]
Observations	186,777	94,037	47,224	36,282
		Females		
Treatment effect	1.019	0.932	1.138	1.078
	[0.917, 1.121]	[0.789, 1.076]	[0.808, 1.469]	[0.726, 1.430]
Observations	179,537	90,881	46,092	34,828

Note: Birth cohorts 1962–1966 are at risk from age 11 onwards until year 2016 (max of age 54). Annual death hazard ratios are reported together with 95% confidence intervals in square brackets. Hazard ratios greater than 1 indicate increased mortality. All specifications include birth cohort dummies and the estimates have been stratified by municipality (i.e. holding the baseline hazard constant within municipality). Standard errors are clustered at the municipal level.

Table A11: Estimated Reform Effect on the Number of Days Spent in Hospital for Mental Health Reasons Between Age 16 and 45.

Outcome	Full sample	Parental education		
		Low	Mid	High
		Males		
a) Hospitalization days	3.1446**	4.1429**	1.1525	0.5366
	(1.4899)	(2.0915)	(2.5885)	(3.2623)
Mean outcome (days)	15.02	15.47	12.48	13.71
Observations	186,777	94,037	47,224	36,282
b) ln(Hospitalization days)	0.0631	0.1248**	-0.0593	0.1577
conditional on being hospitalized	(0.0401)	(0.0626)	(0.1080)	(0.1460)
Mean outcome (days)	150.67	143.63	147.29	186.49
Observations	18,616	10,130	4,001	2,668
		Females		
a) Hospitalization days	1.3853	2.6531	-0.7588	-0.5380
	(1.1449)	(1.8333)	(1.8682)	(2.6052)
Mean outcome (days)	10.86	10.95	9.260	11.78
Observations	179,537	90,881	46,092	34,828
b) ln(Hospitalization days)	0.0057	0.1210	-0.0442	-0.2036
conditional on being hospitalized	(0.0683)	(0.0972)	(0.1427)	(0.1595)
Mean outcome (days)	192.59	184.60	186.01	226.15
Observations	10,125	5,389	2,295	1,814

Note: Each cell reports the estimated effect of the comprehensive school reform from separate models. All models include cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. * = significant at 10%; ** = significant at 5% (all two-sided tests).

Table A12: Effect of the Reform: Treatment Assignment According to the Municipality of Birth

Outcome	Full sample	Parental education		
		Low	Mid	High
		Males		
a) Death by age 45	-0.0010	-0.0023	0.0024	-0.0035*
	(0.0016)	(0.0021)	(0.0020)	(0.0019)
b) Suicide by age 45	-0.0003	-0.0018	0.0016*	-0.0013
	(0.0009)	(0.0011)	(0.0009)	(0.0008)
c) MHD by age 45	0.0019	0.0041	0.0048	-0.0072*
	(0.0028)	(0.0036)	(0.0054)	(0.0041)
		Females		
a) Death by age 45	0.0006	0.0002	0.0017*	-0.0007
	(0.0010)	(0.0012)	(0.0010)	(0.0012)
b) Suicide by age 45	-0.0002	0.0000	-0.0001	0.0004
	(0.0004)	(0.0005)	(0.0004)	(0.0005)
c) MHD by age 45	0.0005	0.0000	-0.0018	0.0064*
	(0.0017)	(0.0026)	(0.0033)	(0.0038)

Note: Here, individuals are classified into treatment based on their municipality of birth instead of municipality at age 11 (cf. results in Tables 3–5). MHD = Mental health disorders requiring hospitalization spell. The full sample also includes individuals for whom information about parents (such as parental education) is missing. All models include cohort and municipality of birth fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. * = significant at 10% (all two-sided tests).

