

Education and Health: Long-Run Effects of Peers, Tracking and Years

Martin Fischer, Ulf-G Gerdtham, Gawain Heckley, Martin Karlsson, Gustav Kjellsson and Therese Nilsson*

Abstract

We investigate two parallel school reforms in Sweden to assess the long-run health effects of education. One reform only increased years of schooling, while the other increased years of schooling but also removed tracking leading to a more mixed socioeconomic peer group. By differencing the effects of the parallel reforms we separate the effect of de-tracking and peers from that of more schooling. We find that the pure years of schooling reform reduced mortality and improved current health. Differencing the effects of the reforms shows significant differences in the estimated impacts, suggesting that de-tracking and subsequent peer effects resulted in worse health.

Keywords: Health Returns to Education, Years of Schooling, School Tracking, Peer Effects, Difference-in-Differences

JEL Classification: *I12, I18, I26*

**Acknowledgements:* The authors would like to thank participants at the Health Economics conference in Essen 2017, NHESG in Finland 2017, iHEA in Boston 2017, the Health and Development conference Gothenburg and the Stockholm-Uppsala Education Economics Workshop 2018 for helpful comments. Financial support from the Centre of Economic Demography (CED), the Crafoord foundation and KEFU is gratefully acknowledged (Nilsson). Martin Fischer gratefully acknowledges financial support by the Ruhr Graduate School in Economics and the German Academic Exchange Service (DAAD). Gerdtham is grateful for financial support from the Swedish Research Council (dnr 2014-646). Financial support from the Swedish Research Council (dnr 2019-03553) (Fischer, Heckley, Nilsson) and (dnr 2019-06292) (Heckley) is gratefully acknowledged. The Health Economics Program at Lund University also receives core funding from Government Grant for Clinical Research, and Region Skåne (Gerdtham). The administrative data used in this paper comes from the Swedish Interdisciplinary Panel (SIP), administered by the Centre for Economic Demography, Lund University, Sweden. All remaining errors are the authors' own.

1 Introduction

Education is strongly associated with better health and longer lives, but we still have very limited knowledge as to whether this gradient captures a causal mechanism. There are several reasons to expect a relationship between education and health: education may have a causal effect through its impact on health production, or through impacting financial resources, preferences or self-empowerment, or through improved understanding or access to information that helps improve health (see e.g. Cutler and Lleras-Muney 2006; Grossman 2006). A very large empirical literature has proxied the measurement of education by years of education. The causal evidence of the impact of education measured in this dimension has relied on quasi-experimental evidence and remains inconclusive (see e.g. Cutler and Lleras-Muney 2012; Grossman 2015; Galama et al. 2018), with results being sensitive to methodology, the source of exogenous variation and to vary by time, place and gender.

Galama et al. (2018) highlight that differences in quality aspects of education are a potential explanation for the conflicting results in the literature on the causal effects of education on health. School reforms often involve a package of changes where the year extension is only one of many components possibly working in different directions. Education as an input can be measured more broadly than just in quantitative terms. Quality aspects of education, such as the composition of peers, tracking and curricula changes, might also be important in determining health outcomes. Peers could share important health information or impact social norms and adoption of health behaviours, particularly in settings where perceptions are in a formative stage. Similarly curriculum changes may affect the ability to process information on health-related behaviours and related risks. Insights into the role of quantity and quality aspects education for health is a topical policy issue. They can help policy makers design policies that improve health in the general population.

In this paper we use a unique quasi-experimental set-up in Sweden that allows us to tease apart years of education effects and compare and contrast these to the effects of the removal of tracking and peer group changes on health. This setting includes two similar natural experiments—implemented during a limited time span and exposing the same target population to two treatments that both include a similar increase in the quantity of education; however only one of them varied the quality of education. The comprehensive schooling reform has been analysed in a number of previous studies,¹ whereas this is the first study to use the coincident compulsory schooling extension. Both reforms generated a sizeable increase in years of education, by 0.28 and 0.54 years respectively, which in terms of size is at the top end of the years of schooling literature (Galama et al., 2018).

Thanks to the rather unique research design with two contemporaneous reforms which both affected the schooling acquired by large parts of the population, we are able to make a number of important contributions to the literature. First, we are able to attribute the health effects of a comprehensive schooling reform including de-tracking,

¹The empirical work analysing the Swedish comprehensive school reform e.g. includes work on incomes (Meghir and Palme, 2005), on financial literacy (Lundborg et al., 2018), on intergenerational effects (Lundborg et al., 2014; Holmlund et al., 2011; Lundborg and Majlesi, 2018) and on crime (Hjalmarsson et al., 2015; Meghir et al., 2012).

to different components of the reform package. Many European countries introduced comprehensive schooling reforms that delayed academic tracking in the 1950-1980 period (Brunello et al., 2015) with the aim to improve inequality of opportunity (Holmlund, 2008; Jones et al., 2014). The large body of research which uses such reforms to study the health effects of education cannot separate out the pure years of education effects from other changes since they happened coincidentally.² Several studies report small overall health effects of such reforms (Spasojevic, 2010; Lager and Torssander, 2012; Palme and Simeonova, 2015; Meghir et al., 2018; Lager et al., 2016; Ravesteijn et al., 2017). To the extent that those estimates are the sum of positive effects of schooling and harmful effects of other reform components, it is deeply troubling if the policy conclusion emerging from this body of literature is that education has no causal effect on health.

There is indeed evidence suggesting that components other than the change in years of education have affected health outcomes after such comprehensive schooling reforms. For example, Lager et al. (2016) study the Swedish reform and show emotional control loss for treated high-SES individuals with potential linkage to psychological health following peer composition changes. Böckerman et al. (2019) evaluate the mental health effects of a comprehensive school reform in Finland and find worsened mental health for some subgroups, which they suggest are due to adverse peer groups and discouragement. Moreover, Meghir and Palme (2005) report that lifetime earnings for high-SES individuals decreased by 6 per cent following the reform and attribute this effect to reduced school quality. If income effects translate into longer and healthier lives (c.f. Lindahl 2005; Mazumder 2012) such economic loss could impact population health. Our paper proposes a simple method for how to disentangle the years of schooling effects and effects operating via these other channels.

A second contribution of this paper is that we present estimates of the effects of extended education on health, which challenge the emerging consensus that the effects of increasing the length of compulsory schooling are small, at least in Scandinavia. We exploit a compulsory schooling reform which is unusually suitable for answering this question: it increased the minimum years of schooling from 7 to 8, but there were no other changes to the school system or to the curriculum and it was rolled out in small geographic units over a time period of 20 years, thus reducing concerns about bottlenecks related to input factors and immediate labour supply shocks. The main result coming out of the analysis is that the reform had a beneficial effect on health. Our main estimate on overall mortality (death up to age 81 for the oldest cohort) suggests that an 8th year of schooling reduced mortality by 1.5 percentage points from a baseline of 16.9 per cent. This estimate is statistically significant at the 5 per cent level and we find a significant reduction in death due to circulatory disease. In addition, we are able to corroborate this finding with survey data, showing that also self-reported health and obesity improved thanks to the reform.

²Indeed this is the reason why some researchers have not chosen to use specific reforms as an instrument for years of education (the reform does not provide a valid instrument). The same problem evidently applies to studies using other types of reforms that involved both quantity and quality changes to examine health (see e.g. Jürges et al. (2011) that evaluate a reform involving year and compositional changes) and non-health outcomes of education, see e.g. Kirdar et al. (2015) and Grenet (2013) that evaluate reforms including both year extensions and certification system changes.

Third, we contribute to the growing literature analysing the role of qualitative aspects of education. Tracking changes, which often go hand in hand with quality changes, have been found to impact health: Basu et al. (2018) examine exposure to two different types of secondary schooling systems in England and Wales and show the transition from early-tracking to non-selective school system increased depression and smoking. The already-mentioned studies on Finland (Böckerman et al., 2019; Ravesteijn et al., 2017) attribute findings on mental health and mortality to de-tracking. Also the literature on quality of education focusing on attendance in dissimilar types of secondary school (see e.g. Jones et al. 2011, 2012, 2014) suggests it gives rise to important differences in health and health-related outcomes, but to separate school quality from the characteristics of students that attend school is inherently very difficult. We exploit the co-existence of two different school systems within the same country in order to analyse how different socio-economic groups are affected by de-tracking. Contrary to several of these previous studies, our findings suggest that de-tracking is harmful to health across the board, and not only in higher socioeconomic groups.

Since school tracking and peer effects are inherently linked, our findings also contribute to the literature on peer effects on health. That literature also suggests important interactions between school environment, peers and health. For example Robalino and Macy (2018) suggest peers affect smoking prevalence among high-school students, and Gaviria and Raphael (2001) find strong peer effects in drug use and alcohol drinking among tenth graders. The literature has so far mainly considered short-term effects and primarily examined the impact on health behaviours (see e.g. Sacerdote et al., 2011, for an overview). It is important to understand whether such peer effects sustain long after the peer groups have dissolved since this impacts the interpretation of the short-term effects. If health behaviours such as smoking are only affected contemporaneously and an individual's long-term health preferences remain unaffected, then smoking behaviour may only be impacted in the short term and not impact long-term health outcomes. This in turn affects policy evaluation of school policies such as the removal or introduction of tracking or school choice that affect peer-group composition.

The next section describes the Swedish school system and the two schooling reforms. Section 3 introduces a conceptual model and illustrates what the estimated causal effects of the two reforms captures. Readers not familiar with the Local Average Treatment Effect (LATE) theorem may skip section 3.2 and go straight to the conclusions of the model in section 3.3. Section 4 introduces the data, while section 5 outlines our empirical strategy. Section 6 and 7 present and discuss the empirical results. Finally, we conclude in section 8.

2 The Swedish School System and the Reforms

During the first half of the 20th century Sweden had a selective two-track school system. Students started school at age seven and attended a common compulsory primary school. After the 4th or 6th grade, good performing students (defined by their marks) had the option to study at lower secondary school, which allowed them to continue

to post compulsory upper secondary school and later university.³ Only a small share of students switched to the academic track. Of those disproportionately many had parents that were professionals and administrators (Jonsson, 1991). The majority of the remaining students continued in primary school up to the 7th grade and then left school. This group of students included about 30 per cent of all high ability students. They left after primary school though they likely would have had completed secondary education without problems (see e.g., Husén and Härnqvist, 2000).

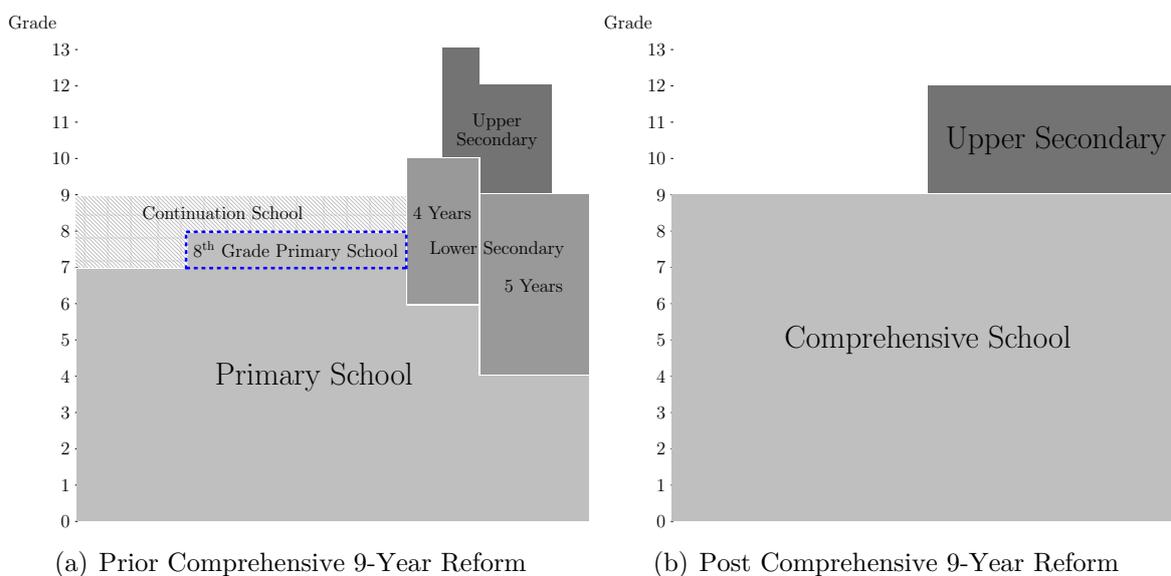


Fig. 1: The Swedish School System

Notes: The figure illustrates the Swedish school system and the reforms considered. Figure (a) shows the one year increase in minimum years of schooling within the old school system for students not matriculating to lower secondary schooling, which we call the 8-year reform. In this system students could take an academic track after grade 4, following a five year track, or after grade 6, following a four year track. Students not enrolling in secondary education took low intensity courses six weeks per year in local continuation schools after finishing compulsory schooling. Figure (b) shows the comprehensive school system which increased minimum years of schooling and postponed tracking, which we call the 9-year reform.

Between 1941 to 1962 a large number of municipalities gradually raised the minimum years of compulsory schooling from 7 to 8 years. We call this the 8-year reform. As illustrated in Figure 1(a) the reform (the first treated birth cohorts were born between 1927 to 1948) was a simple extension of the minimum years of schooling within the municipality for those students who did not go to lower secondary school. The reform was seen as an opportunity to give more time for the students to learn, without any specific changes to the curriculum. About half of all municipalities had introduced a mandatory 8th grade before implementing the comprehensive 9-year reform (see Figure A.1).⁴

³The tracks starting after 4th or 6th grade led to the same degree (*realexamen*) provided the a student passed a written and oral final examination. The examination was developed by the National Board of Education and was the same for the two tracks (Marklund, 1989). The administrative registers do not provide information through which track an individual completed a lower secondary degree.

⁴The previous literature has referred to the 8-year extension as a rare phenomenon mainly occurring

From 1948 to 1969 municipalities also gradually replaced the existing school system with a new comprehensive school system illustrated in Figure 1(b). We call this the 9-year reform.⁵ This reform (affecting birth cohorts born between 1938 and 1959) was very different in character to the 8-year reform: it increased the minimum years of schooling, this time to 9 years, but also postponed tracking of students, with the aim of fostering greater equality of opportunity (Holmlund, 2008).⁶ The removal of tracking broadened the peer group mix (social class and ability composition changes) as all students now shared the same class up to and including the 9th grade.

