School Tracking and Mental Health

Petri Böckerman Mika Haapanen Christopher Jepsen

Alexandra Roulet*

November 9, 2020

Abstract

To understand how the type of education affects long-term mental health, we examine the effects of a comprehensive school reform on mental health-related hospitalizations and deaths. The reform postponed the tracking of students into vocational and academic schools from age 11 to 16, thus affecting the set of peers and curriculum to which these students were exposed. The reform was implemented gradually across Finnish municipalities between 1972 and 1977. We use difference-in-differences variation and administrative data. Our overall results show no discernible effects on mental health-related hospitalizations or deaths, but heterogeneity analysis shows an adverse effect on hospitalizations due to depression for females from highly-educated families.

JEL Codes: I12, I26, I28

Keywords: Tracking age, comprehensive school, mental health, depression, hospitalization, mortality

*Böckerman: University of Jyväskylä, Labour Institute for Economic Research, and IZA Institute of Labor Economics. Haapanen: University of Jyväskylä. Jepsen: University College Dublin and IZA Institute of Labor Economics. Roulet: INSEAD.

1 Introduction

Education leads to monetary (Angrist and Krueger, 1991) and non-monetary (Oreopoulos and Salvanes, 2011) gains at the individual level. Potential positive effect on health is a crucial part of the non-monetary return provided by education. 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 still remains inconclusive (Galama et al., 2018).

We advance the understanding of the education–health relationship by studying the effect of a change at an age in which students are split between academic and vocational education on mental health in adulthood.[1] So far the literature has only focused on physical health. Therefore, in this study, we examine the effects of education on mental health. The lack of evidence on the effects on mental health outcomes is a salient gap, since 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 only the effects of one additional year of

[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 U.S. tracking usually refers to ability grouping within schools (Hall, 2012).

education on health, whereas other relevant aspects of education, such as how long students are exposed to a common curriculum before being split between academic and vocational tracks, might also have an effect on health, especially on mental health. 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 drivers of mental 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 have 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 motivation behind such reform policies was the belief that early tracking systems were unfair to pupils from disadvantaged backgrounds (Jones et al., 2014).

To identify the effect of school tracking on mental health-related hospitalizations and deaths, we use difference-in-differences variation triggered by the Finnish comprehensive school reform, which was implemented gradually across Finnish municipalities during 1972–1977.[3] The phase-in of the reform offers plausibly exogenous variation in the tracking age, and its occurrence as far back as the 1970s allows us to identify long-run health effects. Key to our identification strategy is the fact that the gradual roll-out 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 16 without affecting the length of

[2]Nevertheless, age of tracking differs significantly among 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).

compulsory education. Thus, the reform provides a unique opportunity to study the effect of increasing the age for school tracking, holding fixed the number of years of compulsory education.

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

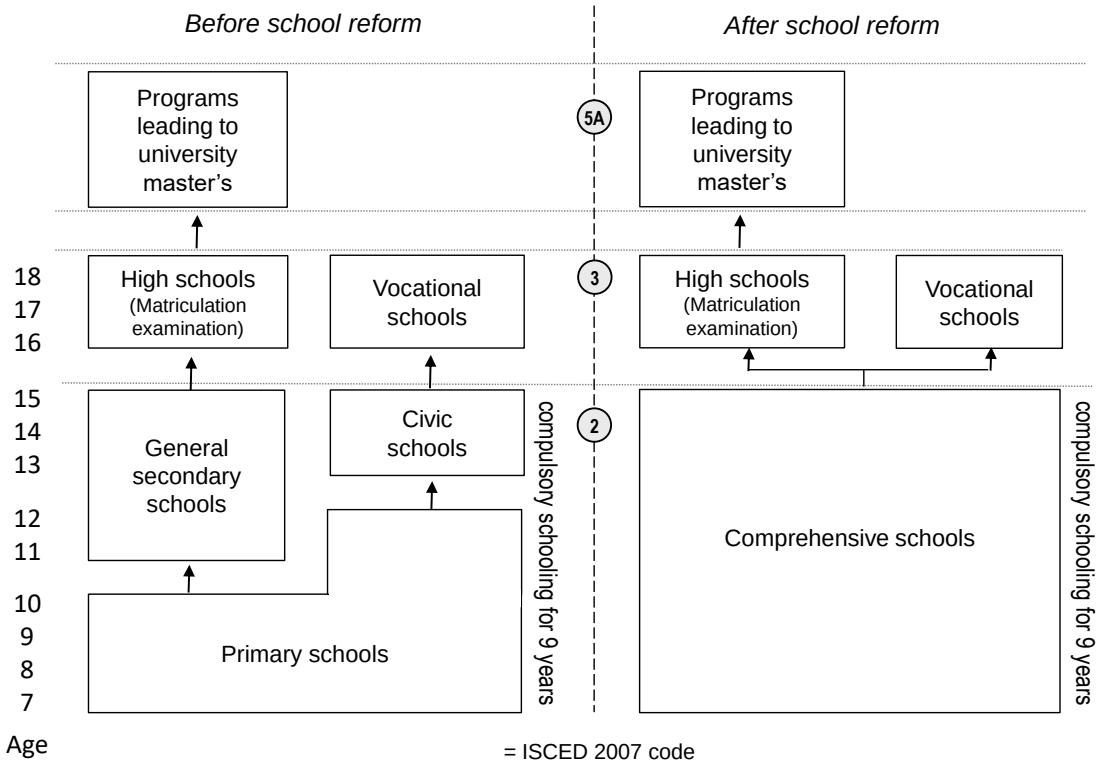
Overall, we do not find significant effects on mental health-related hospitalizations or deaths. This average null result is precisely estimated. We contribute to the debate on whether investing in people's education is an effective way to improve their health as well, which ultimately hinges on whether the correlation between education and health is causal (Jones et al., 2014). Our work brings further evidence to previous research finding negligible causal effects (Meghir et al., 2018). It matters because, as reviewed by Galama et al. (2018), the type of education received seems to affect health behaviors more than the length of education. Hence, our study, in theory, should be better able to detect effects on health than previous studies relying on compulsory schooling reforms that identify the effect of length of education on health. However, we still do not find an economically significant causal effect.

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 structures of both the systems (cf. Sahlberg, 2014).

Figure 1: The Structure of the Finnish Education System Before and After the



The reform postponed the tracking age from 11 to 16 years. Both before and after the reform school starts at age 7 and is compulsory till age 16. In the old system, pupils were taught together in the same class for only four years, from age 7 to 11. Then they 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 whom pupils were exposed between ages 11 and 16.[5] There were no systematic changes in classroom size, gender composition of classes or teacher quality due to the reform.

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 National Board of Education (NBE). Municipalities made suggestions regarding the timing of the reform, but it was the NBE that finally approved and ratified them. Municipalities were in charge of the practical implementation of the reform (in collaboration with the surrounding municipalities). There is spatial correlation in the timing of the reform, because the NBE wanted to make sure that within larger area students would get equal opportunities to move to secondary education (Somerkivi, 1982). Nevertheless, there is meaningful variation in exposure to the new comprehensive school system, both across birth cohorts and across municipalities (see Table 1 below).[6] This variation provides a quasi-experimental research setting.

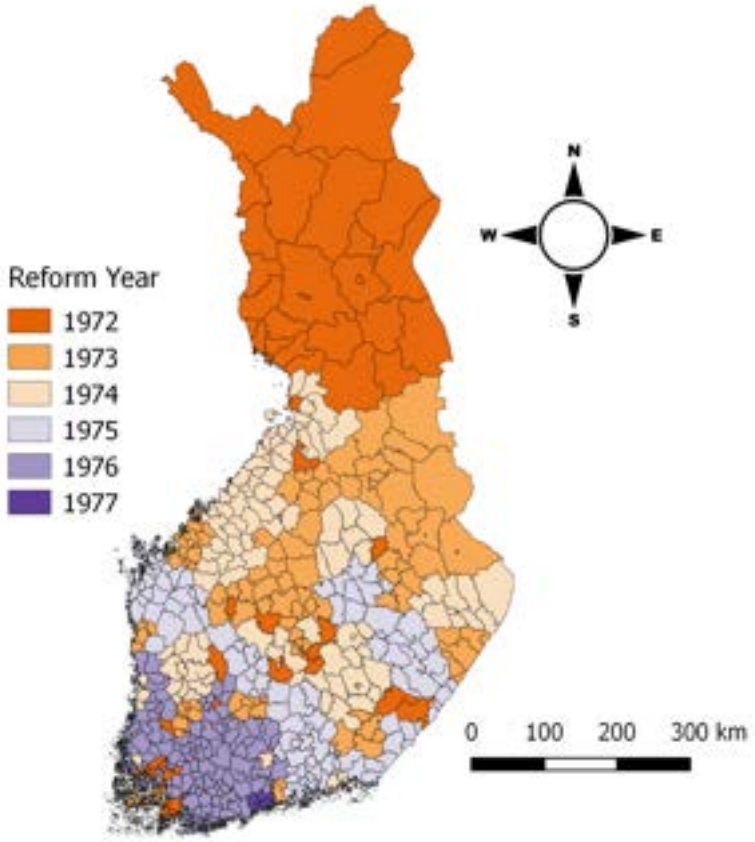
Previous research on the Finnish comprehensive school reform has shown that it

[4] Ability groups in foreign languages and mathematics existed in the comprehensive schools (Grades 7 to 9) until 1985 (Sahlberg, 2014, p. 28).

[5] Schooling from ages 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 curricula in both systems were similar for Grades 1 to 4 (Somerkivi, 1982, p. 28).

[6] Although the reform mostly proceeded from north to south, Figure A6 in Appendix B documents that in the adoption of the reform, substantial within-municipality variation exists among the four major socioeconomic regions.

Figure 2: Adoption of the School Reform during the Period 1972–1977



decreased the intergenerational correlation between earnings of fathers and sons (Pekkarinen et al., 2009). The reform also slightly improved the verbal and mathematical test scores of boys belonging to low socioeconomic backgrounds (Pekkala Kerr et al., 2013). In addition, the reform increased the gender difference in the probability of choosing an academic track and obtaining tertiary education (Pekkarinen, 2008). Finally, Ravesteijn et al. (2017) study the effect of the reform on all-cause mortality. For males, using an 11% random sample of the Finnish population, they find occasionally negative or positive effects, depending on the sub-sample. For females, they do not find any effect on mortality. Using total population data, we focus on the effects of the reform 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. Thus, the reform did not change the minimum school leaving age of 16. This allows us to focus on examining the effects of change in the age of tracking on mental health outcomes while keeping the the number of years of compulsory education fixed.

School tracking reform may positively affect education and income in adulthood. In turn, better human capital and availability of financial resources may improve health (Galama et al., 2018). For this reason, we analyze the potential mediating role of education and income on mental health in Section 6.

Besides education and income effects, other potential mechanisms are also at play. 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 especially on mental health, are ambiguous. After the reform, students aged 11 to 16 now 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 being 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 in turn may improve their health outcomes in the long run. It is also possible that low-achieving students will start adopting and imitating health behaviors of their higher-achieving peers. Conversely, higher-achieving students may be adversely affected by being exposed to lower-achieving peers.[7]

The comprehensive school reform also affected the ability ranking in the classroom. Vocational-track students who scored high in the ability ranking in the pre-reform system are now, on average, ranked lower in comprehensive school classrooms. Cicala et al. (2018) have shown that students' academic achievements and disruptive behaviors depend on their ordinal ranking among their peers. Moreover, Elsner and Isphording (2018) provide evidence that a student's ordinal ability ranking 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 without the reform could have lower health, especially poor mental health, after the reform. It contradicts the positive effect of being exposed to higher-achieving peers.

Moreover, the fact that students in the post-reform system follow a common curriculum from ages 11 to 16, rather than a track-specific curriculum, also means that, after the reform, students in a given classroom are less homogeneous. This makes it more difficult for teachers to tailor their pedagogical approaches to a more heterogeneous mix of students, thereby resulting in negative education and health outcomes (Betts, 2011). Students' achievement may be better when they are

[7]The changes in peer composition could also affect individuals' pool of potential partners and thus marriage outcomes, owing to assortative matching. We leave this issue for future studies.

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 reasons to favor the delay in tracking is that the likelihood of a student being placed in the "wrong" track is reduced and the anxiety associated with tracking lessens, since the amount of pre-tracking information about students' abilities is higher at the time of the tracking decision (Brunello et al., 2007). This suggests that the reform improves mental health. 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 research does not clearly predict the reform's effect on mental health, since different mechanisms are pushing in different directions. However, it does suggest that the effect likely differs based on gender and 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 of permanent residents of 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 since 1971. The data contain almost complete household and parental links. The date and municipality of birth are also recorded. Furthermore, the data also contain detailed information about degrees completed from 1970 onwards.