Table A13: Significance Levels for the Mediator Results after Adjusting for Multiple Hypotheses Testing

	Mediation analysis	
Outcome / Sample	Education	Educ. & Income
Outcome	*Panel (A): Males, Low-educated parents*	
Death by age 45	-0.0040	-0.0042
	[adj. p = 0.241]	[adj. p = 0.144]
Suicide by age 45	-0.0018	-0.0017
	[adj. p = 0.337]	[adj. p = 0.285]
MHD at age 16–45	0.0061	0.0060
	[adj. p = 0.301]	[adj. p = 0.367]
MHD at age 26–35	0.0056	0.0049
	[adj. p = 0.045]	[adj. p = 0.118]
MHD at age 36–45	0.0038	0.0032
	[adj. p = 0.354]	[adj. p = 0.524]
Substance-use disorder	0.0041	0.0033
at age 16-45	[adj. p = 0.184]	[adj. p = 0.370]
Outcome	*Panel (B): Females, Highly-educated parents*	
Death by age 45	-0.0008	-0.0008
	[adj. p = 0.861]	[adj. p = 0.868]
Suicide by age 45	0.0003	0.0005
	[adj. p = 0.925]	[adj. p = 0.843]
MHD at age 16–45	0.0096	0.0097
	[adj. p = 0.092]	[adj. p = 0.086]
MHD at age 26–35	0.0026	0.0025
	[adj. p = 0.633]	[adj. p = 0.654]
MHD at age 36–45	0.0057	0.0061
	[adj. p = 0.147]	[adj. p = 0.113]
Depression at age 16–45	0.0063	0.0065
	[adj. p = 0.022]	[adj. p = 0.017]

Note: We present controlled direct effects based on Acharya et al. (2016), after controlling for education and/or income. The Romano-Wolf step-down adjusted p-values in square brackets are robust to multiple hypotheses testing (jointly for the full sample and three sub-samples by parental background). They have been computed using 2000 bootstrap replications. Significance levels are based on standard errors that are clustered at the municipal level. MHD = Mental health disorders requiring hospitalization spell. See also Table 8.

Table A14: Estimated Reform Effect on Educational Outcomes

Outcome	Full sample	Parental education		
		Low	Mid	High
		Males		
a) Years of schooling	-0.0489**	-0.0630**	-0.1384***	0.0229
	(0.0222)	(0.0266)	(0.0416)	(0.0562)
b) Has a high school degree	-0.0022	-0.0020	-0.0163**	-0.0020
	(0.0037)	(0.0050)	(0.0068)	(0.0082)
High school exam score (percentile)				
c) in native language	0.0007	-0.0047	-0.0036	0.0047
	(0.0059)	(0.0080)	(0.0093)	(0.0086)
d) in advanced math	-0.0173**	-0.0104	-0.0370**	-0.0147
	(0.0073)	(0.0128)	(0.0164)	(0.0094)
e) in basic math	-0.0103	0.0067	-0.0329**	0.0005
	(0.0070)	(0.0132)	(0.0144)	(0.0117)
Highest degree completed				
f) Vocational secondary	-0.0005	-0.0073	0.0143**	0.0040
education	(0.0042)	(0.0067)	(0.0066)	(0.0074)
g) Vocational college	0.0020	0.0041	0.0005	-0.0027
	(0.0027)	(0.0038)	(0.0057)	(0.0069)
h) University master	-0.0048*	-0.0039	-0.0151***	0.0015
	(0.0027)	(0.0028)	(0.0049)	(0.0086)
		Females		
a) Years of schooling	-0.0139	-0.0329	0.0157	0.0126
	(0.0207)	(0.0258)	(0.0443)	(0.0498)
b) Has a high school degree	-0.0082**	-0.0109**	0.0024	-0.0107
	(0.0041)	(0.0053)	(0.0093)	(0.0077)
High school exam score (percentile)				
c) in native language	-0.0083	-0.0084	-0.0190**	0.0019
	(0.0055)	(0.0060)	(0.0086)	(0.0078)
d) in advanced math	-0.0137*	-0.0206*	-0.0242*	-0.0063
	(0.0071)	(0.0124)	(0.0123)	(0.0111)
e) in basic math	-0.0122**	-0.0229***	-0.0241**	0.0159*
	(0.0061)	(0.0082)	(0.0105)	(0.0084)
Highest degree completed				
f) Vocational secondary	-0.0049	-0.0056	-0.0174**	0.0075
education	(0.0045)	(0.0059)	(0.0081)	(0.0074)
g) Vocational college	0.0021	0.0011	0.0110	-0.0052
	(0.0042)	(0.0053)	(0.0074)	(0.0095)
h) University master	-0.0025	-0.0070**	0.0032	0.0043
	(0.0032)	(0.0035)	(0.0054)	(0.0093)

Note: Each cell reports the estimated effect of the comprehensive school reform from separate models. High school exam score is available only for those who have graduated from a high school. All models include dummy variables for the year of reform and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. * = significant at 10%; ** = significant at 5%; *** = significant at 1% (all two-sided tests).