During the period of reform implementation primary education was free of charge and parents were responsible for the fulfilment of compulsory school attendance (Fredriksson, 1971).⁷ The central government funded a very large share of total costs for primary education by state grants, and through the implementation of the reforms no significant change occurred in the funding of primary schools (Skott, 2011; SOU, 1961).

3 Conceptual Model

We provide a simple theoretical framework for how the two schooling reforms might affect health outcomes. The main aim is to highlight potential differences in the impact of the reforms among different sub-populations—and in particular to provide a theoretical basis for the forthcoming empirical analysis. We will make a number of simplifying assumptions, but in section 3.3 we will argue that most of these assumptions can be relaxed without altering the main predictions of the model.

We assume that the school system produces some type of human capital (such as health), H_{ij} , for individual i attending school j based on the inputs individual ability A and a school peer effect P . Each school year generates additional human capital, h_{ij} , according to the production function:

$$h_{ij} = f(A_{ij}, P_j) = \eta A_{ij} + \theta P(\mathbf{A}_j) \quad (1)$$

where A_{ij} is the individual's own ability and \mathbf{A}_j is a vector of the individual abilities of all pupils in school j . Thus, the peer effects depend on the abilities of the other individuals in the classroom. As we do not specify a particular functional form for the function $P_j(\cdot)$, the model is mute on how these peer effects operate. It is consistent with both the canonical “linear-in-means” model in which the peer effects equal the mean ability of all pupils in the classroom (i.e. $P_j = \frac{1}{N_j} \sum_{i=1}^{N_j} A_{ij}$) and other models in which the heterogeneity of pupils in the classroom affects the students either negatively

in the largest cities, but it was in fact a large reform.

⁵Between 1949 and 1962 the reform was introduced as a *social* experiment in certain areas (Marklund, 1982). The National School Board chose the areas from a group of applicants to form a representative set based on observable municipality characteristics (Holmlund, 2008). In 1962 the Swedish parliament decided that all municipalities should be obliged to offer the new comprehensive school system and to have the new system in place in 1969.

⁶The 9-year reform also introduced a few minor changes to the curriculum. With the reform English was made compulsory from 5th grade. In 1953, English was made compulsory even for the old primary school system so all cohorts born after 1945, untreated or treated, had the same curriculum (Fredriksson, 1971).

⁷In 1950, 0.03 per cent of all children of school age were not in primary education (Sweden, 1974)

(“focus”/“boutique” model) or positively (“rainbow” model) (cf. Sacerdote et al., 2011).

Note that we use the term “peer effects” in the widest possible manner. That is, the impact of the classroom ability profile \mathbf{A}_j on human capital acquisition need not be mediated only through interactions between pupils. It may arise if certain school inputs react to the composition of peers; for example, teaching methods may change with the ability mix of the students. Any quality difference between schools that is a result of the student composition would be interpreted as a peer effect.

In a given school year, an individual will attend a school in either a comprehensive or tracking system. For simplicity, we assume that there are only two ability types in the population and that any tracking system imposes perfect separation of the two types. We let $\tau \in \{0, 1\}$ denote both the ability type and the track. Thus, the school system and the ability type determine the ability profile of a pupil’s peer group.⁸ In the presence of tracking, students with ability A_τ attend track τ , i.e., all high-ability students attend the academic track ($\tau = 1$) and all low-ability attend the basic track (i.e. $\tau = 0$). We denote the ability profile of the peer group an individual of type τ is exposed to in a tracking system as $\mathbf{A}_\tau^t = (A_\tau, \dots, A_\tau)$ for $\tau \in \{0, 1\}$. Similarly, \mathbf{A}^c denotes the ability profile of a pupil’s peer group in a pooled classroom under a comprehensive system.

Under these assumption, h_τ^c denotes the additional human capital an individual of type τ will gain from a year in a comprehensive system:

$$h_\tau^c = \eta A_\tau + \theta P(\mathbf{A}^c) \quad (2)$$

h_τ^t denotes that of a year in a tracking system:

$$h_\tau^t = \eta A_\tau + \theta P(\mathbf{A}_\tau^t) \quad (3)$$

Consistent with the historical context, we assume that the school system prescribes S^c years of comprehensive schooling plus S_τ^t years of tracked schooling (cf. Figure 1), either in a basic ($\tau = 0$) or an academic ($\tau = 1$) track. The length of the tracked schooling, S_τ^t , may differ between tracks. In this school system, the accumulated human capital of an individual of ability type τ will be equal to:

$$H_\tau = S^c h_\tau^c + S_\tau^t h_\tau^t \quad (4)$$

Equation (4) imposes the assumption that an individual’s human capital is piecewise linear in years of schooling. This assumption is not necessary to derive our main results below: it is sufficient that the effects of extending the basic track are linear between 7 and 9 years of compulsory schooling.

⁸Note that we drop subscripts i and j , and instead use subscripts indicating the ability type and school system. Further note that the perfect separation of ability types primarily plays the role of defining the types of individuals who in the presence of a tracking system would have attended a specific track. This separation of types allows us to derive simpler theoretical expressions of the 2SLS estimators of the two reforms. In the appendix B.2, we allow for a tracking system that does not perfectly separate ability types.

3.1 Effects of Reforms

We now consider the impact on human capital acquisition of the two school reforms.

- **The 8-year reform** extends the basic track from 3 to 4 years (i.e. increasing total years of schooling $S^c + S_0^t$ from 7 to 8 years for low-ability individuals) while keeping the rest of the system constant ($S^c = 4$ and $S_1^t = 5$).
- **The 9-year reform** extends compulsory schooling from 8 to 9 years (i.e. increasing $S^c + S_0^t$ from 8 to 9 years) and removes tracking (i.e., $S^c = 9$ and $S_\tau^t = 0$).

We start with the 8-year reform. Assuming that the reform does not affect the selection into tracks, there is no effect on human capital for individuals in the academic track. For individuals in the basic track, the additional human capital is generated from another year of schooling among low-ability pupils. That is, the effect of human capital for an individual of type τ equals:

$$\Gamma_8^\tau = \begin{cases} \eta A_0 + \theta P(\mathbf{A}_0^t) & \text{if } \tau = 0 \\ 0 & \text{if } \tau = 1 \end{cases} \quad (5)$$

The 9-year reform increases S^c from 4 to 9 years, and reduces both S_0^t and S_1^t to 0 (from 4 and 5, respectively). According to the model, the effect of such a reform will be:

$$\Gamma_9^\tau = \begin{cases} 4\theta [P(\mathbf{A}^c) - P(\mathbf{A}_0^t)] + \eta A_0 + \theta P(\mathbf{A}^c) & \text{if } \tau = 0 \\ 5\theta [P(\mathbf{A}^c) - P(\mathbf{A}_1^t)] & \text{if } \tau = 1 \end{cases} \quad (6)$$

where the first term reflects the change in the peer composition due to de-tracking. In the reformed system, both high- and low-ability individuals will be exposed to a pooled classroom instead of a type-specific one from year 5 and onward. The second term for the low-ability individuals captures the additional year of schooling in a pooled classroom adding h_0^c to their human capital (cf. equation (2)).

Now we make a comparison of the two reforms at the aggregate level and for each ability group separately. For each ability group, we obtain the difference between the reforms, denoted as Δ_{89}^τ for $\tau \in \{0, 1\}$, by taking the difference of the relevant terms in equation (5) and (6). For each ability group this difference equals the effect of the change in peer composition for 5 years of schooling:

$$\Delta_{89}^\tau = \Gamma_9^\tau - \Gamma_8^\tau = 5\theta [P(\mathbf{A}^c) - P(\mathbf{A}_\tau^t)] \quad (7)$$

On the aggregate level, the difference is simply a weighted sum of the effects for the high- and low-ability type individuals:

$$\Delta_{89} = \alpha \Delta_{89}^1 + (1 - \alpha) \Delta_{89}^0 \quad (8)$$

$$= 5\theta [P(\mathbf{A}^c) - \alpha P(\mathbf{A}_1^t) - (1 - \alpha) P(\mathbf{A}_0^t)] \quad (9)$$

where α equals the share of high-ability individuals.

A comparison of Δ_{89}^τ and Δ_{89} lead to some immediate conclusions. If the peer composition matters (i.e., $P(\mathbf{A}^c) - P(\mathbf{A}_7^t) \neq 0$), there will be a difference between the effect of the two reforms within the ability group: $\Delta_{89}^\tau \neq 0$. This holds regardless of the type of peer effects operating. On the other hand, a zero difference on the aggregate level is compatible with peer effects. For example, if peer effects are linear in means⁹, then the difference in the effect of the reforms for the ability groups will be of opposing signs and cancel out on the aggregate level. That is, $\Delta_{89} = 0$ as long as $P(\mathbf{A}^c) = \alpha P(\mathbf{A}_1^t) + (1 - \alpha) P(\mathbf{A}_0^t)$. It follows that a situation with a difference between the reforms for both ability groups but not in the aggregate ($\Delta_{89}^\tau \neq 0$ and $\Delta_{89} = 0$) is indicative of peer effects being linear in means. Studying sub-populations will therefore be informative of how peer effects operate.

3.2 Estimation

Next, we briefly assess what a Two-Stage Least Squares (2SLS) estimate of returns to schooling picks up in the case of the two reforms, and illustrate how these can be informative of peer effects. Our point of departure is the LATE theorem (cf. Angrist and Pischke, 2008). Readers not at all familiar with this theorem can skip this section and go straight to the conclusion of the model without a loss of understanding. The LATE theorem states that the 2SLS estimator identifies a local average treatment effect for the sub-population of compliers, i.e., individuals who increase their years of schooling when exposed to a reform. This interpretation of the 2SLS is contingent on the instrument being independent of potential outcomes and only affecting the outcome via years spent in school (i.e., the exclusion restriction).

In the simplest case, the 2SLS estimator is the ratio of the difference in the expected level of human capital (in this case health, H_{ij}) dependent on reform assignment Z_j and the difference in the expected level of schooling (denoted as S_{ij}) dependent on reform assignment:

$$\beta_{2SLS} = \frac{\mathbb{E}[H_{ij} | Z_j = 1] - \mathbb{E}[H_{ij} | Z_j = 0]}{\mathbb{E}[S_{ij} | Z_j = 1] - \mathbb{E}[S_{ij} | Z_j = 0]} \quad (10)$$

We now insert the results of our model into this estimator for each of the two reforms. For the 8-year reform, the numerator is equal to the weighted sum of the effect of the 8-year reform for the two ability types (cf. eq.(5)) and the denominator simply equals the proportion of individuals in the basic track:

$$\beta_{8,2SLS} = \frac{(1 - \alpha) [\eta A_0 + \theta P(\mathbf{A}_0^t)] + \alpha \cdot 0}{(1 - \alpha)} = \eta A_0 + \theta P(\mathbf{A}_0^t) \quad (11)$$

In addition to the LATE assumptions, the result in equation (11) requires that the reform had no spillover effects on subsequent schooling decisions. The reform effect being equal to $\eta A_0 + \theta P(\mathbf{A}_0^t)$ among compliers (low-ability individuals) is dependent on the assumption that these spend only one additional year in school.

⁹With our assumption of two ability types, the linear in means peer effects implies $P_j = \alpha_j A_1 + (1 - \alpha_j) A_0$, where α_j is the share of high-ability pupils in school j

The LATE theorem requires an *exclusion restriction* to be satisfied, according to which the reforms only affect health through an increase in years of schooling. If all the other assumptions are satisfied but not the exclusion restriction, a 2SLS estimate of the effects of years of schooling will be biased. As the above exposition makes clear, the exclusion restriction is not satisfied for the 9-year reform (cf. equation (7)). For the 9-year reform the numerator is equal to the weighted sum for the two ability types and again the denominator is the proportion of individuals in the basic track:

$$\begin{aligned}\beta_{9,2SLS} &= \frac{(1 - \alpha) [\eta A_0 + \theta P(\mathbf{A}_0^t)] + 5\theta [P(\mathbf{A}^c) - \alpha P(\mathbf{A}_1^t) - (1 - \alpha) P(\mathbf{A}_0^t)]}{(1 - \alpha)} \\ &= \beta_{8,2SLS} + \frac{5\theta [P(\mathbf{A}^c) - \alpha P(\mathbf{A}_1^t) - (1 - \alpha) P(\mathbf{A}_0^t)]}{(1 - \alpha)}\end{aligned}\quad (12)$$

Hence, for the 9-year reform we can express the estimate as the effect of an additional year of schooling with the pre-reform peer group ($\beta_{8,2SLS}$) for low-ability pupils, *plus* a second term picking up peer group composition. Also the high-ability group experiences changes in peer composition. As the second term includes effects on high-ability students, who are “always-takers” and would have taken nine years of schooling regardless, $\beta_{9,2SLS}$ cannot be interpreted as a causal effect of years of schooling. Despite this violation of the *exclusion restriction*, $\beta_{9,2SLS}$ still represents a causal effect of all the reform components, weighted by the increase in the years of schooling. Analogously to the results in equation (11), the result in equation (12) requires that the 9-year reform had no spillover effects on subsequent schooling.

Note that $\beta_{8,2SLS} = \beta_{9,2SLS}$ is compatible with situations of no peer effects as well as with peer effects of linear-in means variety cancelling out in the aggregate. It is therefore informative to study subpopulations separately. In our empirical analysis we cannot identify the ability type of an individual, but we may identify two distinct groups in the population who differ in their ability distributions. In particular, we consider the case where a high-SES group has a larger proportion of high-ability individuals. We distinguish the population-level proportion of high-ability individuals α from that in the high-SES (α_u) and low-SES (α_l) groups. Under the maintained assumptions, $\beta_{8,2SLS}$ should not change. However, for the 9-year reform, we now get¹⁰

$$\beta_{9,2SLS}^g = \beta_{8,2SLS} + \frac{5\theta [P(\mathbf{A}^c) - \alpha_g P(\mathbf{A}_1^t) - (1 - \alpha_g) P(\mathbf{A}_0^t)]}{(1 - \alpha_g)}\quad (13)$$

where $g \in \{u, l\}$.