The core data include the individuals 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.[8] This sample restriction also allows us to follow all birth cohorts up to the maximum age of 45.[9] 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 implemented in their region of residence. For instance, people who turned 11 in the region in which the reform was implemented in 1975 received post-reform schooling if they were born in 1964, 1965 or 1966 and pre-reform schooling if they were born in 1962 or 1963.

We exclude foreign-born individuals (most of whom immigrated to Finland after the reform) and those living in the Åland Islands (in total, 8% of the original sample) from the estimation sample, retaining individuals living in 465 different municipalities. We also exclude a small number of individuals who at ages 11–15,

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

[9] We followed individuals till age of 45 or their death, whichever came before, resulting in unbalanced panel data. Since mortality is one of the outcomes of interest to us, we did not condition on being in the panel until age 45.

Table 1: Number of Observations and Adoption of the Reform

Birth cohort	Adoption 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 adopted in the years corresponding 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.

migrated between municipalities with a different year of adoption of the reform (less than 2.6% of the original sample)[10] since the reform indicator cannot be assigned unambiguously for these people. 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.[11] The variable is categorized into three values: i) low educated (53%) if neither of the parents completed post-compulsory education, i.e., if they both had a maximum of nine years of schooling; ii) mid educated (27%) if either or both of the parents completed a vocational degree but none studied further; and iii) 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).

[10] In the original sample, between ages 11–16, 5.3% of individuals migrated between municipalities. Less than 2.6% migrated between six waves (years) of adoption (cf. Figure 2). Only these latter individuals (47.3% of the migrants) were excluded from the estimation sample. We have investigated the potential role of endogenous selection of municipality of residence by assigning the individuals to treatment based on their municipality of birth and date of birth, as in Meghir et al. (2018); see Table A18 and discussion in Section 5.3.

[11] The correlation between parents' education and children's academic ability is strong when the ability is measured by high school completion (see Table A1 in Appendix B). Also note 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.

3.2 Mortality and Mental Health Disorders

To relate our findings to previous research on the education-health nexus (Lleras-Muney, 2005; Clark and Royer, 2013; Meghir et al., 2018), we examine mortality outcomes. We use data regarding the year and cause 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). Our mortality outcomes (suicides, mental health-related deaths, and all-cause mortality) are measured until age of 45. We picked age 45 so that all birth cohorts could be followed over the same window.

We use suicides as the primary mortality variable (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, 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). Second, approximately 25% of all deaths by age 45 are suicides. Third, the suicide mortality of young Finns is among the highest in the world (Lahti et al., 2011).

Following Alexander and Schnell (2019), we also consider a broader measure of mental health-related deaths, which includes not only suicides but also injuries of undetermined intent (i.e., fatal injuries about which it is not known whether they occurred accidentally or were purposely inflicted), and accidental deaths involving poisonings, drownings, and deaths involving firearms and trains. In our data, 69.3% of mental health-related deaths by age 45 are suicides, 3.6% are injuries of undermined intent, 19.8% are accidental poisonings[12], 5.7% are accidental drownings, and 1.5% are accidents involving firearms or trains. ICD codes that

[12]For Finnish men, in particular, accidental alcohol poisonings are the most prevalent category in a broader group of accidental poisonings. Accidental drug overdoses are very rare in Finland for the cohorts that we study.

are used to identify these causes of death are given in Appendix A. We utilize this broader measure of mental health-related mortality because not necessarily all deaths caused by mental health disorders are classified as suicides. Finally, all-cause mortality by age 45 is measured by a dummy variable (Table 2). In our data, 4.6% of men and 1.8% of women die by age 45. Moreover, mental health-related death rates for men are higher (1.9%) than that for women (1.2%).

We then study the effects of the reform on mental health-related hospitalizations using inpatient data.[13] We focus on *serious* mental health-related hospitalizations 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, undoubtedly, 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 for 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 HDR data are of high quality from 1972 onwards (Sund, 2012).

In the baseline model, we use a dummy variable to indicate whether the individual had any mental health-related hospitalizations between ages 16 and 45.[14] Ac-

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

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

cording to the data, there are 138,800 mental health-related hospitalization spells between ages 16–45 among 28,700 individuals (representing 7.8% of the total population). Approximately, 10 percent of men and 6 percent of women had mental health problems that resulted in hospitalization (Table 2). Conditional on having a mental health-related hospitalization, the average time spent in the hospital between ages 16 and 45 is 151 days for men and 193 days for women.[15] We then also consider separately whether the individual had any spell starting at ages 11–15, 16–25, 26–35, or 36–45. Again, we record new hospitalization spells until age 45 so that all birth cohorts can be followed during the same age window.

To get a 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 disorder, an affective psychosis involving emotional and mood abnormalities (and manic episodes); iv) depressive disorder, which can include repeated episodes of severe depression or chronic mild-grade depression (dysthymia); v) severe anxiety, stress, and neurotic disorders, which can interfere with daily activities, such as job performance, school work, and social relationships; and vi) substance-use disorder, which 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, broken down by socioeconomic background, in addition to some summary statistics on additional outcomes.

[15]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				
Suicide by age 45	0.013	0.113	0.003	0.055
Mental health-related death by age 45	0.019	0.136	0.012	0.107
Death by age 45	0.046	0.209	0.018	0.131
Hospitalizations due to mental health disorder				
At ages 6–10	0.006	0.080	0.004	0.063
At ages 11–15	0.006	0.079	0.005	0.070
At ages 16–25	0.044	0.205	0.017	0.127
At ages 26–35	0.044	0.206	0.028	0.166
At ages 36–45	0.050	0.219	0.034	0.181
At ages 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:

$$y_{ijc} = \alpha + \eta_j + \tau_c + \beta \cdot REFORM_{jc} + \gamma' X_i + \epsilon_{ijc} \quad (1)$$

where, y_{ijc} is the health outcome of individual i, who was born in year c and schooled in municipality j when entering 5th grade. η_j and τ_c are the municipality and the birth cohort fixed effects.[16] There are permanent regional differences in the outcomes that we need to control for. Similarly, birth cohorts may have also been exposed to different shocks in childhood and adolescence that have an impact on mental health in adulthood. $REFORM_{jc}$ is a dummy that varies across municipalities and cohorts and equals one if individual i has been exposed to the reform, i.e., has 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.

Since we identify the estimates using a difference-in-difference framework, the timing of when municipalities adopt the reform needs to be unrelated to differences in 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 (cf. Bhuller et al., 2017). We also show that there is no evidence of a systematic relationship between the timing of the reform and baseline pre-treatment municipality characteristics that may

[16] We also estimate the models with six year-of-adoption dummies (cf. Figure 2) instead of the full set of 465 municipal dummies. 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) owing to data limitations. The results are robust to this (see Table A4, Panel A). Panel B provides estimation results based on 18 NUTS-3 regional dummies that also utilize within-region variation.

affect mental health disorders (Figure A2).

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

Following Meghir et al. (2018) and insights from Galama et al. (2018), we separately estimate the empirical specifications by gender, since there is substantial variation in all outcomes by gender. For example, suicide mortality is much higher among males. We also estimate the models by level of parental education (three mutually exclusive categories), since 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, which is the level of policy variation. We also report alternative significance levels of the main results using clustering at a larger NUTS-4 level (67 regions; Table A13). In addition, we 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.[17]

[17]Consistent with Attanasio et al. (2017), our baseline Romano and Wolf correction uses four groups: full sample and three sub-samples of low-, mid-, and highly-educated parents. We perform the correction using these four groups, separately for males and females. The medical literature shows that the likelihood of mental health disorders differs substantially by gender (Salk et al., 2017), and thus analyses are separately done for both genders. We also implement the procedure with three groups, i.e., low-, mid-, highly-educated parents, or six groups, i.e., (male, female) $\times$ (low-, mid-, highly-educated parents). Results of these alternative multiple hypotheses testing are reported in Table A12.

We provide robustness checks for our baseline results. First, we control for region-specific linear time trends, which makes identification of the effects 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, since some private schools were operating in the Helsinki region after the reform (Pekkarinen et al., 2009). 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 adoption of the reform (in their municipality of residence). This restriction was imposed so as to rely only on individuals who participated in compulsory schooling closest to the adoption of the reform (i.e., diagonal elements in Table 1); this increases the homogeneity of the treated and untreated birth cohorts (see also Table A6). Fifth, we augment the baseline model by expanding the sample to also include individuals born in 1960–61, as in Pekkarinen et al. (2009). In comparison, our preferred sample focuses on individuals born in 1962–66, the same years as in Pekkala Kerr et al. (2013). These robustness checks are reported in the Appendix for suicides (Table A7), mental health-related deaths (Table A8), all-cause mortality (Table A9), and mental health-related hospitalizations (Table A10), and brief comments regarding them are given in the Results section.[18]

[18]The results are also robust to using additional birth month dummies or birth month by birth year dummies (Table A5), as estimated by Lager et al. (2016).

5 Results

5.1 Graphical presentation of the main results

To check the validity of our research design, as well as to graphically preview the key findings, we estimated the baseline specification using lead and lag year dummies around the reform, omitting the year prior to the reform ($t = -1$), as in Pekkala Kerr et al. (2013):

$$y_{ijct} = \alpha + \eta_j + \tau_c + \sum_{k \neq -1} \beta_k \cdot \mathbb{1}[k = t] + \epsilon_{ijct} \tag{2}$$

where $\mathbb{1}$ is an indicator function and β_k are the parameters of the event time dummies. Time $t = 0$ represents the first birth cohort in the municipality affected by the reform (Table A6). This model allows us to separate the pre-existing trends from policy responses over time.

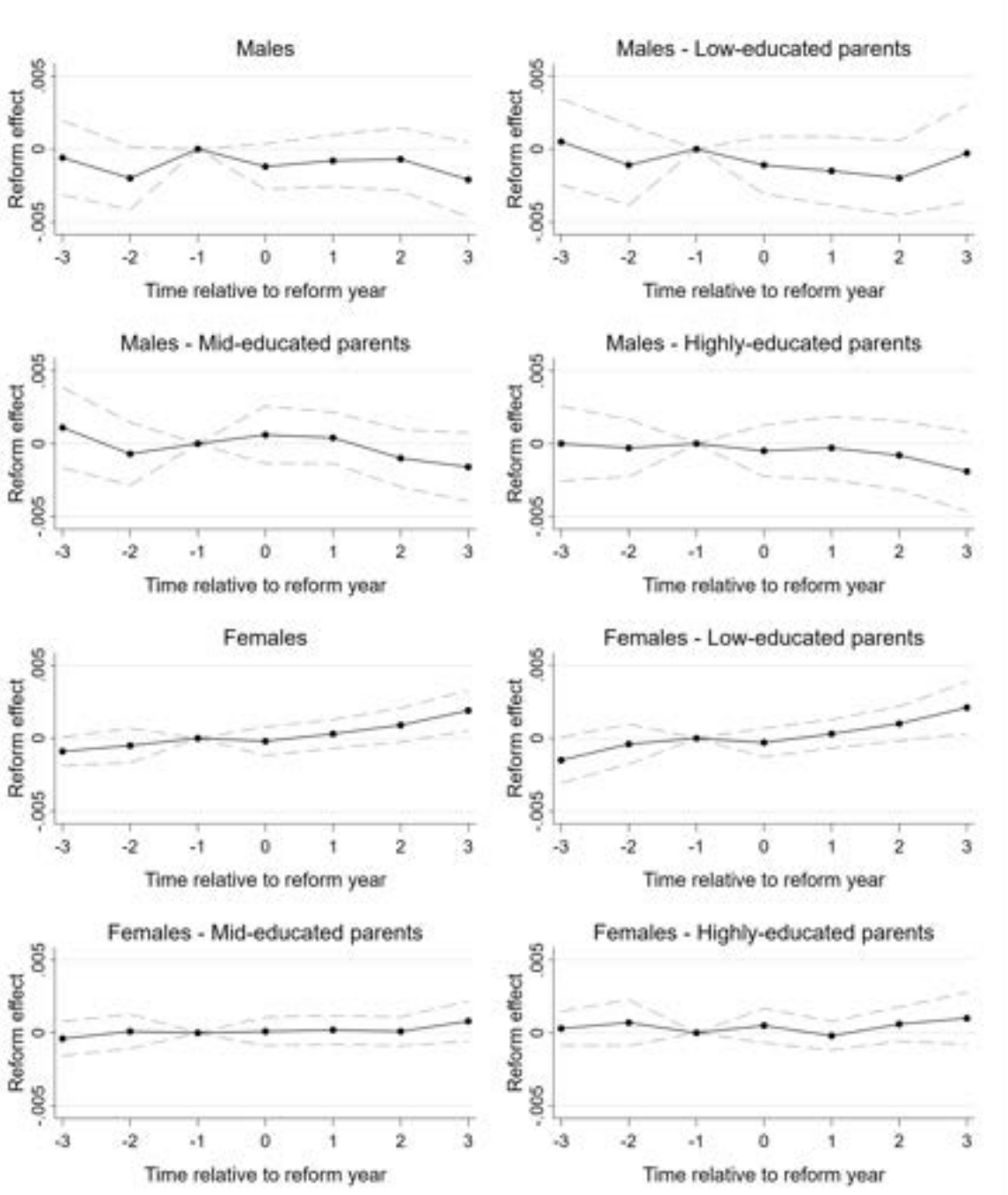
The results reported in Figures 3–4 and Figures A3–A4 of Appendix B suggest that our main findings are not affected by pre-reform trends, which provides evidence in favor of our identification assumption.[19] They also preview that our main findings are going to be mostly null effects. We now turn to regressions where we pool together all the post-reform (and pre-reform) years to precisely estimate the magnitude of these (null) effects.