Table A15: Estimated Reform Effect on Economic Outcomes

Outcome	Full sample	Parental education		
		Low	Mid	High
		Males		
Log of average income				
a) at age 26–45	-0.0092*	-0.0187***	-0.0016	0.0022
	(0.0050)	(0.0065)	(0.0089)	(0.0113)
b) at age 26–35	-0.0022	-0.0144**	0.0096	0.0120
	(0.0048)	(0.0057)	(0.0090)	(0.0097)
c) at age 36–45	-0.0160***	-0.0234***	-0.0121	-0.0048
	(0.0060)	(0.0080)	(0.0107)	(0.0137)
Employment rate				
d) at age 26–45	-0.0029	-0.0057**	-0.0004	0.0044
	(0.0025)	(0.0031)	(0.0044)	(0.0046)
e) at age 26–35	-0.0017	-0.0052*	-0.0002	0.0067
	(0.0027)	(0.0031)	(0.0054)	(0.0049)
f) at age 36–45	-0.0047	-0.0077**	-0.0008	0.0026
	(0.0027)	(0.0036)	(0.0046)	(0.0050)
		Females		
Log of average income				
a) at age 26–45	-0.0083**	-0.0145***	-0.0038	0.0018
	(0.0039)	(0.0049)	(0.0078)	(0.0103)
b) at age 26–35	-0.0036	-0.0079	-0.0014	0.0025
	(0.0039)	(0.0049)	(0.0071)	(0.0097)
c) at age 36–45	-0.0140***	-0.0211***	-0.0102	-0.0006
	(0.0046)	(0.0064)	(0.0093)	(0.0116)
Employment rate				
d) at age 26–45	-0.0060**	-0.0066**	-0.0029	-0.0015
	(0.0025)	(0.0029)	(0.0046)	(0.0055)
e) at age 26–35	-0.0034	-0.0056*	-0.0000	0.0012
	(0.0025)	(0.0033)	(0.0049)	(0.0059)
f) at age 36–45	-0.0085***	-0.0078**	-0.0060	-0.0050
	(0.0031)	(0.0035)	(0.0055)	(0.0059)

Note: Each cell reports the estimated effect of the comprehensive school reform from separate models. All models include cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. * = significant at 10%; ** = significant at 5%; *** = significant at 1% (all two-sided tests).

C Appendix: Testing for Mechanisms

In Section 6, we examine the extent to which post-treatment schooling and income mediates the effect of the comprehensive school reform on mental health. Simply augmented regression model with post-treatment mediator variables can lead to biased estimates (see Acharya et al., 2016).

Acharya et al. (2016) apply a sequential procedure that consistently estimates the treatment effect while holding the values of potential mediators fixed. Adopting their approach, we estimate these average controlled direct effects as follows (cf. Table 8):

1. Estimate an augmented model: $y_{ijt} = \alpha + \eta_j + \tau_t + \beta \times REFORM_{jt} + \delta' X_i^{Post} + \epsilon_{ijt}$, where X_i^{Post} are additional post-treatment controls (i.e. the years of schooling and dummy for having a high school degree, and/or the log of taxable income at ages 26–35).

2. Create a demediated outcome variable: $\tilde{y}_{ijt} = y_{ijt} - \hat{\delta}' X_i^{Post}$.

3. Estimate a model for the demediated outcome: $\tilde{y}_{ijt} = \alpha + \eta_j + \tau_t + \kappa \times REFORM_{jt} + \epsilon_{ijt}$, where κ is the average controlled direct effect (ACDE) of the reform.

Because the final estimation step contains a generated dependent variable ($\tilde{y}_{ijt}$), the standard errors have been bootstrapped using 1,000 replications of the full process (1–3). Additionally, the bootstrap replications have been clustered at the municipal level.

If the estimated ACDE is significantly different from zero, we can conclude that the comprehensive school reform has influenced mental health related hospitalizations through other pathways than the education and income channel. By contrast, if the estimated ACDE is not significant, then the reform has not had an additional effect on mental health related hospitalizations once the proposed mechanisms have been accounted for. In other words, the reform effect would be exclusively driven by the mechanisms related to the changes in education and income.