In equations (12) and (13), the difference between the two reforms—the peer effect term—are scaled by the first stage effect of the 9-year reform on schooling (the denominator with $1 - \alpha$). This makes a comparison across (sub)populations with different first-stage effects difficult. We therefore suggest δ^g an unscaled measure of the peer effect:

$$\delta^g = (\beta_{9,2SLS}^g - \beta_{8,2SLS}^g) (1 - \alpha_g)\quad (14)$$

We will report estimates of δ^g in all regression tables. A comparison of equations (12) and (13) reveals that sub-population estimates of δ^g are informative of the type of peer

¹⁰It should be noted that the assumption that $\beta_{8,2SLS} = \beta_{8,2SLS}^u = \beta_{8,2SLS}^l$ can be relaxed in equation (13) without changing the result.

effects that may operate. While any difference between the reforms, $\delta \neq 0$, is indicative of the 9-year reform affecting human capital acquisition via the peer composition in school, $\delta = 0$ does not rule out such mechanisms. Peer effects of opposite sign for the two ability groups (e.g. linear-in-means) would imply $P(\mathbf{A}^c) = \alpha P(\mathbf{A}_1^t) + (1 - \alpha) P(\mathbf{A}_0^t)$ and thus $\delta = 0$.

3.3 Model Conclusions

Based on the simple theoretical model sketched above, we can draw the following conclusions for the empirical analysis:

- The 2SLS estimate $\beta_{8,2SLS}$ captures the effects of a pure extension of years of schooling.
- The difference $\beta_{9,2SLS} - \beta_{8,2SLS}$ is indicative of effects of the 9-year reform not operating via additional years of schooling: such as de-tracking and peer effects.
- If peer effects are predominantly of the “linear-in-means” variety, we expect $\beta_{9,2SLS} \approx \beta_{8,2SLS}$.
- A comparison of $\beta_{9,2SLS}^g - \beta_{8,2SLS}^g$ across groups with different ability mix will pick up peer effects even in the case of “linear-in-means” peer effects.

The exposition above is based on several assumptions, which we now discuss. First, we assume that peer effects θ are homogeneous across ability and socioeconomic (SES) groups. Many studies have challenged this assumption (Burke and Sass, 2013; Brodaty and Gurgand, 2016; Bertoni et al., 2017; Booij et al., 2017). We also assume that peer effects are additive. In Appendix B.2 we relax these assumptions, allowing peer effects to take on any form and being different across groups. The main conclusions are unaffected; however, it becomes clear that our estimates represent a local effect which applies at a specific level of education.

Second, we assume that in the absence of tracking, each pupil is exposed to an ability profile \mathbf{A}^c which applies uniformly to all pupils. In practice, there will be systematic differences across areas, and random variation in the number of high-ability students sampled in each individual class. However, our empirical analysis is conducted at the municipality level with controls for municipality and time fixed effects. Thus, most of the systematic differences in \mathbf{A}^c across time and space are absorbed by these fixed effects.

Third, we assume that *ceteris paribus*, extending compulsory schooling from 7 to 8 years has the same impact on each individual as extending it from 8 to 9 years. This is potentially a strong assumption. In Figure C.2 in appendix, we show the association between years of schooling and mortality conditional on cohort, gender and municipality specific effects and trends for the sample included in our main analysis. Although these are not causal estimates, the graph is indicative of an approximately linear relationship. Still, it is possible to allow for diminishing returns to schooling within the model. Empirically, we will test the sensitivity of δ to our assumption that the years of schooling effect at 9 years is as large as it is at 8 years by scaling $\beta_{8,2SLS}$ in our calculation of δ to simulate diminishing returns to schooling.

Fourth, our model unrealistically assumes that the tracking system perfectly separates ability types. In appendix B.2, we allow for both ability types to attend any track. Allowing for such misclassification does not affect the main results. Any difference between the reform effects is still indicative of peer effects being present. At the same time, high-ability types attending the basic track in the presence of tracking may react differently to a changed ability profile of peers, than low-ability types in the same track. This may affect the possibility to interpret estimates of peer effects. Specifically the possibility of making statements of the type of peer effect at play based on subgroup analysis is reduced if there is a considerable degree of misclassification or if the peer effects among misclassified individuals are large.

Fifth, inserting the results from the model directly into the 2SLS-estimator in (10) also requires that the reforms only affect education up to the ninth grade and have no spillover effects on higher levels of education. We return to this issue in section 5.3 and assess the plausibility and sensitivity of our results to this assumption.

Sixth, our model assumes the effects of schooling are homogeneous between individuals of the same ability. While this simplification is unlikely to hold, comparability of the 2SLS estimates of the two reforms $\beta_{9,2SLS}$ and $\beta_{8,2SLS}$ requires compliers being similar across reforms. We assess the credibility of this assumption in section 5.4 below.

Finally, the model assumes that the main differences in content between the reforms are the abolishment of the tracking system of the 9-year reform. This is supported by the available documentation of the reforms (cf. appendix A). While the 8-year reform primarily changed the amount of time spent in education, the 9-year reform also included other components such as de-tracking, a new curriculum program and methods. As mentioned in section 2, the changes to the curriculum were minor (Fredriksson, 1971), leaving changes in quality related to the de-tracking component and the change in peer composition as main differences between the reforms. Note that any changes in working methods or pedagogics due to the change in peer composition is in line with the model.

Moreover, there is no support for any differences in the change of funding or available resources between the two reforms. Notably, there is generally little data available—even in non-digitalised archives—to study how any changes in funding or resources may interact with reform implementation. However, the available documentation indicates that there was no change in the funding schemes during the study period.¹¹ We return to this potential issue in section 5.3 below.

¹¹Education historians has pointed out that the reforms took place during a period in which the central government had large influence on content, quality checks, and funding (Skott, 2011). In 1940 about 60 per cent of total costs for primary education was covered by grants from the national government. There were also no change in how schools were funded during this period. According to SOU (1961) and Fredriksson (1971) the division of costs between the municipalities and the central government for compulsory education was more or less constant between 1949 and 1958. In the trial period of the 9-year reform, municipalities that implemented the reform got ear marked state grants to cover the increased costs (Holmlund, 2008).

4 Data Description

To examine the impact of the two education reforms on health outcomes we employ both population-based administrative data and survey data. This section provides a general description of the data and their sources. Table 1 presents the summary statistics of the main variables.

4.1 Reform Data

To measure exposure to the 8-year reform we purposively gathered information on the extension of the old primary school from 7 to 8 years from the Swedish National Archives. For each year we digitized information on whether a municipality had seven or eight years of compulsory primary school.¹²

To measure exposure to the 9-year reform we rely on a dataset as used in Hjalmarsson et al. (2015), of which an earlier version is described in detail in Holmlund (2008). The dataset contains information on the year a specific municipality introduced the new comprehensive school system. Online Appendix A provides detailed information on the reforms, their data sets and their validation.

4.2 Individual Administrative Data

The administrative data is drawn from the Swedish Interdisciplinary Panel (SIP).¹³ We consider the universe of those born in Sweden between 1932 and 1952, who survived to 18 years of age and the year 1970, and who had not emigrated from Sweden by 2012.¹⁴

To identify individuals as exposed or unexposed to the reforms we assign treatment status based on year of birth and municipality of residence obtained from the 1960 and 1965 censuses.^{15,16} We do not use municipality of birth to assign treatment status

¹²At the time the country was divided into school districts that generally perfectly overlapped with a municipality. We collect the reform information on the school district level, but here use the term municipality as reform treatment is assigned based on municipality of residence.

¹³The SIP is based upon Statistics Sweden's Multiple Generation dataset to which all datasets are linked using personal identifiers. The SIP is administered at the Centre for Economic Demography, Lund University, Sweden, and approved by the *Lund University Regional Ethics Committee, DNR 2013/288*.

¹⁴Very limited intergenerational information is available for cohorts born before 1932 explaining our chosen starting point. Our schooling measure comes from the 1970 census and covers up to age 18 education explaining our chosen end point of 1952. Survival up to age 18 is chosen to ensure the impact of the reforms on all individuals' education can be fully captured. Survival to 1970 is chosen because this is when years of schooling is measured.

¹⁵We follow Holmlund (2008) and assume that place of residence in 1960 is the municipality where cohorts born between 1943 and 1948 went to school, and similarly the place of residence as recorded in the 1965 census for cohorts born on or after 1949. For individuals born before 1943 we use place of residence of the mother (father if information is missing for the mother) as recorded in the 1960 census.

¹⁶The number of municipalities is not constant over the period considered. Many mainly rural municipalities merged in 1952. We make municipalities consistent over time by mapping pre-1952 municipalities to municipalities as defined by the 1960 census. As it is mainly cities that implement the 8-year reform early on, the municipality mergers in 1952 do not have any major implication for

as the birth place recorded in the registers for cohorts born before 1947 refers to the location of the hospital in which they were born (Skatteverket, 2007), generating severe measurement error (cf. Fischer et al., 2019).

In the cities of Stockholm, Gothenburg and Malmö, both reforms were rolled out in different parts of the city in different years. Therefore we exclude those resident in these cities, which made up 19 per cent of the total population in 1960 (Statistics Sweden, 1961).¹⁷ From the original sample we also drop individuals without reform assignment and those without information on years of education.

Figure 2 illustrates the sample size across reform years and the birth cohorts considered for each reform year. The figure also shows the number of treated municipalities and how the reforms overlap in time. As an example, in 1955, 10 municipalities adopted the 8-year reform and these 10 municipalities have a sample size of 33,793 students born between 1932 and 1952. In the very same year 3 municipalities adopted the 9-year reform, involving 6,048 students born between 1932 and 1952.

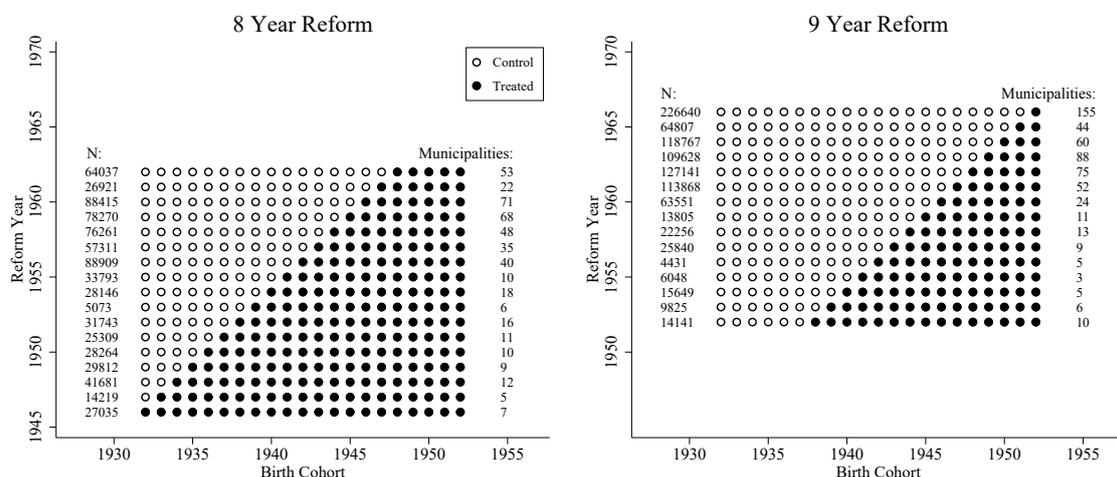


Fig. 2: Observational Sample

Notes: This figure shows the sample size (in the left sidelines) and number of municipalities adopting the reform (in the right sidelines) by reform year for both the 8-year reform and 9-year reform. Total sample size is 1,505,957 and there are 711,361 observations that are in municipalities that never have a 8-year reform and 569,560 observations that are in municipalities that never have a 9-year reform in our sample and these are always used as controls.

From the administrative data we also get information on family SES, defined by father's occupation in 1960. High SES refers to fathers having a high-skilled occupation defined as non-manual employees and professional self-employed using Statistics Sweden's socioeconomic index.¹⁸ Not all individuals have a father recorded in the 1960 census due to emigration or death and therefore our sample split by SES is smaller than our main sample.

our reform assignment. In the few cases when a treated municipality merge with an/other non-treated area(s), we define the whole new municipality as treated.

¹⁷In Appendix A.5 we discuss the implications of this restriction.

¹⁸We take values 3-6 from the variable SEI in the 1960 census.

4.2.1 Education

Data on schooling (primary and upper secondary school) is obtained from the 1970 census and combined with post-mandatory schooling attainments (vocational training and tertiary education) from the Education administrative database. We derive a measure of years of education by assigning the years typically associated with different types of schooling and post-mandatory schooling qualifications and take the sum as an approximation for the total years of education. For example, an individual who takes 8 years of compulsory schooling and then proceeds to a two-year vocational qualification, would be assigned 10 years of education. An individual taking the same qualification after 7 years of compulsory schooling would be assigned 9 years of education.¹⁹

4.2.2 Health Outcomes

Data on mortality comes from the Swedish Cause of Death database. The data covers the years from 1970 up to 2013 and includes cause-specific information up to 2012. We consider the whole observation period and measure the impact on death from the age of 18 up to age 81 for the oldest cohort.

The underlying cause of death and hospitalization is recorded according to the 7th, 8th, 9th and 10th versions of the International Classification of Diseases (ICD) depending on the year of death/admission. We consider the most common causes of death and hospital visits (cancer, circulatory diseases), specific diseases that are related to health behaviours and therefore potentially impacted by education and peer effects: obesity (diabetes), smoking (lung cancer), mental health and alcohol abuse (liver disease).²⁰ Detailed ICD codes are found in Table C.1. Panel A of Table 1 presents the means of years of education and various health outcomes from our population based administrative data based outcomes variables, for the full sample and split by family SES. We observe that on average for our sample that individuals have 9.9 years of education, 17% have died by 2013 and predominantly of cancer. We also note an SES gradient in education and health outcomes with individuals of low SES on average being less educated by about two years and more often having died by 2013 as compared to the high SES group.