5.2 Mortality

Table 3 reports the estimates for the effects of the reform on the incidence of suicide by age 45. Regardless of the sample, we find no evidence that the reform affected

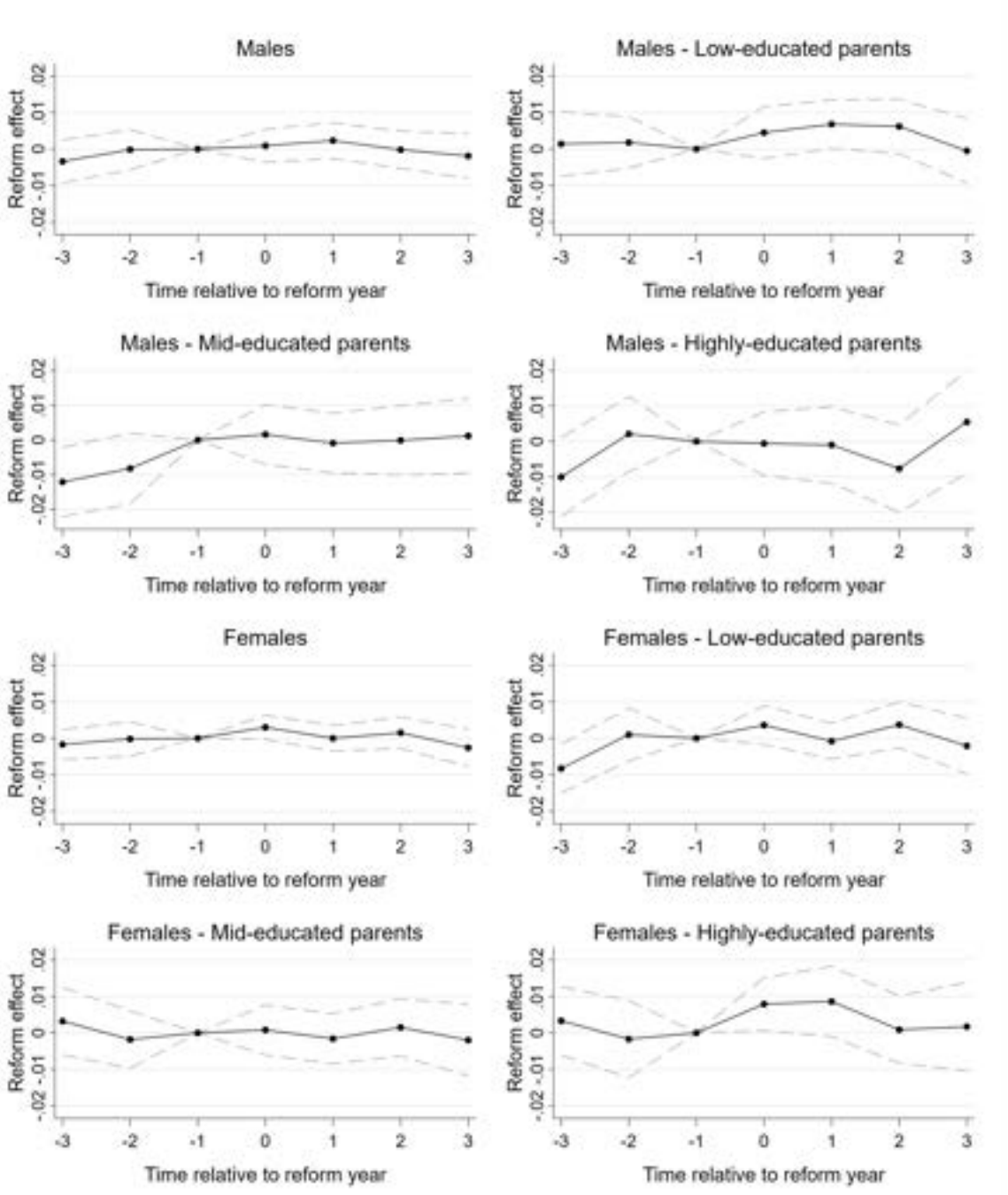
[19] Furthermore, Figure A5 shows that at ages 6–10 there is no evidence for significant effects on mental health-related hospitalizations. However, this test is relatively weak, since it is possible that the replacement of the old system (primary schools) with comprehensive school system also affected 7–10-year-old children even though the curriculum remained largely unchanged (Somerkivi, 1982, p. 28).

Figure 3: Effect of the Reform on Suicide at Ages 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 (based on robust standard errors clustered at the municipal level).

Figure 4: Effect of the Reform Effect on Mental Health-Related Hospitalizations at Ages 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 (based on robust standard errors clustered at the municipal level).

suicides by age 45. 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 A7 shows that the results for the full sample are robust to the variations of the model (e.g., adding controls and excluding some observations).[20]

Next, we consider the effects of the reform on mental health-related deaths that including suicides, injuries of undetermined intent, and accidental deaths. The results reported in Table A2 are similar to those found for suicides. The only notable exception is that for this broader measure of mental health-related deaths, improved survival in males from low-educated families can be seen. The estimate indicates reduction in mortality by 0.3 percentage points (significant at 5% level), relative to the mean outcome of 1.4%. However, after adjusting for multiple hypotheses testing, its significance level is marginally outside 10% level (Table A11). The results for males from low-educated families remain qualitatively intact to the alternative specifications of the model, but the significance levels vary by the model (Table A8). The estimated effects remain insignificant in other sub-samples.

To evaluate the overall health effects of the reform, Table A3 reports the findings on all-cause mortality. Echoing the results for suicides, we find no evidence that the reform affected all-cause mortality by age 45. The estimated effects are close to zero.[21] For males, we can rule out increases of 0.17 percentage points or decreases of 0.5 percentage points at the 5% significance 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%. Despite the universal health care system, the baseline differences in mortality by parental education are

[20]The aggregate results for both genders together are reported in Table A15. None of the estimated effects for suicides or all-cause mortality are significant at the standard 5% level. The reform effect on mental health-related deaths is negative ($p < 0.05$) for individuals from low-educated families.

[21]For females, the null effects are consistent with those of earlier Finnish study of Ravesteijn et al. (2017). Using an 11% random sample of the population, Ravesteijn et al. (2017) find longevity gains (losses) for males from low-income (high-income) families.

Table 3: Effect of the Reform on Suicide by Age 45

	Full sample	Parental education		
		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.

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. Robustness of these results is documented in Table A9.

The conclusions are also not sensitive to the use of Cox proportional hazards model (Table A16). These results provide opportunity to compare the precision of our estimates to prior evidence. Meghir et al. (2018) report the effects of the reform on hazard rates of death by age 45 in their Appendix Table A5 (sample of males and females). Their estimate from a Cox regression is 1.0006 (s.e. 0.0217). We report comparable estimate in the notes to Table A16, which indicates a combined reform effect of 0.985 (s.e. 0.030). Thus, Meghir et al. (2018) can rule out (with 95% confidence) changes in the risk of death by age 45 that are outside -4.2% and +4.3%, whereas we can rule out changes that are outside -7.3% and +4.6%. In both the studies, standard errors are clustered at the municipal level.

5.3 Hospitalization

Table 4 reports the results that use the incidence of mental health-related hospitalizations as an outcome variable 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 4: 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 A11).

The reform led to an increase in mental health-related hospitalizations, but for only some subgroups. The first finding is that males from low-educated families who were exposed to late tracking have a 0.74 percentage points higher probability of having severe mental health disorders that results in hospitalization than those who were 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 ages 16 and 45 (see Table A17). Importantly, the effect for males from low-educated families is no longer statistically significant after adjusting the standard errors for a Romano-Wolf (2005) correction for multiple hypotheses.

The second finding is that owing to the postponement of the tracking age from 11 to 16 for females from highly-educated families the probability of severe mental health disorders increased by 0.95 percentage points. We highlight 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 from highly-educated families (5.2%). Second, the effect remains significant (at 10% level) even after adjusting the standard errors for Romano-Wolf (2005) correction for testing multiple hypotheses (Table A11). 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 A10). Fourth, the finding is also robust to several other sensitivity checks, e.g., accounting for the full set of region-specific linear time trends (Table A10). Fifth, the estimated dynamic response models show that the finding is not driven by pre-reform trends (Figure 4). Therefore, it is unlikely that the result is affected by unobserved regional characteristics that are potentially correlated with the roll-out of the reform. Sixth, the finding remains intact even if individuals are classified into treatment based on their municipality of birth (Table A18) or municipality at age 10 (Table A19) instead of municipality at age 11 (and not eliminating individuals who migrate between ages 11–16).

The historical timing of the reform together with the availability of longitudinal data enable us to examine the effects of the reform over the life cycle. We report

the results for relevant age categories in Table 5. We find that at ages 36 to 45, for females from high-educated families, mental health disorders are significantly more likely to occur in those who had been exposed to the post-reform school system than 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 males from low-educated families, the positive effect of the comprehensive school reform on the probability of hospitalization for mental health reasons peaks at ages 26 to 35, but there is also a significant effect at later ages, i.e., between the ages of 36 and 45.

Table 6 examines whether any specific disorder drives the estimated treatment effects. We observe significant increase in the probability of depressive disorders for females from highly educated families. Conversely, for males 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 gender pattern that we observe is plausible, since in our data, alcohol-related mental disorders and depressive disorders are the most prevalent mental health problems for males and females, 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 without the reform, the key theoretical argument in favor of postponing the tracking is that it would have allowed the system to have more relevant and accurate information about abilities and comparative ad-

Table 5: Effect of the Reform on Mental Health-Related Hospitalizations over the Life Cycle

		Parental education		
	Full sample	Low	Mid	High
		Males		
a) Ages 11–15	-0.0002	0.0011	-0.0016	-0.0016
	(0.0008)	(0.0009)	(0.0012)	(0.0013)
b) Ages 16–25	0.0024	0.0041	-0.0007	0.0027
	(0.0021)	(0.0031)	(0.0040)	(0.0028)
c) Ages 26–35	0.0020	*0.0062***	0.0003	-0.0057
	(0.0016)	(0.0022)	(0.0031)	(0.0035)
d) Ages 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) Ages 11–15	-0.0002	-0.0009	*0.0024***	-0.0017
	(0.0006)	(0.0008)	(0.0010)	(0.0014)
b) Ages 16–25	0.0005	0.0007	-0.0012	0.0009
	(0.0011)	(0.0015)	(0.0021)	(0.0023)
c) Ages 26–35	0.0020	0.0011	0.0034	0.0026
	(0.0012)	(0.0016)	(0.0024)	(0.0029)
d) Ages 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 corresponds to a different sample. See Table A1 for the mean values of the outcome variables by parental education and Figure A5 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 A11).