¹⁹Previous Swedish population administration data based studies have approximated years of education by the average length associated with the *highest* educational qualification (Hjalmarsson et al., 2015; Lager and Torssander, 2012; Lundborg et al., 2014). In our examples above, each individual following these two trajectories would be assigned the same number of years, calculated as average years of education for individuals who exit the education system with that vocational qualification; if years of compulsory schooling vary between 7 and 8 years in this group, the assigned number will be between 9 and 10 years for all. This matters because some of the compliers to the reforms did go on to get a vocational qualification, the length of which were not affected by the reforms. Thus our method captures the impact of the 9-year reform more accurately than the traditional approach to measuring years of education as it distinguishes schooling and post-mandatory schooling achievements (cf. Fischer et al., 2020).

²⁰It is primarily non-insulin-dependent diabetes that relates to health behaviours, but the data does not allow for a separation between diabetes I and II.

Table 1: Descriptive Statistics

	<i>Whole sample</i>	<i>Low SES</i>	<i>High SES</i>
PANEL A: ADMINISTRATIVE DATA			
Years of Education	9.9	9.4	11.2
Dead	0.168	0.161	0.130
<i>Cancer</i>	0.063	0.062	0.052
<i>Lung Cancer</i>	0.010	0.010	0.008
<i>Circulatory Disease</i>	0.040	0.040	0.028
<i>Diabetes</i>	0.003	0.003	0.002
<i>Mental Disorder</i>	0.003	0.003	0.002
<i>Liver Disease</i>	0.003	0.003	0.002
N	1,505,957	931,668	394,211
PANEL B: SURVEY DATA			
Years of Education	10.0	9.5	11.2
N	24,112	15,124	6,597
Fair or Bad Health	0.232	0.244	0.190
N	24,039	15,077	6,580
Smoke Daily	0.236	0.239	0.228
N	23,881	14,972	6,550
Obese	0.110	0.118	0.089
N	14,008	8,770	3,856
Anxiety, Concern etc.	0.148	0.152	0.139
N	16,261	10,164	4,508

Notes: This table shows the means for education and health outcomes for the whole sample and split by family SES based on father's socioeconomic status 1960. High SES refers to individuals with fathers having a high socioeconomic status occupation. Note not all individuals have a father resident in Sweden or who survived up to 1960, therefore the SES split sub samples do not add to the whole sample.

Source: SIP and ULF-survey. Own calculations.

4.3 Individual Survey Data

The survey data stems from the Swedish survey on living standards (ULF) which is linked to administrative registers.²¹ The survey is reported on a yearly basis 1980-2013 and includes self-reported health and health-related behaviour variables which we consider as valuable complements to our population-based data. We therefore have a representative repeated cross section sample of our population across the years 1980-2013. We use binary indicators for smoking behaviour, obesity, anxiety and self-reported fair or bad health (in contrast to good).²² Panel B of Table 1 presents the means of the survey data. The years of education correspond to those for the administrative data in Panel A, which suggests that the sampling frame of the survey data is representative.

²¹The ULF survey (Statistics Sweden, 2008) is a well respected survey used for a wide range of research and in recent years has formed the Swedish part of the European Union Statistics on Income and Living Conditions (EU-SILC). The survey is carried out by face-to-face interviews of a randomly selected sample of the population. The sample size is about 7,500 individuals per year.

²²We define a binary variable *bad or fair health* equal to one if self-reported health is reported as fair or poor. Smoke Daily is a binary indicator, indicating one if smoked daily in the past 30 days prior to interview, zero otherwise. Anxiety is a binary variable, one indicating whether the individual self-reported having heightened anxiety, concern or worry, zero otherwise. Obese is a binary indicator derived from information on height and weight creating a Body Mass Index (BMI), one indicating a BMI of 30 or more, zero otherwise.

5 Empirical Strategy

5.1 Identifying the Impact of the Reforms

The impacts of the reforms on years of education and health outcomes are modelled as linear functions, using either OLS or a Linear Probability Model (LPM) depending on the characteristics of the outcome variable. We identify the impact of the reforms using a DiD strategy. This exploits the fact that they were introduced at different points in time across municipalities in Sweden.²³ Two individuals born in the same year, where one was resident in a municipality that had already implemented the reform and one in a municipality had not yet implemented the same reform, have different exposures to compulsory schooling. This provides us with variation in reform exposure both over time and across municipalities. However, the implementation was not random as discussed in Holmlund (2008). To control for this we difference across municipalities and across birth cohorts by using dummy variables for both. As an additional control we include municipality specific trends. Our linear DiD-model takes the form:

$$H_{i,c,m} = \gamma_0 + \gamma_1 Z_{c,m} + \gamma_2 C + \gamma_3 M + \gamma_4 trend_m + \epsilon_{i,c,m}; \quad (15)$$

for individual i , birth cohort c and municipality m . $Z_{c,m}$ is a vector including our two instruments defined as indicator variables if a cohort and municipality are exposed to the school reform, C is a vector of birth year cohort dummies, M is a vector of municipality dummies, the $trend$ is a vector of municipality specific trends and γ_0 is a constant term. The coefficient γ_1 captures the average impact across municipalities of the reforms on education/health controlling for cohort and municipality specific differences and also municipality specific linear trends.

The two reforms affected different shares of the population and increased schooling by different amounts. In order to get estimates of effect that are comparable across reforms, we apply 2SLS. The first stage equals our linear equation (15) with years of education E in place of H as the dependent variable. The second stage is:

$$H_{i,c,m} = \beta_0 + \beta_1 \hat{E}_{i,c,m} + \beta_2 C + \beta_3 M + \beta_4 trend_m + v_{i,c,m}; \quad (16)$$

where subscripts i , c , and m , and variables C and $trend$ are as for equation (15). \hat{E} is the prediction of years of schooling from the first stage. The coefficient on years of estimated education β_1 is identified by the variation in years of education that comes from the variation generated by the school reforms. We estimate equation (16) separately including the relevant reform as an excluded instrument and the other purely as a control variable.

²³A recent literature has highlighted potential problems of using a DiD-strategy with a staggered introduction, see e.g. Athey and Imbens (2018), de Chaisemartin and d'Haultfoeuille (2019), Goodman-Bacon (2018) and Strezhnev (2018). Our specification includes municipality level trends and therefore, the criticism noted of two way DiD models is not as relevant as we no longer assume common trends, rather our variation now comes from sharp deviations introduced by the school reforms from otherwise linear trends. We also include a large number of never treated which reduces the size of any potential bias. Notably, we also provide RDD-based results that are comparable to our DiD results. RDD should not be subject to the problems discussed in this literature and suggests that the issue raised is less of a concern in this empirical application.

For our 2SLS coefficients to capture the causal effect of another year of education, it is necessary to assume that the reforms affected health outcomes only via their effects on years of education. This exclusion restriction will be violated if the reforms had other impacts on students over and above their impact on years of education that then impacted on health. We argue that the 8-year reform is a pure years of schooling reform because neither the structure of the education system nor the curriculum were altered during the implementation. In that case, the 8-year reform satisfies the exclusion restriction requirements and our 2SLS estimate has a causal interpretation.

In the conceptual model, we illustrate that the 2SLS estimate of the 9-year reform is biased as an estimate of the effects of additional years of schooling: it equals the sum of this schooling effect and any additional health effects of de-tracking. This reform abolished the old school structure that tracked students from fourth/sixth grade and instead, students were kept in the same school and classes until the ninth grade.^{24,25} The removal of tracking by the 9-year reform led to a change in the peer composition, increasing the heterogeneity in the mix of ability and socio-economic status the students were exposed to. This potentially impacted health outcomes, through a change in teaching methods or through peer interactions. Health-related behaviours as peer effects have been found to impact behaviours such as drinking, smoking, criminal behaviour and drug use in the short-term (see e.g. Sacerdote et al., 2011, for an overview on peer effects). Such effects lead to a violation of the exclusion restriction, invalidating the 9-year reform as an instrument for years of schooling.

We estimate the 2SLS model in equation (16) separately for the two reforms, for both the total population and subgroups based on the SES background (father's socioeconomic status). Guided by the conceptual model, we obtain an unscaled estimate of the average peer effect, δ , by subtracting the 8-year reform β_1 from the 9-year reform β_1 , and multiplying the difference by the coefficient γ_1 from the first stage of the 9-year reform (see equation (14)). We obtain such estimates for the total population, as well as for the subgroups with high and low SES background. Standard errors for δ are obtained by cluster bootstrap with 999 repetitions.

To perform an additional statistical test of peer effects, we also estimate a combined version of equation (16) where both reforms are included as excluded instruments for years of education. By combining the reforms we can perform a Sargan-Hansen type of over-identification test; testing the null hypothesis that both reforms are exogenous (Sargan, 1958; Baum et al., 2003). A rejection of this test will strongly suggest that one of the reforms fails the exclusion restriction. Under the assumption that the exposure to the 8-year reform is a valid instrument, we can use this as a statistical test of whether the removal of tracking and consequent peer group composition effects of the 9-year reform had significant independent effects on health.

In summary, we are able for the first time to provide credible evidence from Sweden of

²⁴There were some exceptions, where tracking was maintained within the same physical school for some subjects in the very final 9th grade during the first years of the trial period, but overall students were much more mixed.

²⁵Normally peer group effects are a concern as they lead to violation of the Stable Unit Treatment Value Assumption. However, as our analysis is at the municipality level our DiD results are internally valid for the combined impact of the years of schooling and peer group composition effects.

the causal impact of *years* of education on health using the 8-year reform as an IV. In addition, by comparing the two reforms using the guidance of our conceptual model we can also assess whether the 9-year reform had wider effects on health other than through just years of education.

5.2 First-Stage Results

Our identification strategy builds upon our method of treatment status assignment performing well. In addition to this assumption and the exclusion restriction discussed above, our 2SLS estimates requires an impact of the reforms on years of education and that reform assignment impact is as good as random given our control strategy. In this section we establish the existence of a first stage.

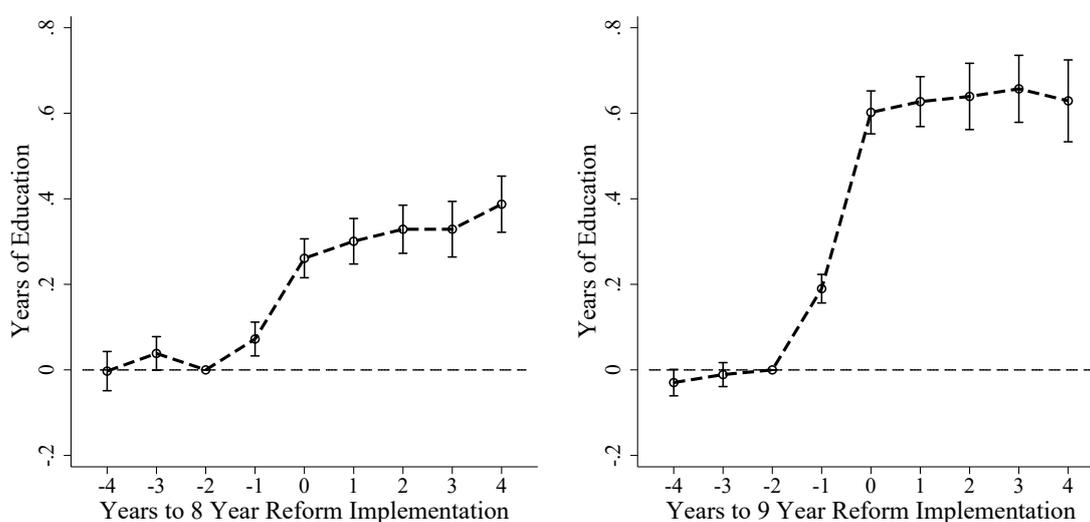


Fig. 3: Event Study Analysis of the Reforms on Years of Education

Notes: Event study scatter plots that depict the difference in average years of education by birth cohort relative to the birth cohort born two years prior the first cohort impacted by the reform in their municipality. Left panel is for the 8-year reform, right panel the 9-year reform. Reform implementation is at time zero. Point estimate 95% significance level confidence intervals are represented by the spikes. *Source:* SIP. Own calculations.

Figure 3 presents one event study graph for each reform. These graphs depict the difference in average years of schooling by birth cohort relative to the birth cohort born two years prior to the first cohort impacted by the reform in their municipality. The graphs show a jump in the average years of education at the reform year for both reforms, and that this jump is larger for the 9-year reform. This is expected because the 9-year reform increased years of schooling by two years for students who were in municipalities offering only 7 years of primary school and one year for those offering 8 years, whereas the 8-year reform increased years of schooling by one year in all municipalities affected.

Note that we also see a jump in years of schooling in period $t-1$ and this is also much clearer for the 9-year reform than the 8-year reform. Hjalmarsson et al. (2015) suggest that the pre-reform increase in schooling is due to either measurement error in the

exposure variable or due to pupils being in the wrong grade based on their age. They cite evidence that grade repetition was not a common occurrence for those in the old 7 year primary school system but was for those who were tracked into the junior secondary school. Those at junior secondary school who were born a year too early but had dropped out would have normally gone back to old primary school, but because of the reform they would have instead been caught by the 9-year school reform and would as a consequence be a year older than their peers in the same class. This story fits with what we see in the data: A small jump in t-1 for the 8-year reform where grade repetition was not very common in the old primary school system and a larger jump in t-1 for the 9-year reform where grade repetition at junior secondary school was more common. Accepting that this type of measurement error is causing the jump one year prior to the reforms, implies that the jump should be considered as a part of the reform effect. Notably, our first-stage regression would in that case underestimate the effect of the reform on years of education. It will however also mean we have the same measurement error in our reduced form estimates and therefore the measurement error should cancel out in our 2SLS estimates.

Table 2 presents the regression results of the impact of the reforms on years of education. In column (1) of Table 2 we present the results for the 8-year reform on years of education for all individuals and find an increase of 0.28 years of education. In column (4) we present the results for the 9-year reform for all individuals and we find an impact of 0.54 years. Note that the 9-year reform estimates presented here are much larger than previously documented (see e.g. Holmlund 2008; Lundborg et al. 2014; Meghir et al. 2018). This follows from the fact that our measure of years of education is different from what has previously been used, and this new measure better captures the impact of increased compulsory schooling on years of education (see Fischer et al., 2020, for a detailed discussion).