Table 6: Effect of the Reform on Specific Mental Health-Related Hospitalizations at Ages 16–45

		Parental education		
	Full sample	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	-0.0005	0.0013	-0.0007	-0.0028
psychosis	(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	0.0010	0.0026	-0.0015	-0.0013
disorder	(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	0.0006	0.0015	-0.0016	-0.0000
psychosis	(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	0.0001	0.0003	0.0006	-0.0010
disorder	(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 A11).

vantages when matching students to a particular education (Brunello et al., 2007). Thus, increasing the tracking age should 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 find that the reform did not lead to an improvement in education or labor market outcomes (Tables A20 and A21). For students from low-educated families, we even find some negative effects of the reform on educational achievements 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 further pin down the potential mechanisms, 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 of females from highly-educated families remain intact even when controlling for the education and income mediators, as can be seen in Table 7. 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 ages 11 to 16 and exposed only to high-ability peers. However, after the reform, they are now exposed to comprehensive school and to peers who would have pursued the vocational track without the reform. Presumably, this new set of peers, on average, is often less well-behaved and of lower academic ability. This may result in a higher probability of being exposed to disruptive behaviors. Prior work has shown the negative association between being bullied at school and mental health in adulthood (Sigurdson et al., 2014). In particular, bullied girls

Table 7: 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*		
Suicide by age 45	-0.0016	-0.0018	-0.0017*
	(0.0011)	(0.0012)	(0.0010)
Mental health-related	-0.0030**	*-0.0033***	-0.0027**
death by age 45	(0.0014)	(0.0014)	(0.0013)
Death by age 45	-0.0034	-0.0040*	-0.0042**
	(0.0022)	(0.0022)	(0.0021)
MHD at ages 16–45	0.0074**	0.0061*	0.0060
	(0.0036)	(0.0036)	(0.0037)
MHD at ages 26–35	*0.0062****	*0.0056***	0.0049*
	(0.0022)	(0.0022)	(0.0022)
MHD at ages 36–45	0.0046*	0.0038	0.0032
	(0.0023)	(0.0024)	(0.0024)
Substance-use disorder	0.0048**	0.0041*	0.0033
at ages 16-45	(0.0023)	(0.0024)	(0.0023)
Outcome	*Panel (B): Females, Highly-educated parents*		
Suicide by age 45	0.0002	0.0003	0.0005
	(0.0007)	(0.0007)	(0.0007)
Mental health-related	0.0005	0.0005	0.0007
death by age 45	(0.0008)	(0.0009)	(0.0008)
Death by age 45	-0.0008	-0.0008	-0.0008
	(0.0012)	(0.0013)	(0.0012)
MHD at ages 16–45	*0.0095***	*0.0096***	*0.0097***
	(0.0041)	(0.0041)	(0.0042)
MHD at ages 26–35	0.0026	0.0026	0.0025
	(0.0029)	(0.0029)	(0.0029)
MHD at ages 36–45	0.0058**	0.0057**	0.0061**
	(0.0027)	(0.0028)	(0.0029)
Depression at ages 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 3–6 and A2–A3. See Table A23 for corresponding results for a full sample of males or females. 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 A22 for more details).

rather than bullied boys, even infrequently bullied ones, are more likely to suffer from depression symptoms in adulthood (Brunstein Klomek et al., 2007, p. 43). Further empirical research is needed to establish whether changes in the exposure to disruptive behaviors is driving our results.

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 backgrounds 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).

In summary, we do not find evidence supporting the role for the education/income channels (see also Table A23). Since the reform had a slightly negative effect on income in adulthood, this could push towards lower mental health and it could be that, once we control for income, the reform had a positive effect on mental health. However, since our earlier results remain intact after controlling for education/income, we interpret these results as further evidence that the reform, overall, had no effect on mental health rather than multiple effects of different signs cancelling each other out.

7 Conclusion

We contribute to the limited 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 during 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. The reform also changed the type of education received, providing five additional years of general education to students before the start of tracking. Consequently, our study also contributes to a growing literature illustrating the importance of human capital measures other than simply the quantity of education. In their review, Galama et al. (2018) highlight that for health-related behaviors, the type of education received – general, vocational, or academic – is more important than the length of education. 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).

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. Furthermore, we believe that our results can provide potentially valuable insights for other types of school tracking systems, such as those which identify ”gifted” or ”special education” students early during schooling and either provide additional opportunities or segregated instruction alongside similar students.

Overall, we find no significant effects on mental health-related hospitalizations or deaths, even though the average zero effects are precisely estimated. However, since we use register data on hospitalizations and deaths, our results do not rule out the possible effects of the reform on less serious mental health disorders. Heterogeneity analysis shows that postponing the age of tracking had an adverse long-run effect on mental health outcomes for females from highly-educated families, as they were more likely to be hospitalized for depression after the reform. Thus, increasing the age of tracking may come at a cost of negative mental health effects

for some groups. Furthermore, we find that the result for females is not accounted for by changes in education or income induced by the reform. Instead, we propose that peer effects can possibly explain our finding regarding the affected females. Exploring this mechanism is a promising avenue for future research.

Acknowledgments

We are thankful for the valuable comments from Jay Bhattacharya and Isaac Ehrlich (the editors), the three anonymous reviewers, Artturi Björk, Hannu Karhunen, Sari Kerr, Kristian Koerselman, Jani-Petri Laamanen, Krista Riukula, Kerstin Schneider, Tuomo Suhonen, Annual Conference of the Royal Economic Society (2019), EALE (2018), IIPF (2018), Irish Economic Association (2018), seminar participants at the Queen's University Belfast, ESRI, Labour Institute for Economic Research, University of Tampere, University of Jyväskylä, and University of Queensland. 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 Foundation (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.

Alexander, D. and M. Schnell (2019): "Just what the nurse practitioner ordered: Independent prescriptive authority and population mental health," *Journal of Health Economics*, 66, 145–162.

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

Attanasio, O., A. Guarn, C. Medina, and C. Meghir (2017): "Vocational training for disadvantaged youth in Colombia: A long-term follow-up," *American Economic Journal: Applied Economics*, 9, 131–143.

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.

Bhuller, M., M. Mogstad, and K. G. Salvanes (2017): "Life-cycle earnings, education premiums, and internal rates of return," *Journal of Labor Economics*, 35, 993–1030.

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*, 16, 983–1020.

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*, 53, 356–381.

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. van Kippersluis (2018): "The effect of education on health and mortality: A review of experimental and quasi-experimental evidence," *Oxford Research Encyclopedia of Economics and Finance*.

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. Kuop-

PASALMI, 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.

LLERAS-MUNEY, A. (2005): "The relationship between education and adult mortality in the United States," *The Review of Economic Studies*, 72, 189–221.

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, 159–184.

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.

Salk, R. H., J. S. Hyde, and L. Y. Abramson (2017): "Gender differences in depression in representative national samples: Meta-analyses of diagnoses and symptoms," *Psychological Bulletin*, 143, 783–822.

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.

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*, Cambridge, MA: MIT Press.

A Appendix: ICD-8, ICD-9, and ICD-10 Diagnosis Codes (Appendices are not for print publication)

A.1 Mental Health-Related Deaths

i) Suicides

ICD-8 and ICD-9: E950–E959;

ICD-10: X60–X84, Y87.0

ii) Injuries of undetermined intent (i.e., undetermined whether accidentally or purposely inflicted)

ICD-8 and ICD-9: E980–E989;

ICD-10: Y10–Y34, Y87.2, Y89.9

iii) Accidental deaths

a) Poisoning (i.e., accidental poisoning and exposure to noxious substances)

ICD-8: E850–E877; ICD-9: E850–E869;

ICD-10: X40–X49

b) Accidental drowning and submersion

ICD-8 and ICD-9: E910;

ICD-10: W65–W74

c) Firearms (i.e., accidental discharge of firearms)

ICD-8 and ICD-9: E922;

ICD-10: W32–W34

d) Trains (i.e., railway accidents and motor vehicle accident involving collision with train)

ICD-8 and ICD-9: E800–E807 & E810;

ICD-10: V05, V15, V80.6, V81.2–V81.9 & V25, V35, V45, V55, V65, V75, V81.0, V81.1, V87.6, V88.6.

A.2 Mental Health-Related Hospitalizations

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 disorder

ICD-8: 2961, 2963; ICD-9: 2962–2967;

ICD-10: F30–F31

iv) Depressive disorder

ICD-8: 2960, 2962; ICD-9: 2961, 3004A;

ICD-10: F32–F33, F341

v) Anxiety, stress, neurotic disorder

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

ICD-10: F40–F42, F430–F431

vi) Substance-use disorder

ICD-8: 291, 303–304; ICD-9: 291–292, 303–305;

ICD-10: F10–F19.

B Appendix: Tables and Figures

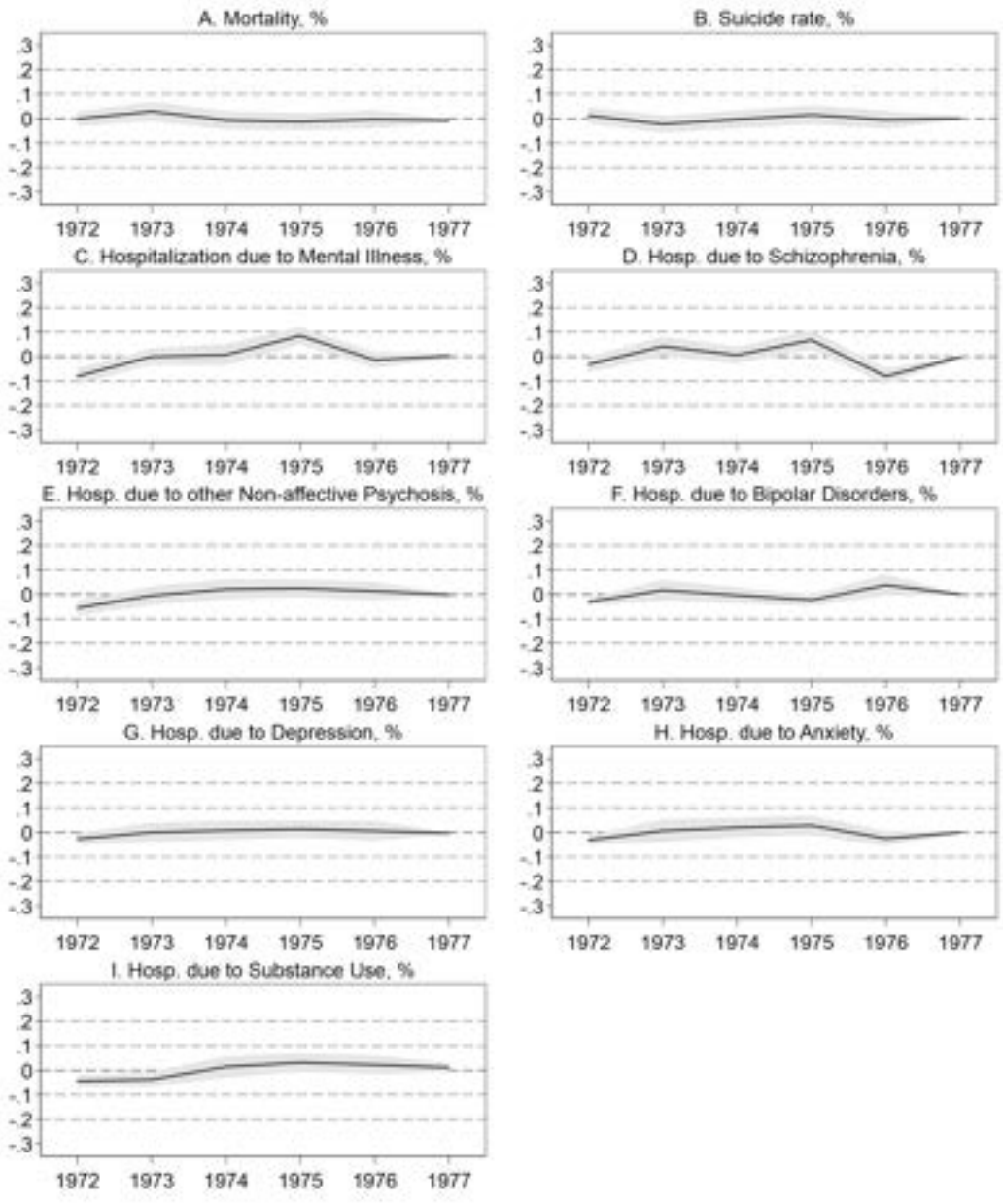
Table A1: Mean Values by Sex and Parental Background

Variable	Males			Females		
	Low-educated parents	Mid-educated parents	Highly-educated parents	Low-educated parents	Mid-educated parents	Highly-educated parents
Mortality						
Suicide by age 45	0.0099	0.0034	0.0023	0.0020	0.0007	0.0009
Mental health-related death by age 45	0.0143	0.0051	0.0036	0.0031	0.0012	0.0012
Death by age 45 (all-cause mortality)	0.0339	0.0132	0.0102	0.0122	0.0050	0.0047
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