Table 2: Compulsory Schooling Reforms' Impact on Education

	(1)	(2)	(3)	(4)	(5)	(6)
	8-YEAR REFORM			9-YEAR REFORM		
	ALL	LOW-SES	HIGH-SES	ALL	LOW-SES	HIGH-SES
Coefficient	0.277*** (0.021)	0.284*** (0.020)	0.174*** (0.031)	0.542*** (0.024)	0.654*** (0.023)	0.346*** (0.025)
F-stat	178.29	206.40	31.49	521.19	798.55	196.97
N	1,505,957	931,668	394,211	1,505,957	931,668	394,211

Notes: This table shows the impact of the 8-year and 9-year school reforms on years of education from a DiD regression controlling for cohort and municipality fixed effects gender and municipality trends. SES is defined as whether father has a non-manual or professional self-employed occupation. Robust standard errors clustered at the municipality level are in parentheses. Testing the null of the coefficient: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Source: SIP. Own calculations.

The results split by socio-economic background show that the 9-year reform had a larger impact on years of education compared to the 8-year reform for both high- and low-SES individuals. These results also show that the impact for low-SES individuals is larger than for high-SES individuals for both reforms, but of similar proportion across

the reforms. The F statistics suggest that we have a strong first stage across the board, based on the the rule of thumb for weak instruments of an F statistic above 10 (Stock et al., 2002).²⁶

5.3 *Spillover Effects and Resources*

Our conceptual model in Section 3 relies on the assumption that there are no spillover effects of the reforms. Such spillover effects would arise if either compliers to the reforms were incited to take more than compulsory schooling, or if the always-takers took additional schooling in order to further differentiate themselves from the compliers. Since both behaviours can be rationalised with reference to economic theory (cf. Fischer et al., 2019), their importance needs to be assessed empirically. Appendix Table C.3 presents estimates of how different levels of education are affected by the reforms. It shows the expected reduction in the previous compulsory level(s) to the new level. It also shows the absence of spillover effects into tertiary education. As regards upper secondary schooling, Table C.3 shows that there are no spillover effects of the 8-year reform at these levels. For the 9-year reform, we do find a small increase in the propensity to take two- and three-years of secondary level education. The joint increase of about 4 percentage points corresponds to the difference between the decrease of individuals with the previous level of mandatory schooling (7 or 8 years) and the increase of individuals with the new mandatory level (9 years), and is equivalent to an increase of years of schooling by about 0.1.

In Appendix B we discuss the implications of this spillover effect due to the 9-year reform for the interpretation of our results: we conclude that our estimate of δ includes a bias term which will depend on how much the effects of schooling on health deviate from linearity; and that for a wide range of such deviations, the bias will be small in magnitude.

Another important issue is whether the reforms had an impact on the resources devoted to each pupil-year. If there are differences between the two reforms in this regard, they may lead to biases in our results. Using available information on the number of teachers in a municipality from the censuses in 1950, 1960, and 1970, we estimate the effects of the two reforms on pupil-teacher ratios using our main DiD specification, cf. Appendix Table A.2. These results suggest that the increase in the number of students was larger than the increase in number of teachers for both reforms. The recorded increase in pupil-teacher ratio is larger for the 8-year reform -- implying that to the extent that these differences in resources are relevant for health production, our estimates of δ will represent a lower bound.

5.4 *Diagnostic Tests*

We have shown that the reforms coincide with substantial increases in years of education and therefore that our method of reform assignment is working well. In addition to a strong first stage, our 2SLS estimates require exposure to reform to be as good as random, conditional on our control strategies. This may be violated if selective

²⁶In the appendix we provide a corresponding table to Table 2, but calculated using the survey data and find similar coefficient estimates.

migration to and from reform municipalities occurred, either to escape or gain access to the reform. In previous work assessing the 9-year reform, both Meghir and Palme (2005) and Holmlund (2008) have tested for selective migration and have found that it was not a problem. Similarly, Fischer et al. (2019) show that the previous 7-year reform had no impact on migration either and that the probability to move to another school district was close to zero in the ages 6-16 for cohorts born 1930--1940. We are not able to test selective migration regarding the 8-year reform, but make the assumption that the results of these two other reforms apply to the 8-year reform as well. We view this as a plausible assumption given that the 8-year reform affected similar cohorts but was just a pure years of schooling change and would have provided much less of a reason to move compared to the comprehensive 9-year school reform.

Table 3: Diagnostics: Balancing Test for Differences in Father's Predetermined Characteristics by Reform Status

	(1)	(2)	(3)	(4)	(5)	(6)
	8-YEAR REFORM			9-YEAR REFORM		
	OLS	DiD	DiD+TRENDS	OLS	DiD	DiD+TRENDS
Years of Education	0.304*** (0.037)	0.023* (0.012)	-0.008 (0.011)	0.182*** (0.034)	0.037*** (0.011)	0.003 (0.008)
N	663,313	663,313	663,313	663,313	663,313	663,313
No Occupation	-0.006*** (0.001)	0.004*** (0.002)	0.001 (0.001)	0.002 (0.001)	-0.001 (0.002)	0.001 (0.001)
Agricultural Worker	-0.158*** (0.013)	0.005* (0.002)	-0.002 (0.002)	-0.073*** (0.013)	0.002 (0.003)	-0.001 (0.002)
Blue Collar Worker	0.050*** (0.009)	-0.016*** (0.004)	-0.001 (0.003)	0.022*** (0.008)	-0.010*** (0.004)	-0.001 (0.002)
White Collar Worker	0.114*** (0.011)	0.007*** (0.002)	0.001 (0.002)	0.048*** (0.011)	0.008*** (0.002)	0.001 (0.002)
N	1,325,239	1,325,239	1,325,239	1,325,239	1,325,239	1,325,239
High Socioeconomic Status	0.112*** (0.011)	0.006*** (0.002)	0.004 (0.003)	0.045*** (0.011)	0.007*** (0.002)	-0.001 (0.003)
N	1,325,879	1,325,879	1,325,879	1,325,879	1,325,879	1,325,879

Notes: This table shows impact of reform status on various predetermined characteristics. Columns (1) and (4) are simple associations controlling for year of birth. Columns (2) and (5) are estimates from a DiD regression controlling for cohort and municipality fixed effects and columns (3) and (6) are the same as columns (2) and (5) plus linear municipality trends. Robust standard errors clustered at the municipality level are shown in parentheses. Testing the null of the coefficient: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Source: SIP. Own calculations.

To assess whether reform assignment is as good as random, we perform a batch of balancing tests of predetermined characteristics and reform assignment in Table 3. The results show that when we control only for birth cohort fixed effects (columns 1 and 4) our predetermined characteristics are predicted by reform status. The correlations also go the way we might expect: the reforms were introduced earlier in areas where parents were better educated and had better jobs. Also after introducing cohort and municipality fixed effects (columns 2 and 5) most predetermined characteristics are predicted by reform status. In columns (3) and (6) we use our DiD specification which also adds linear municipality trends to identify the impact of the reforms. We find that the size of the coefficients here tend towards zero and become insignificant. Whilst

this evidence is just indicative that our reforms are not correlated with the error terms after applying our identification strategy, they provide credibility to our strategies. Conversely, a specification which does not remove the correlation with observable characteristics, is very likely to suffer from bias due to unobserved characteristics as well. In sum, our assignment method works as implied, and there is also evidence supporting our claim that reform exposure in our DiD strategy is as good as random.²⁷

Finally, we characterise and compare the compliers to the two reforms. Results presented in appendix sections C.3.1 and C.3.2 also confirm that individuals responding to the two reforms are very similar in terms of observables, which further strengthens the case for comparing the two reforms.

6 Results

This section presents our main analysis of the impact of both school reforms on health. We first present the results for mortality and cause-specific mortality, and then findings for self-reported health and health behaviours.

6.1 Mortality

Figure 4 shows an event study analysis for the two reforms depicting the difference in mortality between birth cohorts within the municipality. The reference group is again the birth cohort born two years prior to the first cohort impacted by the reform (in their municipality). Although no coefficients are individually significant there is a pattern in the two graphs that indicate a negative mortality effect of the 8-year reform and a (smaller) positive effect of the 9-year reform.

This pattern is confirmed in Table 4 that presents the reduced-form DiD regression estimates of the impact of both reforms (rows 1 and 2) and the 2SLS estimates using each reform separately (rows 3 and 4) for mortality. Column (1) presents the results for our main sample. In columns (2)-(3) we consider heterogeneity by father's socioeconomic status. The two SES groups likely include different shares of individuals who were directly affected by the change in the mandatory schooling induced by the reforms (i.e. in terms of the conceptual model, the share of high ability types, α , is likely to be larger in the high-SES group). This is confirmed by the heterogeneous effect of an increase in the years of schooling. As spelled out in the conceptual model these groups are likely exposed to different changes in the peer composition induced by the de-tracking.

The reduced-form estimates of the 8-year reform indicate a statistically significant negative effect. The impact on mortality in the main sample is -0.4 percentage points. Although this result to some degree is driven by a larger effect among high-SES

²⁷In Appendix table A.1 we complement this analysis with regressions of reform years on some baseline municipality characteristics. It confirms previous findings that the reforms correlated with urban status and local earnings; however, we extend this to show that also life expectancy at baseline predicts reform year for both reforms. Taken together, these results provide some intuition as to why the municipality trends are required to achieve identification: secular trends in life expectancy may be correlated with reform status, and previous research shows that the 1940-80 period was characterised by strong regional convergence in incomes (Enflo and Rosés, 2015).

individuals, the estimate for the low-SES group is of similar size as for the full sample. The reduced form estimates for the 9-year reform are all smaller than what is found for the 8-year reform, of opposite sign, and insignificant.

Since the 2SLS estimates can be interpreted as the reduced-form estimates weighted by the change in years of schooling from the first stage regression in Table 2, these allow us to compare the effect of the two reforms in a standardised way. The difference in magnitude of the first stage estimates causes the difference between the 2SLS estimates of the two reforms to be much larger than the difference in the reduced form estimates. 2SLS estimates based on the 8-year reform indicate that an additional year of schooling reduces mortality in the full sample by 1.5 percentage points, and by 1.5 and 3.2 p.p. in the low and high SES groups respectively. The corresponding 2SLS estimates for the 9-year reform are all much smaller in magnitude, of opposite sign, and insignificant.²⁸

Table 4 also presents our estimates of the (unscaled) de-tracking and peer effects on mortality, δ , which are computed as the difference between the 2SLS estimates of the two reforms multiplied by the first stage of the 9-year reform (see Table 2). These estimates are positive on the aggregate level as well as in the two subgroups, but this is less precisely estimated for the SES subgroups. In magnitude, these results imply that the adverse effects of de-tracking and peer group composition changes associated with the 9-year reform correspond to 2/3 of the effect of one additional year of schooling in the full population (i.e., the 2SLS-estimates of the 8-year reform).

Our additional strategy to statistically test the hypothesis of an effect of the de-tracking and peers using the Hansen J statistic test for over-identification confirms these results. *Overid* presents the p-value from such a test statistic from a 2SLS regression including both reforms as instruments. We interpret this as a test of the validity of the 9-year reform as an instrument. A rejection of the null implies the exclusion restriction is unlikely to hold, suggesting that peer and de-tracking effects are different from zero—on the aggregate level. We find such evidence for mortality in the main sample and for both SES subgroups, which suggests that the overall impact of de-tracking and peer group composition changes on long-term health is in fact significant and has worked against the health improving increase in years of schooling element of the 9-year school reform.

²⁸Note that the documented grade repetition may cause a minor underestimation of the 9-year reform 2SLS estimate. Importantly this does not affect our comparison of the two reforms, as δ rescales the difference between the two 2SLS estimates by the first stage

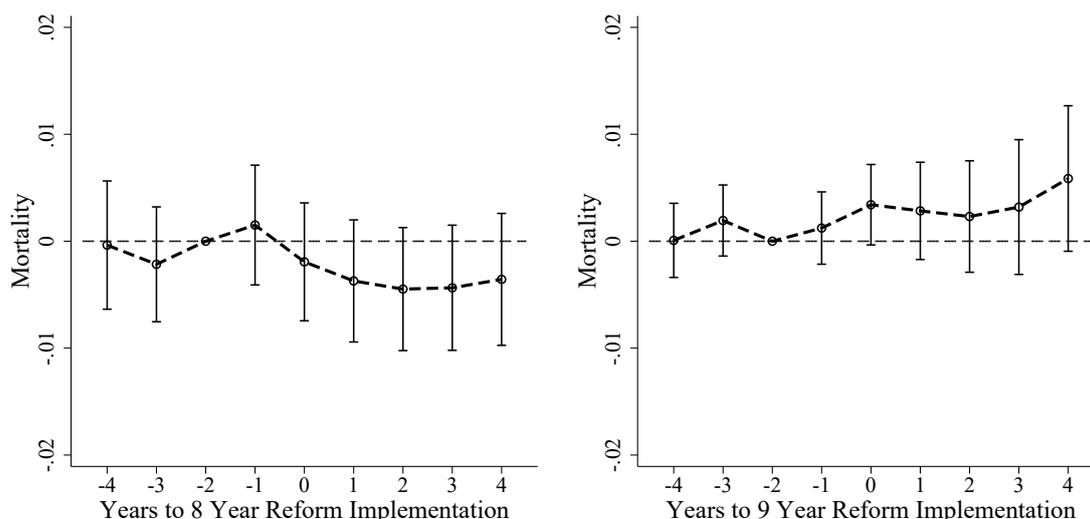


Fig. 4: Event Study Analysis of the Reforms on Mortality

Notes: Event study scatter plots that depict the difference in death rates by birth cohort relative to the birth cohort born two years prior the first cohort impacted by the reform in their municipality. Left panel is for the 8-year reform, right panel the 9-year reform. Reform implementation is at time zero. Point estimate 95% significance level confidence intervals are represented by the spikes.

Source: SIP. Own calculations.

Table 4: Education Effects on Mortality

	(1)	(2)	(3)
	ALL	LOW-SES	HIGH-SES
RF 8-Year Reform Impact	-0.004** (0.002)	-0.004* (0.002)	-0.006** (0.003)
RF 9-Year Reform Impact	0.002 (0.002)	0.001 (0.002)	0.002 (0.002)
IV 8-Year Reform	-0.015** (0.006)	-0.015* (0.008)	-0.032** (0.016)
IV 9-Year Reform	0.003 (0.003)	0.001 (0.003)	0.005 (0.007)
δ	0.010*** (0.004)	0.010* (0.006)	0.013** (0.006)
Overid	0.008	0.078	0.031
Mean	0.168	0.161	0.130
N	1,505,957	931,668	394,211

Notes: This table presents the impact of the reforms and years of education on mortality in the full sample (column 1) and by father's SES (columns 2-3). The sample is cohorts born 1932-1952. All coefficients are from separate regressions (RF=Reduced form; IV=Two Stage Least Squares). IV estimates include the relevant reform as an excluded instrument and are based on DiD with cohort, municipality fixed effects and municipality specific trends. "Overid" is the p value from a Hansen J statistic testing the validity of the instruments when both are used simultaneously as instruments for years of schooling in the same regression (Baum et al., 2003). Standard errors clustered at the municipality level in parenthesis. δ is our measure of peer effects calculated as in equation 14 and standard errors are provided by parametric bootstrap accounting for clustering with 999 repetitions. Testing the null hypothesis: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: SIP. Own calculations.

6.2 Cause-Specific Results

Figure 5 presents the 2SLS results for mortality by the main cause of death. Causes considered are cancer, circulatory disease, diabetes, mental disorder, lung cancer and liver disease. Cancer and circulatory disease are the most common causes of death for the cohorts we consider and diabetes, lung cancer and liver disease are particularly amenable to changes in health-related behaviours, a key channel education and peer group changes could influence. Cancer, especially lung cancer and circulatory disease, are strongly predicted by smoking.²⁹ Liver disease and diabetes are very strongly linked to heavy alcohol consumption and poor diet/lack of exercise respectively, behaviours linked to short-term peer group effects. Mental health is potentially impacted both by health behaviours but also peer effects directly.

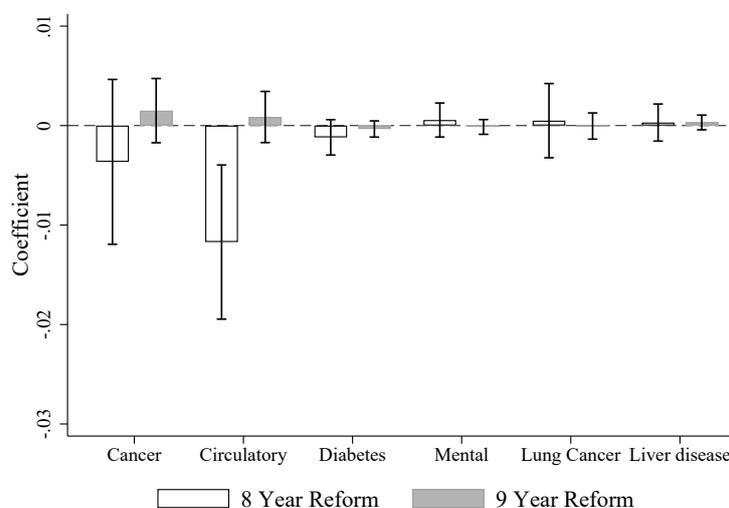


Fig. 5: Impact of the Reforms on Cause-Specific Mortality by 2013

Notes: This figure plots the 2SLS coefficient of years of education on mortality by main diagnosis. Lung cancer is included in cancer. 95% confidence intervals are represented by the capped vertical lines.

Source: SIP. Own calculations.

The mortality impacts by cause are all statistically insignificant with the exception of circulatory diseases which appears to be the driver behind the reduction in mortality caused by the 8-year reform. The 9-year reform 2SLS estimates are smaller than the 8-year reform 2SLS estimates, nearly all are positive and all are insignificant. The differences in estimated effects between the reforms support the interpretation of a health penalty from de-tracking due to the 9-year reform, driven by death from cancer and, in particular, circulatory disease. Note that, all-cause mortality is our primary endpoint whereas the analysis of specific death causes is exploratory in character. Nevertheless, it is of interest to know whether the conclusion emerging from figure 5 is robust to adjustment for multiple hypothesis testing. When we apply the two-step procedure proposed by Benjamini et al. (2006) to control for the false discovery rate, none of the conclusions regarding statistical significance changes.

²⁹Cancer and circulatory disease are also linked to nutrition and health early in life (cf. Bhalotra et al., 2017).

We also explore the impact on mortality by cause split by father's SES. As illustrated in Figure 6 the significant reduction in mortality is due to circulatory disease, although the impact is smaller and not significant at the 5% level for the low SES group compared to the high SES group. The high SES group also observe a large and significant effect on death due to cancer.

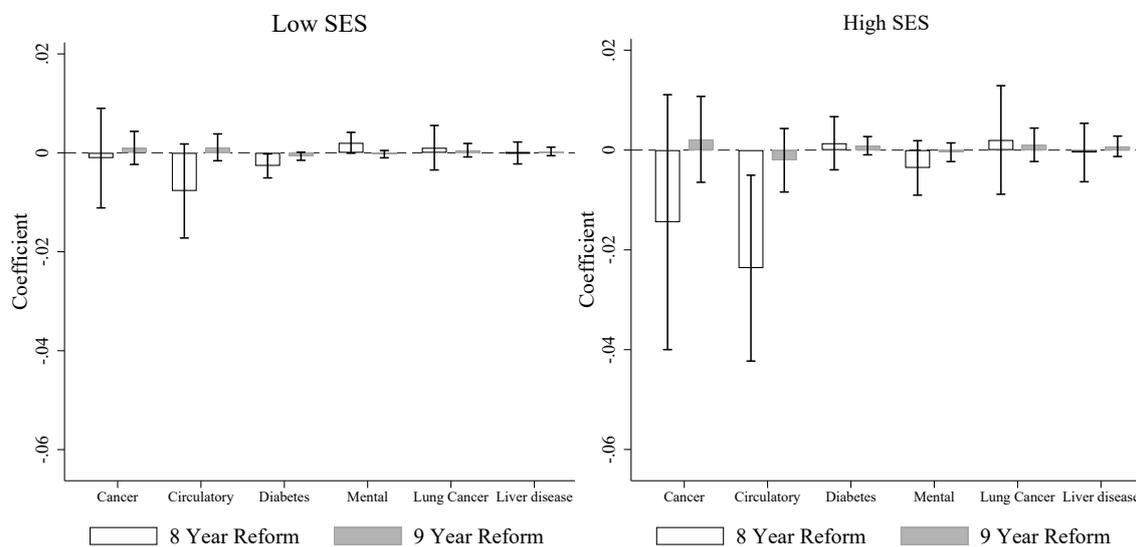


Fig. 6: Impact of the Reforms on Cause Specific Mortality by Family SES

Notes: This figure plots the 2SLS coefficient of years of education on mortality by main diagnosis and fathers education high/low. Lung cancer is included in cancer. 95% confidence intervals are represented by the capped vertical lines.

Source: SIP. Own calculations.

6.3 Self-Reported Health and Health Behaviours

Next, we examine the effects on self-reported health and health behaviours using survey data. In Table 5 the reduced-form estimates of the 8-year reform in the first row indicate a non-significant health improving impact for fair/bad health of -2.2 percentage points.³⁰ The corresponding reduced form estimate for the 9-year reform is also insignificant, of about the same magnitude, but of opposite sign.

Turning to the 2SLS estimates, the 8-year reform results indicate that an additional year of schooling reduces fair/bad health by 6.2 percentage points. The analogous estimate for obesity is also imprecisely estimated, but corresponds to a reduction by 6.6 percentage points. The 2SLS estimates for fair/bad health and obesity for the 9-year reform are all much smaller in magnitude, of opposite sign and insignificant. The impact on smoking is positive but small and statistically insignificant for both reforms. The unscaled difference between the 2SLS-estimates of the two reforms, δ ,

³⁰We cannot rule out that these estimates are affected by reform effects on reporting bias. However, the literature on the relationship between education and bias in self-reported health is mixed and does not provide any clear evidence (Humphries and Van Doorslaer, 2000; Van Doorslaer and Gerdtham, 2003; Bago d'Uva et al., 2008; Doiron et al., 2015). Specifically, Van Doorslaer and Gerdtham (2003) use a subset of the survey data in our paper showing that the association between self-assessed health and mortality risk does not depend on indicators of socioeconomic status like income or education.

is positive and large for self-reported health bad/fair and obesity. Whereas δ is only significant at the 10% level for the general health measure, the *Overid* test rejects the null for both outcomes; the estimate for self-reported health is significant at the 5 per cent level. These results suggest that the overall impact of de-tracking and peer group composition on self reported-health has worked against the health improving increase in years of education element of the 9-year school reform. Notably, the difference in obesity between the reforms corresponds to the above noted reduction of cardiovascular and (marginally significant) diabetes related mortality of the 8-year reform. The lack of a difference between the effect on smoking is also in line with a very similar effect on lung cancer in Figure 5.

In column (4) we examine the reform effects on anxiety. As suggested by Böckerman et al. (2019) and the work by Cicala et al. (2017) changed student ability ranking in a group could harm psychological health. The reduced form and 2SLS estimates however, show an insignificant average negative effect of both the 8-year and the 9-year reforms (and larger for the latter).

In appendix Table C.6 we examine the reform effects on long-term sick leave using government paid sick leave administrative data for spells lasting more than 14 days (short-term sick leave is paid by the employer for the first 14 days). The results show small effects with similar sign as to the mortality effects but all coefficients are insignificant.

Table 5: Education Effects on Self-Reported Health and Health Behaviours

	Fair/bad health (1)	Smoking (2)	Obesity (3)	Anxiety (4)
RF 8-year reform	-0.022 (0.014)	-0.001 (0.017)	-0.027 (0.017)	-0.007 (0.018)
RF 9-year reform	0.016 (0.014)	0.001 (0.015)	0.011 (0.015)	-0.014 (0.016)
IV 8-year reform	-0.062 (0.041)	-0.001 (0.048)	-0.066 (0.046)	-0.018 (0.043)
IV 9-year reform	0.033 (0.030)	0.001 (0.030)	0.022 (0.032)	-0.027 (0.032)
δ	0.048* (0.029)	0.001 (0.031)	0.045 (0.044)	-0.005 (0.031)
Overid	0.048	0.958	0.086	0.862
Mean	0.232	0.236	0.110	0.148
N	24,038	23,880	13,997	16,256

Notes: This table presents the impact of the reforms and years of education on self-reported health and health behaviours. The sample is cohorts born 1932-1952. All coefficients are from separate regressions (RF=Reduced form; IV=Two Stage Least Square). IV estimates include the relevant reform as an excluded instrument and are based on DiD with cohort, municipality fixed effects and municipality specific trends. Standard errors clustered at the municipality level in parenthesis.

"Overid" is the p-value from a Hansen J-statistic testing the validity of the instruments when both are used simultaneously as instruments for years of schooling in the same regression (Baum et al., 2003). δ is our measure of peer effects calculated as in equation 14 and standard errors are from a parametric bootstrap accounting for clustering with 10,000 repetitions and 0.25% of the most extreme values are Winsorized. We use a Winsorized bootstrap in order to align the standard error based 95% Confidence interval estimate with the robust 95% Percentile Confidence Interval. Doing this avoids the adverse influence of extreme bootstrapped values obtained from bootstrapping a relatively small sample size, where a few clusters have only a few sample observations which occasionally have undue influence for particular simulation draws. Testing the null hypothesis: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: ULF-Survey. Own calculations.

6.4 Sensitivity Analysis

To explore the robustness of our findings we have performed several sensitivity checks. Table 6 shows reduced-form estimates using RDD estimated using a quadratic polynomial in age and month as the forcing variable estimated either side of the cut-off, plus month of birth and gender dummies and find that the coefficients are stable to this alternative identification strategy. This is important as the identifying assumption of RDD (that the functional form of the relationship between the outcome variable and the forcing variable is adequately captured) is different to that of DiD yet the results are very similar. As a consequence of a different identifying assumption, RDD is not subject to problems related to DiD and two-way fixed effects models with staggered implementations (Athey and Imbens, 2018; de Chaisemartin and d'Haultfoeuille, 2019; Goodman-Bacon, 2018; Strezhnev, 2018). Together, our RDD results and event study figures support our main DiD modelling specification and its identifying assumption of time-invariant treatment effects. An alternative DiD specification with quadratic trends also finds results similar to our main specification. The results for the 8-year reform indicate a statistically significant health improving impact around -0.4 percent-

age points, and in general strikingly similar to the main estimates. Estimates for the 9-year reform are nearly always small, positive and insignificant.