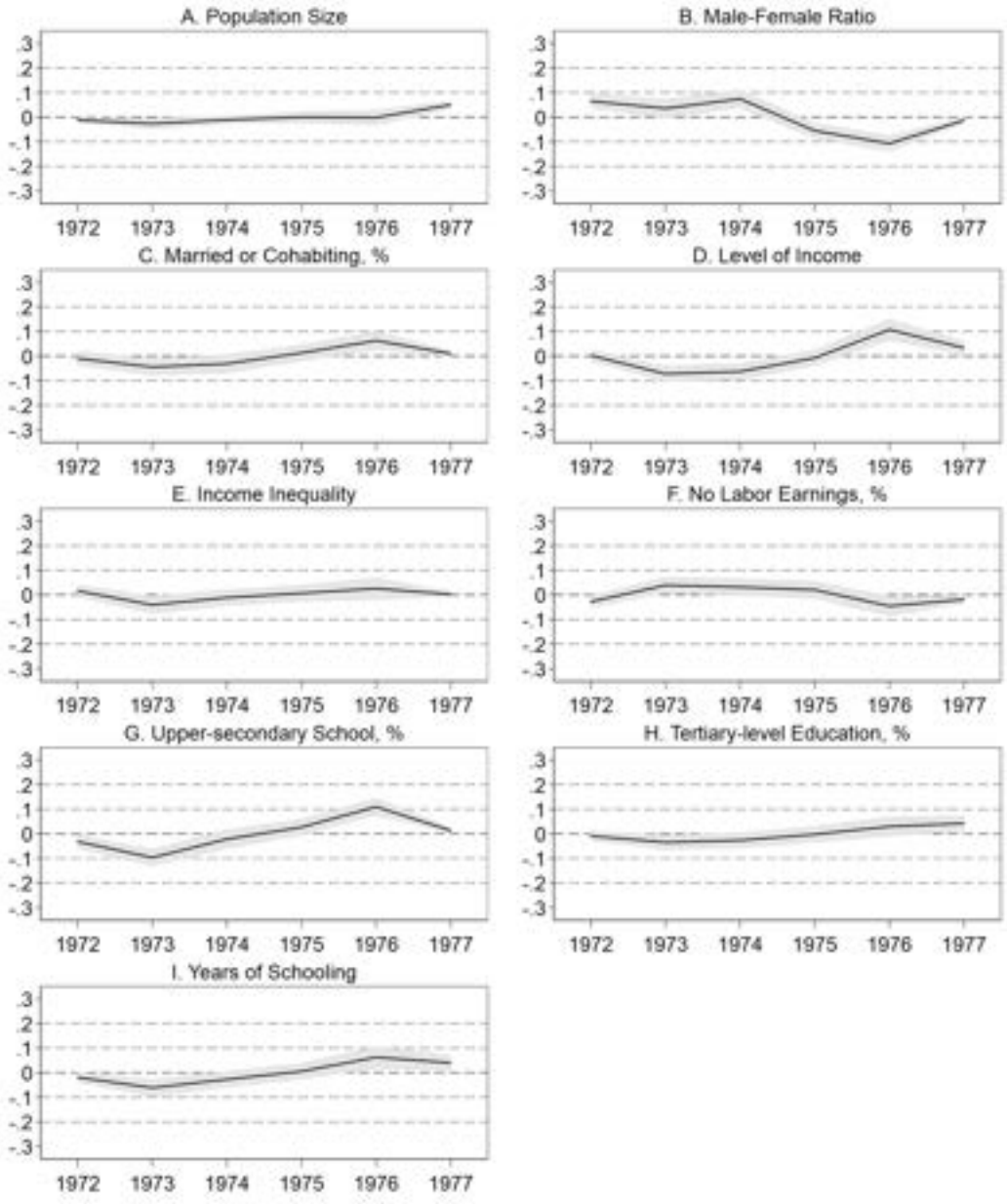
Variable	Males			Females		
	Low-educated parents	Mid-educated parents	Highly-educated parents	Low-educated parents	Mid-educated parents	Highly-educated parents
Educational outcomes						
Years of schooling	12.183	12.790	14.577	12.978	13.532	14.977
Has a high school (HS) degree	0.177	0.265	0.636	0.383	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



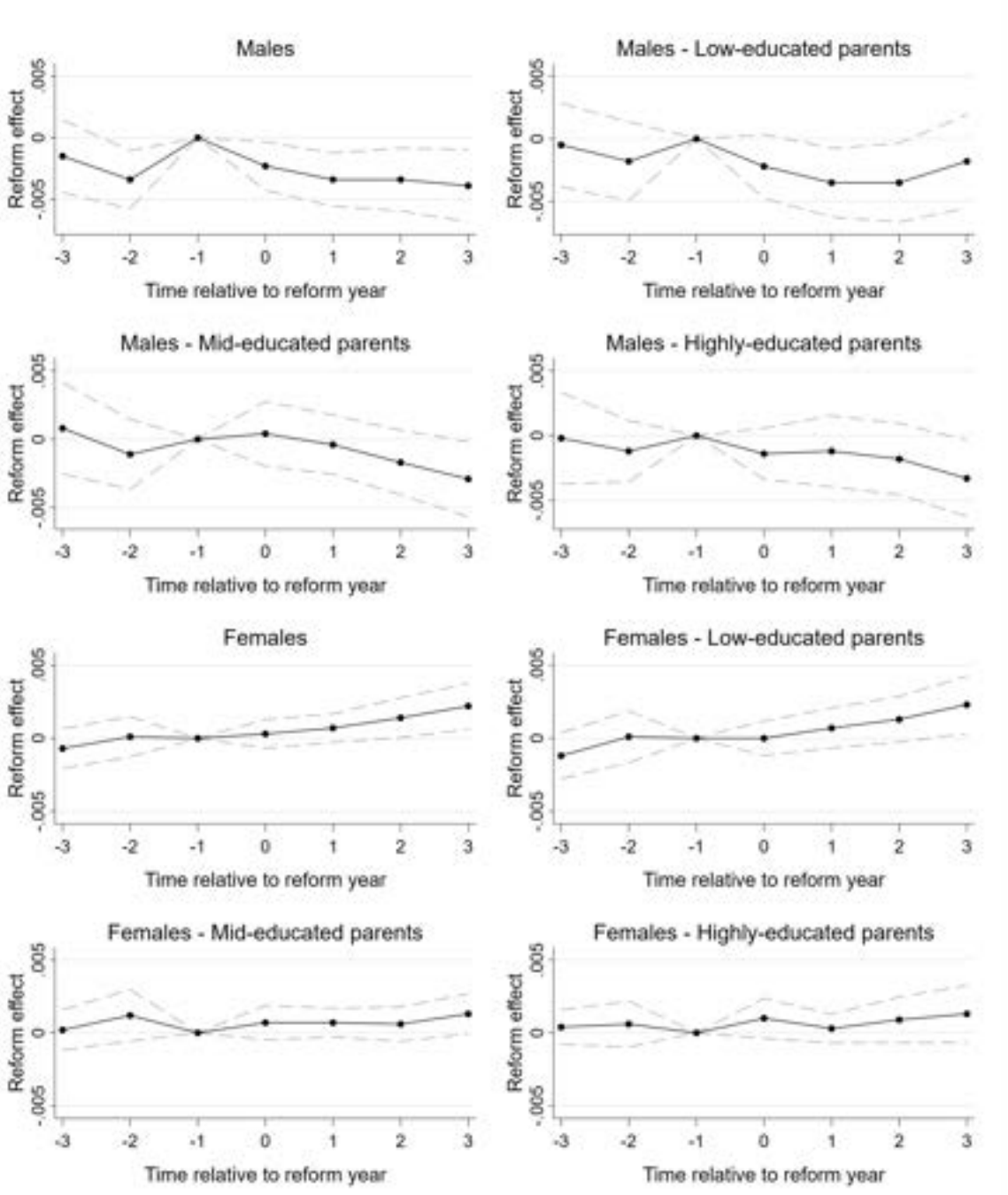
Note: We report results from separate regressions: $R_{jt} = (YEAR_t \times X_j)'\zeta_t + \tau_t + \epsilon_{jt}$, where the dependent variable R_{jt} is an indicator of the timing of the reform (1 if the reform was implemented in year t in municipality j). Explanatory variables contain year fixed effects 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



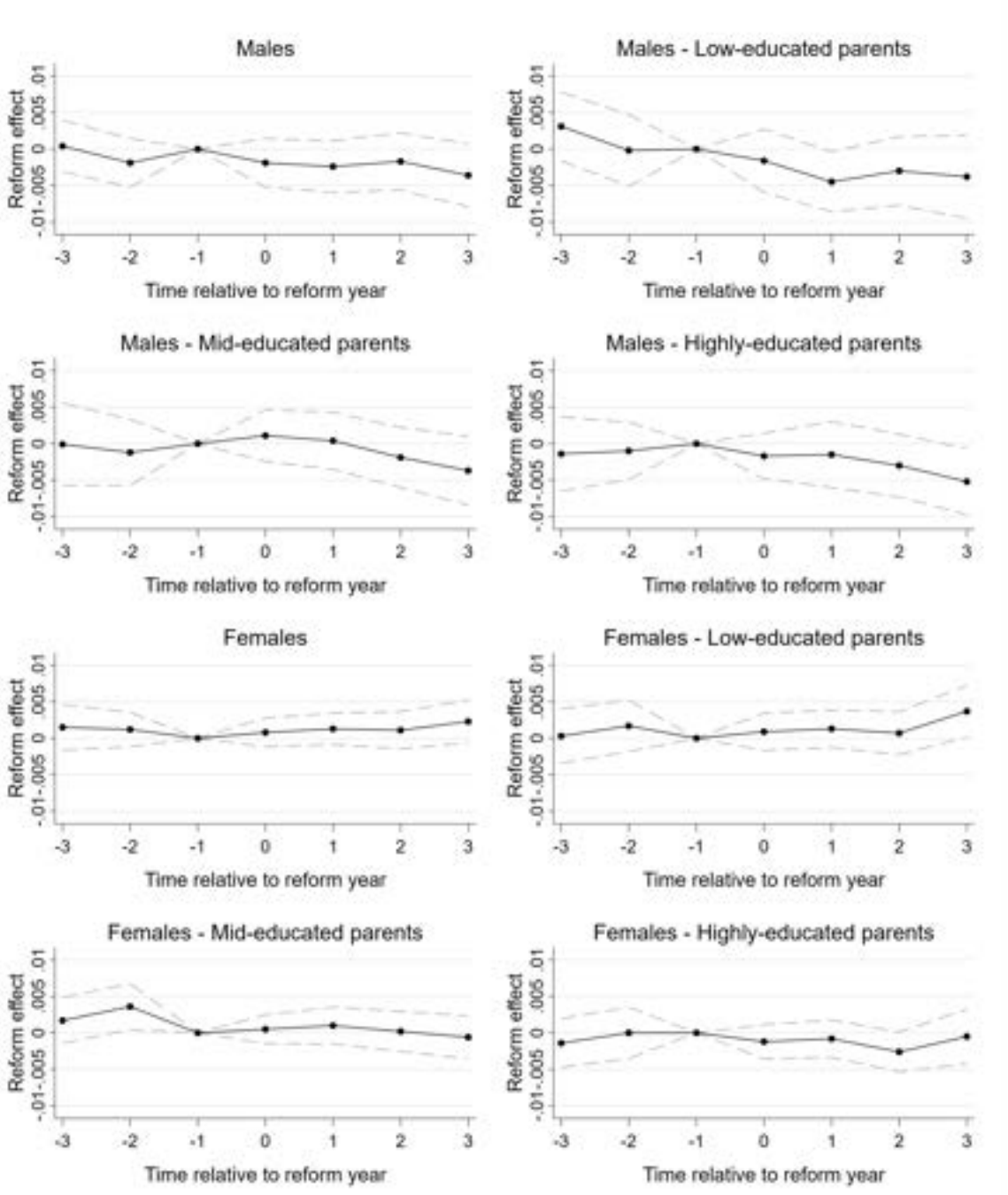
Note: See notes to Figure A1. The municipal characteristics are measured in 1971 or the nearest available year.

Figure A3: Effect of the Reform on Mental Health-Related Deaths 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 (based on robust standard errors clustered at the municipal level).

Figure A4: Effect of the Reform 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 (based on robust standard errors clustered at the municipal level).