Table 6: Sensitivity Analysis of First Stage and Reduced Form Estimates

	Education (1)	Mortality (2)	Fair/bad health (3)	Smoking (4)	Obesity (5)	Anxiety (6)
WHOLE SAMPLE						
<i>8-Year Reform</i>	0.277*** (0.021)	-0.004** (0.002)	-0.022 (0.014)	-0.001 (0.017)	-0.027 (0.017)	-0.007 (0.018)
<i>9-Year Reform</i>	0.542*** (0.024)	0.002 (0.002)	0.016 (0.014)	0.001 (0.015)	0.011 (0.015)	-0.014 (0.016)
N	1,505,957	1,505,957	24,038	23,880	13,997	16,256
WHOLE SAMPLE, REGRESSION DISCONTINUITY DESIGN						
<i>8-Year Reform</i>	0.253*** (0.018)	-0.005** (0.002)	-0.019 (0.014)	-0.007 (0.015)	-0.021 (0.014)	-0.008 (0.017)
N	794,596	794,596	24,039	23,881	14,008	16,261
<i>9-Year Reform</i>	0.548*** (0.018)	0.001 (0.002)	0.002 (0.014)	-0.009 (0.015)	0.001 (0.011)	-0.026* (0.016)
N	1,505,957	1,505,957	24,039	23,881	14,008	16,261
WHOLE SAMPLE, DiD WITH QUADRATIC TRENDS						
<i>8-Year Reform</i>	0.277*** (0.021)	-0.004** (0.002)	-0.022 (0.014)	-0.001 (0.017)	-0.027 (0.017)	-0.007 (0.018)
<i>9-Year Reform</i>	0.542*** (0.024)	0.002 (0.002)	0.016 (0.014)	0.001 (0.015)	0.011 (0.015)	-0.014 (0.016)
N	1,505,957	1,505,957	24,038	23,880	13,997	16,256
MUNICIPALITIES THAT IMPLEMENTED BOTH REFORMS						
<i>8-Year Reform</i>	0.262*** (0.020)	-0.004** (0.002)	-0.021 (0.014)	0.002 (0.018)	-0.024 (0.017)	-0.006 (0.018)
<i>9-Year Reform</i>	0.415*** (0.024)	0.000 (0.002)	0.019 (0.018)	0.010 (0.021)	0.024 (0.020)	-0.029 (0.023)
N	794,596	794,596	12,556	12,485	7,297	8,462
COHORTS IMPACTED BY BOTH REFORMS (BORN BEFORE 1949)						
<i>8-Year Reform</i>	0.233*** (0.019)	-0.004** (0.002)	-0.030** (0.015)	-0.005 (0.019)	-0.029 (0.018)	-0.021 (0.020)
<i>9-Year Reform</i>	0.574*** (0.039)	0.001 (0.002)	-0.000 (0.023)	-0.021 (0.023)	-0.007 (0.026)	0.001 (0.021)
N	1,180,067	1,180,067	18,944	18,815	11,110	12,891

Notes: Testing the null hypothesis: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

Source: SIP (Education, mortality) and ULF-Survey (all other outcomes). Own calculations.

The same table reports the sensitivity of our results by limiting the analysis to municipalities implementing both reforms, and by restricting the sample to cohorts impacted by reforms, born before 1949 (see Figure 2). These restrictions do not make a major difference to our baseline findings. In line with these results, Figure C.3 in the appendix, shows that 2SLS coefficients for sub-samples of municipalities from a moving window based on reform time are all in the same range. The 2SLS estimate from the main sample is included in the confidence interval for 10 out of 11 cases. These results are reassuring for our interpretation of the difference between the reforms

which relies on homogeneity in the effect of schooling between the reforms. Indeed, our main motivation for comparing the two reforms is that they are implemented for an overlapping period of time affecting similar cohorts who were faced with very similar conditions.

7 Discussion

We have considered two major parallel school reforms and used a unique quasi-experimental set-up that allows us to tease apart years of schooling effects from tracking and changes in peer group composition effects on health outcomes and health behaviours.

The clean analysis and comparisons of the two reforms and the large and detailed data are key strengths of our approach yielding important results with strong internal validity, but how relevant are the findings outside of the Swedish context? Medical services play an important part in determining health outcomes and education can impact this health channel through its impact on ability to pay for medical services. We are unable to assess the importance of this channel, but for many European countries like Sweden that grant access to health care at the point of need paid for through general taxation, this channel is not important. Instead our results speak to education's impact on individual health behaviours and investments and this is likely to be of relevance for a wide set of institutional and culturing settings. Finally, we argue that the labour market is unlikely to be an important mediator of the effects we identify; the effects on earnings are small for both reforms (see appendix Table C.6). This possibly represents a limitation regarding the external validity of our findings: the cohorts we study were exposed to a labour market characterised by full employment and a very compressed wage structure (Bhalotra et al., 2018). In a different context, the reforms we study here could have had different and larger effects on labour market outcomes.

Our analysis generates two main insights, both of high policy relevance. We will now discuss each one of them in turn.

7.1 The Health Effects of Extended Compulsory Schooling

First, the empirical analysis shows that there was a positive long-run health effect from the 8-year reform. Using the reform as an instrument for years of schooling, the results suggest that longevity increased following the quantitative change in education, and there are indications of health gains also when using alternative health measures like self-assessed health and obesity. As discussed above, the large literature using education policy reforms to isolate the causal health effects of education produces mixed evidence (Cutler and Lleras-Muney, 2012; Galama et al., 2018). However, it is rare to have an education reform that only introduced one additional year of schooling without changing other components that could affect the production and protection of health. The significance of our finding lies in the fact that we have a pure year extension (and that reform assignment is thus a credible instrument for years of schooling), but also that the reform was not introduced overnight.

The existing research on the health effects of more education that most closely aligns to that of ours is the population based research of Clark and Royer (2013) for Britain and that of van Kippersluis et al. (2011) for the Netherlands. Our finding of a significant reduction in mortality by 1.5 percentage points is smaller than the 2.5 percentage point decrease in mortality observed from the Dutch reform, but larger than the potential reduction recorded after the British reforms.³¹ For both studies there may be concerns that the reforms were implemented overnight nationwide. Such overnight nationwide reforms can potentially create temporary labour supply shocks or affect educational quality due to resource limitations. They therefore may not be capturing only the increased years of schooling impact.³² The cohorts affected by the Dutch reform and the 1947 UK reform were also both impacted by the World Wars which may limit their external validity. This is less of a concern in our setting, as Sweden was neutral in the wars.³³

Thus, our estimate of a 1.5 p.p. reduction in mortality per year of schooling appears plausible. As Figure 7 illustrates, it also aligns well with estimates for the same population using alternative estimation strategies. In the figure, we use our main analysis sample and estimate the effect of an additional year of education using OLS, sibling fixed effects, twin fixed effects, same-sex twin fixed effects -- and the 8-year reform. Each point estimate is in the interval -1 to -1.5, which suggests there is a great degree of validity outside the population of compliers to this reform.

³¹Direct comparison between coefficients should be done with caution as the studies observe mortality at different age span. van Kippersluis et al. (2011) observe mortality between 81 to 87 years of age, conditional on survival to 80. Clark and Royer (2013) observe mortality up to between 36 and 74 for the 1947 reform and between 12 to 50 for the 1972 reform. Notably, Clark and Royer (2013) rule out a mortality reducing effect larger than 1% from an additional year of schooling, while Davies et al. (2018) find a significant decrease in the risk of death due to the second reform for individuals in the UK biobank being observed between 42 and 62 years. Related to our results, Davies et al. (2018) also finds effects on several health behaviour and life-style related diseases. Using the same reform to study chronic conditions, Janke et al. (2020) finds no effect on a range of outcomes except for diabetes.

³²Our estimates will likely suffer to less of a degree from bottlenecks related to school inputs and more immediate labour supply effects because the reform was rolled-out over time and space. Major education reforms affecting full cohorts simultaneously may imply that wider changes in the economy start to ensue, e.g. affecting skill prices, educational resources per student. It seems reasonable that effects channelled through educational resources are less likely with a progressive roll out in small geographical units compared to when reforms are implemented overnight. A similar argument can be made regarding the labour market. A reform that is introduced overnight with a whole cohort getting more years of schooling may generate a major shift in supply of more educated workers that have wage effects already from the first year of labour market participation of affected cohorts.

³³The cohorts studied in this paper were born across the span 1938--1954. Like the cohorts exposed to the school reform 1947 in Clark and Royer (2013), some of the cohorts in our study were impacted by the Second World War (WWII). Using Sweden as our study case has the advantage that, although WWII was an unusual period, Sweden was a neutral part and as such much less affected than other countries. There are also historical sources supporting that the war did not cause any educational disruptions (Bhalotra et al. 2018 discuss that schools could cancel classes in case of a threat during the WWII, but any time lost had to be caught up later on and that a substitute teacher had to be called in if a teacher was called for military service), and there were no policies to move children out of the city to extended family or volunteers on the countryside, as in Britain. We argue therefore that our results should not be specific to the cohorts we consider, but instead being of relevance beyond the study period.

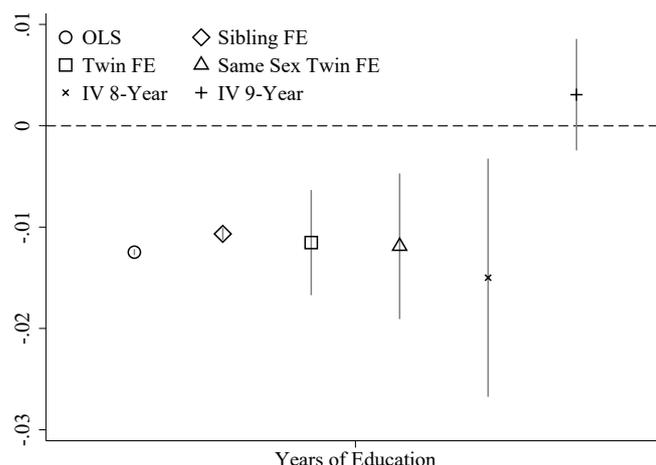


Fig. 7: Years of education estimates compared across estimation strategies

Notes: The graph plots estimates of the effect of years of schooling on mortality from various estimation strategies for the same sample. The 2SLS estimates from section 6.1 are compared to estimate using OLS, sibling fixed effects, twin fixed effects, and same sex twin fixed effects. Estimations for OLS, sibling and twin FE are based on the same sample. We identify siblings and twins by the biological mother from the multi-generational register. Notably, the sibling/twin fixed effect estimates are comparable to the estimates in Lundborg et al. (2016) who use a sample of about 50,000 Swedish twins born 1886-1958. Source: SIP. Own calculations.

7.2 Estimating the Impact of De-Tracking

The striking exception in Figure 7 is the IV estimate for the 9-year reform, which is positive and insignificant. Which leads us to the second main contribution of this paper: that de-tracking and peer composition may have important long-term effects on health. Several articles have examined the 9-year reform under study in this paper (Meghir et al. 2018; Palme and Simeonova 2015; Lager and Torssander 2012; Lager et al. 2016; Spasojevic 2010) but the impact of different components of schooling affected by the reform have not previously been disentangled. We have put forward the argument that the parallel implementation of the two reforms can be used to back out the pure years-of-schooling effect, which leaves a residual effect which may be attributed to other reform components. Using this approach, we estimated that de-tracking is potentially responsible for a 1 percentage point increase in mortality, which is also reflected in an increased propensity to report fair/poor health and obesity.

We recommend some caution in the interpretation of this result, since there are a number of reasons why our estimate of the effects of de-tracking may be biased. We reach our conclusion after having considered a number of alternative mechanisms and, if possible, tested their empirical relevance. One obvious such candidate would be resources, to the extent that the two reforms had different effects on the resources expended per pupil-year. However, we have provided suggestive evidence that both reforms lead to an increase in pupil/teacher ratios, but the impact of the 8-year reform is larger. This suggests that our estimated effect of de-tracking is if anything downward biased as a result of the difference in change of resources. We also provide results on the effects of the two reforms on lifetime earnings (appendix Table C.6). In line with the general conclusion in the literature, the returns to compulsory schooling are moderate

for both reforms, and certainly not large enough to be realistically considered as an important channel for the different health effects we observe between the reforms.

An important assumption we make in our interpretation of the de-tracking effects is that the relationship between health and years of education is linear. There are good reasons to question this assumption since we would expect, in general, that there are diminishing returns to education. We have shown that for the levels of education driving our results, it is indeed approximately linear. Figures C.1 and C.2 show that our IV estimates put weight in the linear segment of the conditional expectation function of education and mortality. Nevertheless, given that linearity is possibly the most restrictive assumption we impose, we have conducted a sensitivity analysis, considering how diminishing returns to education in the health domain would affect our conclusions. Results from progressively increasing the assumed deviation from linearity are presented in Figure 8. Even though δ , the harmful effect of de-tracking, is mechanically diminished by alterations of this assumption, we would reach a qualitatively very similar conclusion even under strong deviations from linearity. Only a 9th-year effect smaller than 20 per cent of the 8th-year effect would lead to the conclusion that other reform components had no harmful effects on health.³⁴

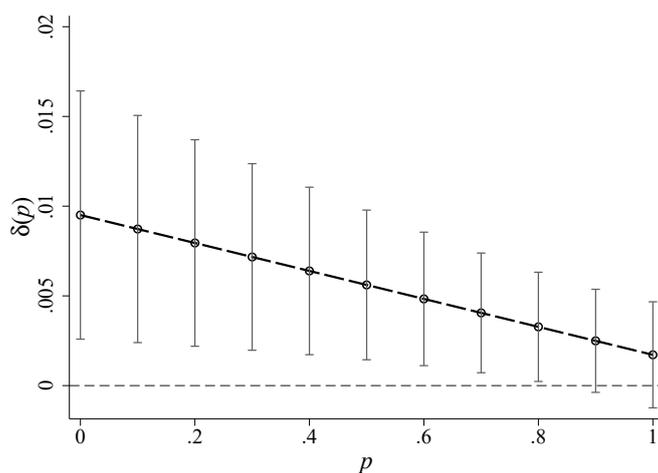


Fig. 8: Sensitivity of δ to the Years of Education Effect Derived From the 8-year reform

Notes: δ is calculated as $[\beta_{IV9} - (1 - p)\beta_{IV8}] \cdot \gamma_{FS9}$ for different values of p . Thus, we downscale the pure years-of-schooling effect by 10% increments down to 0% of the original estimate to show how sensitive our estimate of δ is to linearity assumption. Note that for $p = 1$, δ equals the reduced-form estimate of the 9-year reform presented in Table 4. As in the main analysis standard errors are obtained from a parametric bootstrap of 999 repetitions clustering at the municipality level.

Source: SIP. Own calculations.