Table A2: Effect of the Reform on Mental Health-Related Deaths by Age 45

	Full sample	Parental education		
		Low	Mid	High
		Males		
Treatment effect	-0.0014	-0.0030**, [a]	0.0010	-0.0001
	(0.0011)	(0.0014)	(0.0012)	(0.0012)
R squared	0.0033	0.0056	0.0086	0.0147
Mean outcome	0.0187	0.0143	0.0051	0.0036
Observations	186,777	94,037	47,224	36,282
		Females		
Treatment effect	0.0002	-0.0003	0.0004	0.0005
	(0.0005)	(0.0006)	(0.0005)	(0.0008)
R squared	0.0025	0.0050	0.0089	0.0129
Mean outcome	0.0043	0.0031	0.0012	0.0012
Observations	179,537	90,881	46,092	34,828

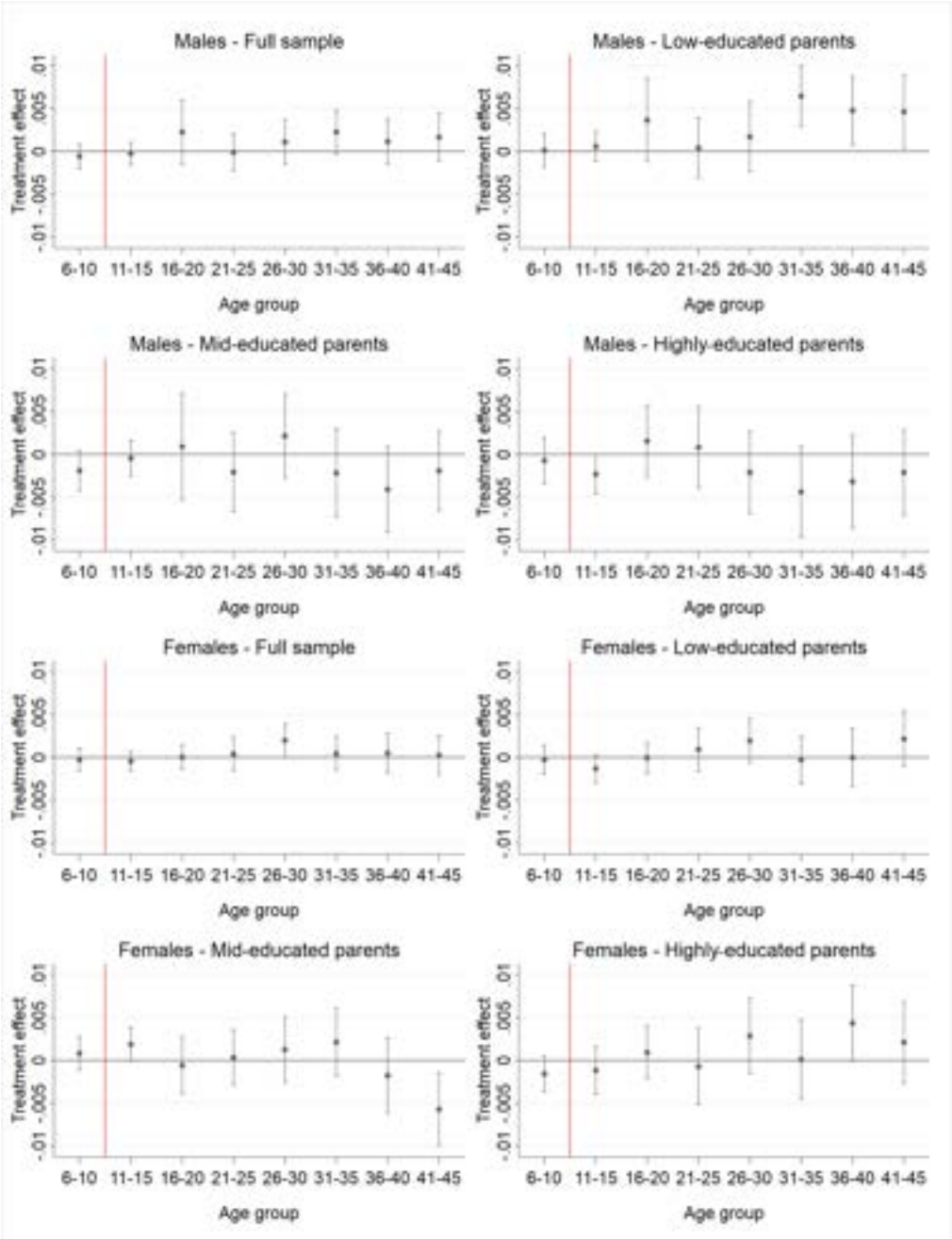
Note: The full sample also includes individuals for whom information about parents (such as parental education) is missing. All models include cohort and municipality fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. ** = significant at 5% level (two-sided test). [a] Significance level after Romano-Wolf (2005) correction for multiple hypotheses testing is 0.104 (see Table A11).

Table A3: Effect of the Reform on All-Cause Mortality by Age 45

	Full sample	Parental education		
		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.

Figure A5: 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 A4: Estimated Reform Effect After Controlling for Alternative Regional Effects

Outcome	Full sample	Parental education		
		Low	Mid	High
Panel A: Controls for the Six Years of Adoption				
		Males		
a) Suicide by age 45	-0.0001	-0.0015	0.0015	0.0005
	(0.0008)	(0.0011)	(0.0010)	(0.0009)
b) Mental health-related	-0.0013	-0.0027**	0.0011	0.0002
death by age 45	(0.0011)	(0.0013)	(0.0012)	(0.0011)
c) Death by age 45	-0.0012	-0.0028	0.0030*	-0.0011
	(0.0017)	(0.0022)	(0.0018)	(0.0017)
d) 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) Suicide by age 45	-0.0002	-0.0005	-0.0001	0.0004
	(0.0005)	(0.0005)	(0.0004)	(0.0007)
b) Mental health-related	0.0003	-0.0003	0.0003	0.0007
death by age 45	(0.0005)	(0.0006)	(0.0005)	(0.0008)
c) Death by age 45	0.0007	-0.0004	0.0008	-0.0007
	(0.0011)	(0.0013)	(0.0011)	(0.0012)
d) Mental health-related	0.0031*	0.0024	-0.0004	0.0101**
hospitalizations at age 16–45	(0.0019)	(0.0025)	(0.0034)	(0.0040)
Panel B: Controls for the 18 NUTS-3 Regions				
		Males		
a) Suicide by age 45	0.0001	-0.0014	0.0007	0.0011
	(0.0008)	(0.0009)	(0.0008)	(0.0009)
b) Mental health-related	-0.0009	-0.0026**	-0.0002	0.0009
death by age 45	(0.0010)	(0.0011)	(0.0010)	(0.0011)
c) Death by age 45	-0.0010	-0.0033*	0.0009	-0.0007
	(0.0015)	(0.0018)	(0.0016)	(0.0014)
d) Mental health-related	0.0003	0.0036	-0.0059	-0.0055
hospitalizations at age 16–45	(0.0028)	(0.0039)	(0.0041)	(0.0036)
		Females		
a) Suicide by age 45	-0.0003	-0.0006	-0.0002	0.0002
	(0.0004)	(0.0004)	(0.0003)	(0.0006)
b) Mental health-related	-0.0002	-0.0009*	-0.0000	0.0005
death by age 45	(0.0004)	(0.0005)	(0.0004)	(0.0007)
c) Death by age 45	0.0002	-0.0013	0.0002	-0.0007
	(0.0009)	(0.0011)	(0.0010)	(0.0011)
d) Mental health-related	-0.0003	-0.0014	-0.0017	0.0043
hospitalizations at age 16–45	(0.0016)	(0.0024)	(0.0030)	(0.0037)

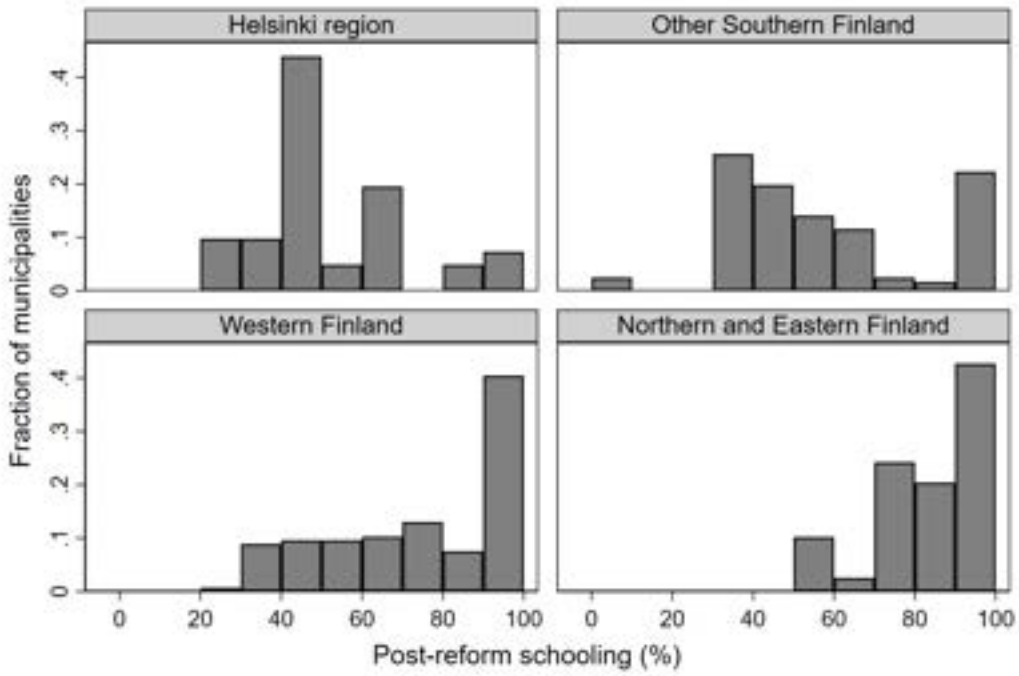
Note: Each cell reports the estimated reform effect from separate models. In Panel A, we have replaced the full municipal fixed effects with six dummies for the years of adoption (1972 to 1977). In Panel B, we have replaced the municipal fixed effects with 18 NUTS-3 regional dummies. All models include 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 A5: Estimated Effect of the Reform after Controlling for Month of Birth

Outcome	Full sample	Parental education		
		Low	Mid	High
Panel A: Controlling for birth month				
		Males		
a) Suicide by age 45	-0.0002	-0.0016	0.0013	0.0003
	(0.0008)	(0.0011)	(0.0010)	(0.0010)
b) Mental health-related	-0.0014	-0.0030**	0.0010	-0.0001
death by age 45	(0.0011)	(0.0013)	(0.0012)	(0.0012)
c) Death by age 45	-0.0016	-0.0034	0.0022	-0.0014
	(0.0017)	(0.0022)	(0.0018)	(0.0017)
d) 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) Suicide by age 45	-0.0003	-0.0006	-0.0001	0.0002
	(0.0005)	(0.0005)	(0.0004)	(0.0007)
b) Mental health-related	0.0002	-0.0003	0.0004	0.0005
death by age 45	(0.0005)	(0.0006)	(0.0005)	(0.0008)
c) Death by age 45	0.0007	-0.0006	0.0009	-0.0008
	(0.0011)	(0.0013)	(0.0011)	(0.0012)
d) 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) Suicide by age 45	-0.0002	-0.0016	0.0013	0.0003
	(0.0008)	(0.0011)	(0.0010)	(0.0010)
b) Mental health-related	-0.0013	-0.0030**	0.0010	-0.0001
death by age 45	(0.0011)	(0.0013)	(0.0012)	(0.0012)
c) Death by age 45	-0.0016	-0.0034	0.0022	-0.0014
	(0.0017)	(0.0022)	(0.0018)	(0.0017)
d) 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) Suicide by age 45	-0.0003	-0.0006	-0.0001	0.0002
	(0.0005)	(0.0005)	(0.0004)	(0.0007)
b) Mental health-related	0.0002	-0.0003	0.0004	0.0005
death by age 45	(0.0005)	(0.0006)	(0.0005)	(0.0008)
c) Death by age 45	0.0007	-0.0006	0.0008	-0.0009
	(0.0011)	(0.0013)	(0.0011)	(0.0012)
d) 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 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).

Figure A6: Illustration of the Reform Status in Municipalities across NUTS-2 Regions



Note: Municipal level data

Note: Data are aggregated to the municipal level. Municipalities have been classified according to the NUTS-2 regional level. Figure illustrates the regional allocation of individuals with pre- and post-reform schooling at age 11. For example, in Helsinki region and Other Southern Finland regions (outside Helsinki), most municipalities have roughly the same number of individuals with pre- and post-reform schooling. In Northern and Eastern Finland, the sample is not as well balanced between individuals with pre- and post-reform schooling. Population size of the NUTS-2 regions range from 74,500 to 114,000 inhabitants. Average number of observations per municipality is 788 (median is 410).

Table A6: Timing of Schooling Relative to the Year of Adoption of the Reform in the Municipality of Residence

Birth cohort	Adoption 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 adopted 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 adopted in 1974, entered the 5th grade two years after the first post-reform schooling cohort in the region. Also see Figure 2.

Table A7: Robustness Checks of the Estimated Effect of the Reform on Suicide

	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 3. 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 A6). * = significant at 10%; ** = significant at 5% (all two-sided tests).

Table A8: Robustness Checks of the Estimated Effect of the Reform on Mental Health-Related Deaths

		Parental education		
	Full sample	Low	Mid	High
		Males		
Baseline results	-0.0014	-0.0030**	0.0010	-0.0001
	(0.0011)	(0.0014)	(0.0012)	(0.0012)
a) Controlling for prior mental	-0.0014	-0.0030**	0.0010	-0.0001
health disorders	(0.0011)	(0.0014)	(0.0012)	(0.0012)
b) Controlling for prior mental	-0.0014	-0.0030**	0.0010	-0.0001
health and parents' mental health	(0.0011)	(0.0013)	(0.0012)	(0.0012)
c) Controlling for prior mental	-0.0014	-0.0030**	0.0010	-0.0001
health disorders and its interaction	(0.0011)	(0.0014)	(0.0012)	(0.0012)
with treatment status				
d) Controlling for region-specific	-0.0019	-0.0017	0.0004	-0.0013
linear time trends	(0.0013)	(0.0016)	(0.0015)	(0.0014)
e) Excluding Helsinki	-0.0026*	-0.0043***	0.0019	-0.0012
metropolitan area	(0.0013)	(0.0016)	(0.0013)	(0.0015)
f) Excluding if $t < -3$ or $t > 3$	-0.0013	-0.0024	0.0020	-0.0005
	(0.0012)	(0.0016)	(0.0013)	(0.0013)
g) Extending birth cohorts to 1960–66,	-0.0009	-0.0023	0.0020*	-0.0002
excluding if $t < -3$ or $t > 3$	(0.0011)	(0.0014)	(0.0012)	(0.0012)
		Females		
Baseline results	0.0002	-0.0003	0.0004	0.0005
	(0.0005)	(0.0006)	(0.0005)	(0.0008)
a) Controlling for prior mental	0.0002	-0.0003	0.0004	0.0005
health disorders	(0.0005)	(0.0006)	(0.0005)	(0.0008)
b) Controlling for prior mental	0.0002	-0.0003	0.0004	0.0005
health and parents' mental health	(0.0005)	(0.0006)	(0.0005)	(0.0008)
c) Controlling for prior mental	0.0002	-0.0004	0.0004	0.0005
health disorders and its interaction	(0.0005)	(0.0006)	(0.0005)	(0.0008)
with treatment status				
d) Controlling for region-specific	0.0001	-0.0001	0.0004	0.0008
linear time trends	(0.0006)	(0.0007)	(0.0006)	(0.0010)
e) Excluding Helsinki	-0.0001	-0.0008	0.0002	0.0011
metropolitan area	(0.0006)	(0.0007)	(0.0006)	(0.0010)
f) Excluding if $t < -3$ or $t > 3$	-0.0006	-0.0009	0.0002	0.0005
	(0.0006)	(0.0008)	(0.0006)	(0.0010)
g) Extending birth cohorts to 1960–66,	-0.0003	-0.0009	0.0003	0.0011
excluding if $t < -3$ or $t > 3$	(0.0006)	(0.0006)	(0.0006)	(0.0009)

Note: Baseline results are those reported in Table A2. 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 A6). * = significant at 10%; ** = significant at 5%; *** = significant at 1% (all two-sided tests).

Table A9: Robustness Checks of the Estimated Effect of the Reform 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 A3. 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 A6). * = significant at 10%; ** = significant at 5% (all two-sided tests).

Table A10: Robustness Checks of the Estimated Effect of the Reform 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 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 A6). * = significant at 10%; ** = significant at 5%; *** = significant at 1% (all two-sided tests).

Table A11: 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 death by age 45	Males, low-educated parents	-0.0030	0.025	0.104
b) Mental health-related hospitalizations at ages 16–45	Males, low-educated parents	0.0074	0.039	0.146
c) Mental health-related hospitalizations at ages 16–45	Females, highly-educated parents	0.0095	0.020	0.083
Panel B: Timing of hospitalization				
a) Ages 26–35	Males, low-educated parents	0.0062	0.005	0.022
b) Ages 36–45	Males, low-educated parents	0.0046	0.052	0.183
c) Ages 11–15	Females, mid-educated parents	0.0024	0.014	0.048
d) Ages 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 3–6 and Tables A2–A3. 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 2,000 bootstrap replications.

Table A12: Multiple Hypotheses Testing with Alternative Romano-Wolf (RW) p-values

Outcome	Sample	(1) Baseline RW p-value	(2) RW p-val. with 3 sub-samples	(3) RW p-val. with 6 sub-samples
Panel A: Main outcomes				
a) Mental health-related death by age 45	Males, low-educated parents	0.104	0.075	0.145
b) Mental health-related hospitalizations at ages 16–45	Males, low-educated parents	0.146	0.112	0.185
c) Mental health-related hospitalizations at ages 16–45	Females, highly-educated parents	0.083	0.059	0.119
Panel B: Timing of hospitalization				
a) Ages 26–35	Males, low-educated parents	0.022	0.021	0.038
b) Ages 36–45	Males, low-educated parents	0.183	0.147	0.232
c) Ages 11–15	Females, mid-educated parents	0.048	0.044	0.065
d) Ages 36–45	Females, highly-educated parents	0.144	0.102	0.217
Panel C: Type of mental disorder				
a) Substance-use disorder	Males, full sample	0.062	n/a	n/a
b) Substance-use disorder	Males, low-educated parents	0.106	0.102	0.207
c) Bipolar disorder	Females, highly-educated parents	0.205	0.165	0.288
d) Depressive disorder	Females, highly-educated parents	0.022	0.015	0.035

Note: We only report estimated reform effects that are significant in Tables 3–6 and Tables A2–A3. Romano-Wolf (RW) step-down adjusted p-values are based on standard errors that are clustered at the municipal level. The adjusted p-values have been calculated using 2,000 bootstrap replications. In column (1), the baseline RW p-values are have been computed for the full sample and three sub-samples by parental background (see Table A11). In column (2), RW p-values are have been computed for the three sub-samples by parental background only. In column (3), RW p-values are have been computed for all the six sub-samples by parental background and gender.

Table A13: Significance Levels for the Main Results after Clustering at the NUTS-4 Regional Level

Outcome	Sample	Reform effect	Conventional p-value	Regional clustering p-value
Panel A: Main outcomes				
a) Mental health-related death by age 45	Males, low-educated parents	-0.0030	0.025	0.066
b) Mental health-related hospitalizations at ages 16–45	Males, low-educated parents	0.0074	0.039	0.058
c) Mental health-related hospitalizations at ages 16–45	Females, highly-educated parents	0.0095	0.020	0.015
Panel B: Timing of hospitalization				
a) Ages 26–35	Males, low-educated parents	0.0062	0.005	0.003
b) Ages 36–45	Males, low-educated parents	0.0046	0.052	0.055
c) Ages 11–15	Females, mid-educated parents	0.0024	0.014	0.028
d) Ages 36–45	Females, highly-educated parents	0.0058	0.039	0.057
Panel C: Type of mental disorder				
a) Substance-use disorder	Males, full sample	0.0038	0.017	0.007
b) Substance-use disorder	Males, low-educated parents	0.0048	0.038	0.037
c) Bipolar disorder	Females, highly-educated parents	0.0026	0.058	0.014
d) Depressive disorder	Females, highly-educated parents	0.0062	0.005	0.004

Note: We only report estimated reform effects that are significant in Tables 3–6 and Tables A2–A3. Conventional p-values are based on standard errors that are clustered at the municipal level (465 municipalities). Alternative p-values utilize clustering at the NUTS-4 regional level (i.e., municipalities aggregated to 67 regions). The precision of the estimates does not improve by introduction of an additional control for parents' income (see Table A14).

Table A14: Main Results after Controlling for Parents' Income

Outcome	Sample	Reform effect in the baseline	After controlling for parents' income
Panel A: Main outcomes			
a) Mental health-related death by age 45	Males, low-educated parents	-0.0030 (p = 0.025)	-0.0031 (p = 0.020)
b) Mental health-related hospitalizations at ages 16–45	Males, low-educated parents	0.0074 (p = 0.039)	0.0071 (p = 0.051)
c) Mental health-related hospitalizations at ages 16–45	Females,highly-educated parents	0.0095 (p = 0.020)	0.0095 (p = 0.020)
Panel B: Timing of hospitalization			
a) Ages 26–35	Males, low-educated parents	0.0062 (p = 0.005)	0.0060 (p = 0.007)
b) Ages 36–45	Males, low-educated parents	0.0046 (p = 0.052)	0.0043 (p = 0.064)
c) Ages 11–15	Females, mid-educated parents	0.0024 (p = 0.014)	0.0024 (p = 0.015)
d) Ages 36–45	Females, highly-educated parents	0.0058 (p = 0.039)	0.0056 (p = 0.038)
Panel C: Type of mental disorder			
a) Substance-use disorder	Males, full sample	0.0038 (p = 0.017)	0.0036 (p = 0.020)
b) Substance-use disorder	Males, low-educated parents	0.0048 (p = 0.038)	0.0046 (p = 0.048)
c) Bipolar disorder	Females, highly-educated parents	0.0026 (p = 0.058)	0.0026 (p = 0.057)
d) Depressive disorder	Females, highly-educated parents	0.0062 (p = 0.005)	0.0062 (p = 0.005)

Note: We only report estimated reform effects that are significant in Tables 3–6 and Tables A2–A3. Each cell reports the estimated effect of the comprehensive school reform from separate models. All models include cohort as well as municipality fixed effects. Additional controls are parents' income percentile (0–100) within the child's birth year and whether this information is missing. Parents' income is computed as average over measurement from 1975 and 1985 for non-missing years and for non-missing parents. P-values are based on standard errors clustered at the municipal level.

Table A15: Estimated Effect of the Reform in the Combined Sample of Males and Females

Outcome	Full sample	Parental education		
		Low	Mid	High
Panel A: Main outcomes				
a) Suicide by age 45	-0.0002	-0.0011*	0.0007	0.0004
	(0.0005)	(0.0006)	(0.0005)	(0.0006)
b) Mental health-related	-0.0006	-0.0016**	0.0007	0.0003
death by age 45	(0.0006)	(0.0007)	(0.0006)	(0.0007)
c) Death by age 45	-0.0005	-0.0019	0.0018*	-0.0009
	(0.0010)	(0.0013)	(0.0010)	(0.0011)
d) Mental health-related	0.0028*	0.0051**	0.0002	0.0023
hospitalizations at ages 16–45	(0.0015)	(0.0022)	(0.0026)	(0.0030)
Panel B: Timing of hospitalization				
a) Ages 6–10	-0.0006	-0.0005	-0.0007	-0.0013
	(0.0004)	(0.0007)	(0.0007)	(0.0008)
b) Ages 11–15	0.0002	0.0001	0.0004	-0.0016***
	(0.0004)	(0.0006)	(0.0008)	(0.0006)
c) Ages 16–25	0.0015	0.0025	-0.0013	0.0021
	(0.0011)	(0.0016)	(0.0020)	(0.0018)
d) Ages 26–35	0.0021**	0.0038***	0.0018	-0.0012
	(0.0010)	(0.0013)	(0.0020)	(0.0023)
e) Ages 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 A16: Cox Proportional Hazards Model Results for All-Cause Mortality

		Parental education		
	Full sample	Low	Mid	High
		Males		
Treatment effect	0.965	0.899	1.191	0.865
	(0.036)	(0.059)	(0.160)	(0.143)
	[0.896, 1.038]	[0.789, 1.023]	[0.915, 1.548]	[0.625, 1.196]
Observations	186,777	94,037	47,224	36,282
		Females		
Treatment effect	1.035	0.953	1.257	0.834
	(0.063)	(0.102)	(0.282)	(0.213)
	[0.919, 1.166]	[0.774, 1.174]	[0.810, 1.951]	[0.506, 1.375]
Observations	179,537	90,881	46,092	34,828

Note: Birth cohorts 1962–1966 are at risk from age 11 onwards until age 45. Annual death hazard ratios are reported together with 95% confidence intervals in square brackets. Standard errors reported in parenthesis are clustered at the municipal level. 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). In a combined sample of males and females the treatment effect is 0.985 (0.030) with 95% confidence interval [0.927, 1.046].

Table A17: Estimated Effect of the Refrom on the Number of Days Spent in Hospital for Mental Health Reasons Between Ages 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 A18: Effect of the Reform: Treatment Assignment According to the Municipality of Birth

Outcome	Full sample	Parental education		
		Low	Mid	High
		Males		
a) Suicide by age 45	-0.0004	-0.0017	0.0016*	-0.0010
	(0.0009)	(0.0011)	(0.0009)	(0.0008)
b) Mental health-related	-0.0008	-0.0028**	0.0016	-0.0010
death by age 45	(0.0010)	(0.0013)	(0.0012)	(0.0010)
c) Death by age 45	-0.0014	-0.0026	0.0027	-0.0033*
	(0.0016)	(0.0020)	(0.0019)	(0.0018)
d) MHD by age 45	0.0021	0.0039	0.0048	-0.0071*
	(0.0027)	(0.0035)	(0.0052)	(0.0039)
Observations	194,020	97,366	48,530	37,554
		Females		
a) Suicide by age 45	-0.0002	-0.0001	-0.0001	0.0005
	(0.0004)	(0.0005)	(0.0004)	(0.0005)
b) Mental health-related	0.0002	-0.0001	0.0007	0.0007
death by age 45	(0.0005)	(0.0005)	(0.0005)	(0.0005)
c) Death by age 45	0.0007	0.0000	0.0016	-0.0002
	(0.0010)	(0.0011)	(0.0010)	(0.0012)
d) MHD by age 45	0.0002	-0.0005	-0.0022	0.0066*
	(0.0017)	(0.0026)	(0.0032)	(0.0037)
Observations	186,629	94,138	47,363	36,075

Note: Here, individuals are classified into treatment based on their municipality of birth instead of municipality at age 11 (cf. results in Tables 3–4, Tables A2–A3 and Table A19). Individuals who migrate between ages 11–16 are not dropped from the sample. 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%; ** = significant at 5% (all two-sided tests).

Table A19: Effect of the Reform: Treatment Assignment According to the Municipality of at Age 10

Outcome	Full sample	Parental education		
		Low	Mid	High
		Males		
a) Suicide by age 45	-0.0003	-0.0015	0.0013	0.0000
	(0.0006)	(0.0011)	(0.0010)	(0.0011)
d) Mental health-related	-0.0010	-0.0027**	0.0012	-0.0001
death by age 45	(0.0008)	(0.0013)	(0.0012)	(0.0012)
c) Death by age 45	-0.0004	-0.0023	0.0030*	-0.0012
	(0.0013)	(0.0022)	(0.0018)	(0.0017)
b) MHD by age 45	0.0026	0.0065*	0.0000	-0.0055
	(0.0024)	(0.0034)	(0.0047)	(0.0041)
Observations	181,136	95,620	48,189	37,327
		Females		
a) Suicide by age 45	-0.0003	-0.0006	-0.0003	0.0003
	(0.0003)	(0.0005)	(0.0004)	(0.0007)
b) Mental health-related	0.0001	-0.0002	0.0004	0.0005
death by age 45	(0.0004)	(0.0006)	(0.0005)	(0.0008)
c) Death by age 45	-0.0001	-0.0004	0.0009	-0.0008
	(0.0007)	(0.0013)	(0.0011)	(0.0012)
d) MHD by age 45	0.0023	0.0021	-0.0023	0.0080**
	(0.0019)	(0.0025)	(0.0035)	(0.0039)
Observations	175,304	92,465	46,998	35,841

Note: Here, individuals are classified into treatment based on their municipality at age 10 instead of municipality at age 11 (cf. results in Tables 3–4, Tables A2–A3 and Table A18). Individuals who migrate between ages 11–16 have not been dropped from the sample. 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-at-age-10 fixed effects. Standard errors reported in parenthesis are clustered at the municipal level. * = significant at 10%; ** = significant at 5% (all two-sided tests).

Table A20: Estimated Effect of the Reform 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)
	[12.777]	[12.183]	[12.790]	[14.577]
b) Has a high school degree	-0.0022	-0.0020	-0.0163**	-0.0020
	(0.0037)	(0.0050)	(0.0068)	(0.0082)
	[0.291]	[0.177]	[0.265]	[0.636]
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)
	[0.463]	[0.431]	[0.429]	[0.503]
d) in advanced math	-0.0173**	-0.0104	-0.0370**	-0.0147
	(0.0073)	(0.0128)	(0.0164)	(0.0094)
	[0.538]	[0.500]	[0.511]	[0.571]
e) in basic math	-0.0103	0.0067	-0.0329**	0.0005
	(0.0070)	(0.0132)	(0.0144)	(0.0117)
	[0.506]	[0.485]	[0.493]	[0.535]
Highest degree completed				
f) Vocational secondary	-0.0005	-0.0073	0.0143**	0.0040
education	(0.0042)	(0.0067)	(0.0066)	(0.0074)
	[0.459]	[0.536]	[0.502]	[0.232]
g) Vocational college	0.0020	0.0041	0.0005	-0.0027
	(0.0027)	(0.0038)	(0.0057)	(0.0069)
	[0.140]	[0.116]	[0.157]	[0.192]
h) University master	-0.0048*	-0.0039	-0.0151***	0.0015
	(0.0027)	(0.0028)	(0.0049)	(0.0086)
	[0.113]	[0.055]	[0.090]	[0.301]

Table A20: Continued

Outcome	Full sample	Parental education		
		Low	Mid	High
		Females		
a) Years of schooling	-0.0139	-0.0329	0.0157	0.0126
	(0.0207)	(0.0258)	(0.0443)	(0.0498)
	[13.489]	[12.978]	[13.532]	[14.977]
b) Has a high school degree	-0.0082**	-0.0109**	0.0024	-0.0107
	(0.0041)	(0.0053)	(0.0093)	(0.0077)
	[0.498]	[0.383]	[0.505]	[0.790]
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)
	[0.497]	[0.457]	[0.472]	[0.568]
d) in advanced math	-0.0137*	-0.0206*	-0.0242*	-0.0063
	(0.0071)	(0.0124)	(0.0123)	(0.0111)
	[0.484]	[0.447]	[0.455	[0.528]
e) in basic math	-0.0122**	-0.0229***	-0.0241**	0.0159*
	(0.0061)	(0.0082)	(0.0105)	(0.0084)
	[0.502]	[0.472]	[0.487]	[0.556]
Highest degree completed				
f) Vocational secondary	-0.0049	-0.0056	-0.0174**	0.0075
education	(0.0045)	(0.0059)	(0.0081)	(0.0074)
	[0.379]	[0.454]	[0.394]	[0.194]
g) Vocational college	0.0021	0.0011	0.0110	-0.0052
	(0.0042)	(0.0053)	(0.0074)	(0.0095)
	[0.254]	[0.244]	[0.275]	[0.264]
h) University master	-0.0025	-0.0070**	0.0032	0.0043
	(0.0032)	(0.0035)	(0.0054)	(0.0093)
	[0.148]	[0.086]	[0.131]	[0.337]

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). Mean values of dependent variables are reported in square brackets below the standard error. The full sample also includes individuals for whom information about parents (such as parental education) is missing.

Table A21: Estimated Reform Effect on Economic Outcomes

Outcome	Full sample	Parental education		
		Low	Mid	High
		Males		
Log of average income				
a) at ages 26–45	-0.0092*	-0.0187***	-0.0016	0.0022
	(0.0050)	(0.0065)	(0.0089)	(0.0113)
	[3.307]	[3.234]	[3.332]	[3.528]
b) at ages 26–35	-0.0022	-0.0144**	0.0096	0.0120
	(0.0048)	(0.0057)	(0.0090)	(0.0097)
	[3.111]	[3.063]	[3.135]	[3.247]
c) at ages 36–45	-0.0160***	-0.0234***	-0.0121	-0.0048
	(0.0060)	(0.0080)	(0.0107)	(0.0137)
	[3.444]	[3.350]	[3.462]	[3.707]
Employment rate				
d) at ages 26–45	-0.0029	-0.0057**	-0.0004	0.0044
	(0.0025)	(0.0031)	(0.0044)	(0.0046)
	[0.766]	[0.765]	[0.806]	[0.815]
e) at ages 26–35	-0.0017	-0.0052*	-0.0002	0.0067
	(0.0027)	(0.0031)	(0.0054)	(0.0049)
	[0.740]	[0.739]	[0.775]	[0.776]
f) at ages 36–45	-0.0047	-0.0077**	-0.0008	0.0026
	(0.0027)	(0.0036)	(0.0046)	(0.0050)
	[0.803]	[0.799]	[0.841]	[0.858]

Table A21: Continued

Outcome	Full sample	Parental education		
		Low	Mid	High
		Females		
Log of average income				
a) at ages 26–45	-0.0083**	-0.0145***	-0.0038	0.0018
	(0.0039)	(0.0049)	(0.0078)	(0.0103)
	[3.047]	[2.998]	[3.069]	[3.204]
b) at ages 26–35	-0.0036	-0.0079	-0.0014	0.0025
	(0.0039)	(0.0049)	(0.0071)	(0.0097)
	[2.861]	[2.825]	[2.876]	[2.977]
c) at ages 36–45	-0.0140***	-0.0211***	-0.0102	-0.0006
	(0.0046)	(0.0064)	(0.0093)	(0.0116)
	[3.178]	[3.113]	[3.198]	[3.351]
Employment rate				
d) at ages 26–45	-0.0060**	-0.0066**	-0.0029	-0.0015
	(0.0025)	(0.0029)	(0.0046)	(0.0055)
	[0.726]	[0.732]	0.760]	[0.764]
e) at ages 26–35	-0.0034	-0.0056*	-0.0000	0.0012
	(0.0025)	(0.0033)	(0.0049)	(0.0059)
	[0.673]	[0.668]	[0.701]	[0.704]
f) at ages 36–45	-0.0085***	-0.0078**	-0.0060	-0.0050
	(0.0031)	(0.0035)	(0.0055)	(0.0059)
	[0.782]	[0.790]	[0.821]	[0.824]

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). Mean values of dependent variables are reported in square brackets below the standard error. The full sample also includes individuals for whom information about parents (such as parental education) is missing.

Table A22: 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*	
Suicide by age 45	-0.0018	-0.0017
	[adj. p = 0.337]	[adj. p = 0.285]
Mental health-related	-0.0033	-0.0027
death by age 45	[adj. p = 0.0620]	[adj. p = 0.1139]
Death by age 45	-0.0040	-0.0042
	[adj. p = 0.241]	[adj. p = 0.144]
MHD at ages 16–45	0.0061	0.0060
	[adj. p = 0.301]	[adj. p = 0.367]
MHD at ages 26–35	0.0056	0.0049
	[adj. p = 0.045]	[adj. p = 0.118]
MHD at ages 36–45	0.0038	0.0032
	[adj. p = 0.354]	[adj. p = 0.524]
Substance-use disorder	0.0041	0.0033
at ages 16-45	[adj. p = 0.184]	[adj. p = 0.370]
Outcome	*Panel (B): Females, Highly-educated parents*	
Suicide by age 45	0.0003	0.0005
	[adj. p = 0.925]	[adj. p = 0.843]
Mental health-related	0.0005	0.0007
death by age 45	[adj. p = 0.897]	[adj. p = 0.711]
Death by age 45	-0.0008	-0.0008
	[adj. p = 0.861]	[adj. p = 0.868]
MHD at ages 16–45	0.0096	0.0097
	[adj. p = 0.092]	[adj. p = 0.086]
MHD at ages 26–35	0.0026	0.0025
	[adj. p = 0.633]	[adj. p = 0.654]
MHD at ages 36–45	0.0057	0.0061
	[adj. p = 0.147]	[adj. p = 0.113]
Depression at ages 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 2,000 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 7.

Table A23: The Estimated Reform Effect, Controlling for Education and Income Mediators Later in Life (by Gender)

Outcome / Sample	Baseline estimates	Mediation analysis	
		Education	Educ. & Income
Outcome		*Panel (A): Males*	
Suicide by age 45	-0.0002	-0.0004	-0.0004
	(0.0008)	(0.0008)	(0.0007)
Mental health-related	-0.0014	-0.0016	-0.0015
death by age 45	(0.0011)	(0.0011)	(0.0010)
Death by age 45	-0.0016	-0.0022	-0.0022
	(0.0017)	(0.0017)	(0.0016)
MHD at ages 16–45	0.0027	0.0018	0.0021
	(0.0025)	(0.0026)	(0.0025)
MHD at ages 26–35	0.0020	0.0016	0.0016
	(0.0016)	(0.0016)	(0.0016)
MHD at ages 36–45	0.0018	0.0013	0.0014
	(0.0016)	(0.0016)	(0.0016)
Substance-use disorder	*0.0038***	0.0033**	0.0031*
at ages 16-45	(0.0016)	(0.0016)	(0.0016)
Outcome		*Panel (B): Females*	
Suicide by age 45	-0.0003	-0.0003	-0.0001
	(0.0005)	(0.0005)	(0.0004)
Mental health-related	0.0002	0.0002	0.0002
death by age 45	(0.0005)	(0.0005)	(0.0005)
Death by age 45	0.0007	0.0006	0.0005
	(0.0011)	(0.0011)	(0.010)
MHD at ages 16–45	0.0029	0.0027	0.0029
	(0.0019)	(0.0019)	(0.0020)
MHD at ages 26–35	0.0020	0.0019	0.0019
	(0.0012)	(0.0012)	(0.0013)
MHD at ages 36–45	0.0012	0.0011	0.0013
	(0.0014)	(0.0015)	(0.0016)
Depression at ages 16–45	0.0015	0.0014	0.0016
	(0.0009)	(0.0010)	(0.0010)

Note: We report the results only on outcomes parallel to Tables 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 (using estimates for full sample and three sub-samples of low, mid and high parental education within gender).

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-related hospitalizations (and deaths). 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 7):

1. Estimate an augmented model: $y_{ijt} = \alpha + \eta_j + \tau_t + \beta \cdot REFORM_{jt} + \delta' X_i^{Post} + \epsilon_{ijt}$, 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: $\tilde{y}_{ijt} = \alpha + \eta_j + \tau_t + \kappa \cdot REFORM_{jt} + \epsilon_{ijt}$, 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 zero, 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.