If the mortality penalty that δ picks up is indeed related to harmful effects of peer group composition, we would expect the reform effects to depend on the peer composition in

³⁴A related issue that has been discussed in the literature is that education's effect on health diminishes with time or age (Galama et al., 2018, cf.). Since we are studying two parallel reforms, this is less likely to be a concern here. The fact that our results are robust to using only cohorts or municipalities impacted by both reforms, and further that compliers are similar across reforms, should eliminate concerns that this issue biases our results.

the municipality. Since our analysis includes a large number of reform municipalities, we are able to test this empirically. In appendix figure C.4 we let the reform effects depend on the proportion of pupils in a municipality who take the old secondary school form before the reform. It shows that in municipalities with a larger share of pre-reform attendance of the academic track, there is a smaller increase in education (as expected as more students already study 9 or more years) and positive effects on mortality but also that the mortality effects tail off for the most homogeneous municipalities. This provides some evidence that heterogeneity in peer group composition matters for long-term health outcomes. By contrast, the education effect is larger and the mortality effect is negative in municipalities with a lower share of pre-reform realskola attendance. There is no such pattern for the 8-year reform. These results are consistent with the change in peer composition reversing any positive health effects of education, at least in municipalities with above-average heterogeneity.

The work most closely aligned with the peer effect interpretation of our results are the results from the school reform evidence literature of tracking. Basu et al. (2018) find a fraction of individuals' health and health related behaviours are negatively impacted by the removal of tracking. However, because their paper had a greater emphasis on assessing heterogeneous effects their results rely on a weaker and less convincing identification strategy compared to ours. Also, the work by Böckerman et al. (2019) and Ravesteijn et al. (2017) on the removal of tracking in Finland suggest worsened mental health and increased mortality for some subgroups.

Notably, our results suggest—at least for the more general health measures—that de-tracking and a broader set of peers have an adverse effect on health for both high and low SES groups. If we are willing to accept an interpretation where our estimates for the effects of de-tracking (δ) represent peer effects, then the possibilities for how peer effects operate and the mechanisms at play narrow down quite substantially.³⁵ Our results are consistent with the “boutique” and “focus” models, according to which classroom heterogeneity is harmful to everyone (Sacerdote et al., 2011).³⁶ For example, mixing can lead to less optimal education outcomes across types as the learning environment better satisfies students' type if the group is homogeneous. It becomes more difficult for teachers to customize pace of learning, and tailor teaching materials and/or pedagogic approaches. It could also be that peer learning (where students learn from each other)

³⁵Notably a deeper understanding of the mechanisms at play would require a different and more sophisticated empirical approach specifically allowing to identify peers and their effects more directly. Also, the possibility to interpret noted peer effects based on sub-group analyses are reduced in case of misclassification, see section 3.3.

³⁶In a review of the literature, Sacerdote et al. (2011) lists six additional models of peer effects. Our empirical results are not consistent with any of these. The already-mentioned “linear-in-means” model would predict that the reform had no noteworthy effects in the aggregate, and that the low-SES group benefits from the new peer group composition. The “shining light” model predicts that all students benefit from having high-ability students around, but in our case no students benefit. According to the “invidious comparison” model, outcomes are harmed by the presence of better-achieving peers. This appears to indeed have happened in the low-SES group, but it is inconsistent with the adverse effects in the high-SES group. The “rainbow model” suggesting that heterogeneity is good for everyone is obviously inconsistent with our results, as well as the “single-crossing” model which would suggest that low-SES students benefit. Finally, the results are inconsistent with the “bad apple” model, which suggests that one disruptive student harms everyone -- since we would expect low-SES students to be exposed to fewer bad apples in the de-tracked system.

only takes place when students are of similar background/ability. The mixing of types may thus decrease the efficiency of the school system to generate human capital. Assuming that human capital is good for health (which our evidence for the 8-year reform suggests), a broader set of peers could this way generate worse human capital and health outcomes across all types, as compared to when students are tracked.

Also, the results are at odds with the interpretation provided by Böckerman et al. (2019) who argue that the removal of tracking affects the ability ranking in the classroom, with low-ability students that did not take on academic tracking in the old system having a lower rank in the new system. This interpretation is well in line with Cicala et al. (2017) and Elsner and Isphording (2018) who report that the ordinal rank among peers affect academic achievement but also problem behaviour and engagement in risky behaviour. However, it is difficult to reconcile a story based on rank with high-SES students suffering health penalties when tracking is removed. In fact, δ is slightly larger for the high-SES group than the low-SES group in Table 4 (though the difference is not statistically significant). This could suggest that on top of the efficiency argument there may be an additional penalty for high-SES students. Previous research on the 9-year reform has documented a loss in emotional control for boys from high socio-economic backgrounds (Lager et al., 2016). Low emotional control may affect risk taking and has proven important for later life health risks (Moffitt et al., 2011).

It has often been suggested that increasing the mix of high-ability and high-SES students with lower ability and lower SES students could help reduce inequalities in a number of long-term outcomes, yet the evidence from these natural experiments on long-term health outcomes does not support this. Our finding is novel and suggests that changes in compulsory school peers based on their SES background may have a long-lasting impact on health behaviour or actual health. Importantly, these results do not rule out positive effects of de-tracking -- either in terms of efficiency or improved equality of opportunity -- but it is difficult to argue that abolishment of de-tracking is instrumental in order to achieve health improvements.

The analysis included in this paper has considered two Swedish school reforms that were rolled out during the same time period, ensuring that the cohorts exposed to the two reforms later acted under similar welfare and labour market institutions. The way the reforms were rolled out also means that concerns about resource shocks such as lack of teachers and schools and even immediate labour supply shocks that may change skill prices are of less concern.

8 Conclusion

The literature documenting an education gradient in health is vast yet the causal effect literature using compulsory school reforms as instruments for years of education has produced varied results. Compulsory school reforms are indeed a powerful way to assess the causal impact of education on health. They often impact a large population and are often directly targeting sub-populations of specific public health interest, the low educated who live on average shorter lives and have poorer health outcomes generally. More importantly, these reforms can provide exogenous variation of exposure to education, and under certain parametric assumptions also variation in years of

education. However, very few of these reforms actually satisfies these assumptions as the implementation has often mixed increases in mandatory schooling with changes in school structure.

Our quasi-experimental setting provides us with one reform that only changes the quantity aspects of education, and one reform with a similar change in quantity accompanied by changes in quality aspects (de-tracking and peer exposure). Combining this set-up with detailed data from surveys and registers, we are able to provide important—and policy-relevant—insights on both quantity and quality aspects of education. Our findings suggest that the increasing quantity of education improves both objective and subjective health outcomes, but such health improvements may be undone by increased heterogeneity in the classroom.

Even though our study goes a long way toward unpacking the health effects of comprehensive schooling reforms, we are unable to give a detailed answer as to the mechanisms giving rise to adverse health effects from de-tracking. We are able to show that our results are inconsistent with a majority of popular approaches to model peer effects; however, it clearly has to be left as a topic for future research to determine whether long-term health penalties from such reforms arise from peer interactions or from other aspects of the learning situation.

References

- Angrist, J. and Fernandez-Val, I. (2010). Extrapolate-ing: External validity and overidentification in the late framework. Technical report, National Bureau of Economic Research.
- Angrist, J. D. and Pischke, J.-S. (2008). *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.
- Athey, S. and Imbens, G. W. (2018). Design-based analysis in difference-in-differences settings with staggered adoption. Technical report, National Bureau of Economic Research.
- Bago d'Uva, T., Van Doorslaer, E., Lindeboom, M., and O'Donnell, O. (2008). Does reporting heterogeneity bias the measurement of health disparities? *Health Economics*, 17(3):351--375.
- Basu, A., Jones, A. M., and Dias, P. R. (2018). Heterogeneity in the impact of type of schooling on adult health and lifestyle. *Journal of Health Economics*, 57:1--14.
- Baum, C. F., Schaffer, M. E., and Stillman, S. (2003). Instrumental variables and gmm: Estimation and testing. *The Stata Journal*, 3(1):1--31.
- Benjamini, Y., Krieger, A. M., and Yekutieli, D. (2006). Adaptive linear step-up procedures that control the false discovery rate. *Biometrika*, 93(3):491--507.
- Bertoni, M., Brunello, G., and Cappellari, L. (2017). Parents, siblings and schoolmates: The effects of family-school interactions on educational achievement and long-term labor market outcomes.
- Bhalotra, S., Karlsson, M., and Nilsson, T. (2017). Infant health and longevity: Evidence from a historical intervention in sweden. *Journal of the European Economic Association*, 15(5):1101--1157.
- Bhalotra, S. R., Karlsson, M., Nilsson, T., and Schwarz, N. (2018). Infant health, cognitive performance and earnings: Evidence from inception of the welfare state in sweden. *HEDG Working Paper 18/6*. University of York.
- Björklund, A., Edin, P.-A., Fredriksson, P., and Krueger, A. (2004). *Education, equality, and efficiency: An analysis of Swedish school reforms during the 1990s*. Institutet för arbetsmarknadspolitisk utvärdering (IFAU).
- Böckerman, P., Mika, H., Christopher, J., and Alexandra, R. (2019). School tracking and mental health. *IZA Working Paper*.
- Booij, A. S., Leuven, E., and Oosterbeek, H. (2017). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies*, 84(2):547--578.
- Brodaty, T. and Gurgand, M. (2016). Good peers or good teachers? evidence from a french university. *Economics of Education Review*, 54:62--78.

- Brunello, G., Fort, M., Schneeweis, N., and Winter-Ebmer, R. (2015). The causal effect of education on health: What is the role of health behaviors? *Health Economics*.
- Burke, M. A. and Sass, T. R. (2013). Classroom peer effects and student achievement. *Journal of Labor Economics*, 31(1):51--82.
- Cicala, S., Fryer, R. G., and Spenkuch, J. L. (2017). Self-selection and comparative advantage in social interactions. *Journal of the European Economic Association*, 16(4):983--1020.
- Clark, D. and Royer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *The American Economic Review*, 103(6):2087--2120.
- Cutler, D. M. and Lleras-Muney, A. (2006). Education and health: evaluating theories and evidence. *National Bureau of Economic Research*, (No. w12352).
- Cutler, D. M. and Lleras-Muney, A. (2012). Education and health: insights from international comparisons. *National Bureau of Economic Research*, (No. w17738).
- Davies, N. M., Dickson, M., Smith, G. D., Van Den Berg, G. J., and Windmeijer, F. (2018). The causal effects of education on health outcomes in the UK Biobank. *Nature Human Behaviour*, 2(2):117--125.
- de Chaisemartin, C. and d'Haultfoeuille, X. (2019). Two-way fixed effects estimators with heterogeneous treatment effects. Technical report, National Bureau of Economic Research.
- Doiron, D., Fiebig, D. G., Johar, M., and Suziedelyte, A. (2015). Does self-assessed health measure health? *Applied Economics*, 47(2):180--194.
- Ecklesiastikdepartementet (1935). *Betänkande och förslag angående obligatorisk sjuårig folkskola*, SOU 1935:58. Ivar Hagströms Boktryckeri A.B.
- Edgren, H. (2011). *Folkskolan och grundskolan*. In Larsson, E. & Westberg, J. (Ed.) *Utbildningshistoria*, Studentlitteratur. Lund.
- Elsner, B. and Ispording, I. E. (2018). Rank, sex, drugs, and crime. *Journal of Human Resources*, 53(2):356--381.
- Enflo, K. and Rosés, J. R. (2015). Coping with regional inequality in Sweden: structural change, migrations, and policy, 1860--2000. *The Economic History Review*, 68(1):191--217.
- Fischer, M., Heckley, G., Karlsson, M., and Nilsson, T. (2020). Education reform and equality of opportunity, revisited. *Mimeo*.
- Fischer, M., Karlsson, M., and Nilsson, T. (2013). Effects of compulsory schooling on mortality: evidence from Sweden. *International Journal of Environmental Research and Public Health*, 10(8):3596--3618.
- Fischer, M., Karlsson, M., Nilsson, T., and Schwarz, N. (2019). The long-term effects of long terms: Compulsory schooling reforms in Sweden. *Forthcoming in Journal of the European Economic Association*.

- Folkskolläraryrörbundet, S. (1943). Folkskolans årsbok 1943. *Stockholm*.
- Folkskolläraryrörbundet, S. (1949). Folkskolans årsbok 1949. *Stockholm*.
- Folkskolläraryrörbundet, S. (1952). Folkskolans årsbok 1952. *Stockholm*.
- Fredriksson, V. A. (1950). *Svenska folkskolans historia*, volume 5. Albert Bonniers förlag, Stockholm.
- Fredriksson, V. A. (1971). *Svenska folkskolans historia*, volume 6. Albert Bonniers förlag, Stockholm.
- Galama, T. J., Lleras-Muney, A., and van Kippersluis, H. (2018). The effect of education on health and mortality: A review of experimental and quasi-experimental evidence. Technical report, National Bureau of Economic Research.
- Gaviria, A. and Raphael, S. (2001). School-based peer effects and juvenile behavior. *Review of Economics and Statistics*, 83(2):257--268.
- Goodman-Bacon, A. (2018). Difference-in-differences with variation in treatment timing. Technical report, National Bureau of Economic Research.
- Grenet, J. (2013). Is extending compulsory schooling alone enough to raise earnings? evidence from french and british compulsory schooling laws*. *The Scandinavian Journal of Economics*, 115(1):176--210.
- Grossman, M. (2006). Education and nonmarket outcomes. *Handbook of the Economics of Education*, 1:577--633.
- Grossman, M. (2015). The relationship between health and schooling: What's new? *Nordic Journal of Health Economics*, 3(1):1--7.
- Hjalmarsson, R., Holmlund, H., and Lindquist, M. J. (2015). The effect of education on criminal convictions and incarceration: Causal evidence from micro-data. *The Economic Journal*.
- Holmlund, H. (2008). A researcher's guide to the swedish compulsory school reform. Technical report, Centre for the Economics of Education, London School of Economics and Political Science.
- Holmlund, H., Lindahl, M., and Plug, E. (2011). The causal effect of parents' schooling on children's schooling: A comparison of estimation methods. *Journal of Economic Literature*, 49(3):615--651.
- Humphries, K. H. and Van Doorslaer, E. (2000). Income-related health inequality in canada. *Social Science & Medicine*, 50(5):663--671.
- Husén, T. and Härnqvist, K. (2000). *Begåvningsreserven: en återblick på ett halvsekels forskning och debatt*. Fören. för svensk undervisningshistoria.
- Janke, K., Johnston, D. W., Propper, C., and Shields, M. A. (2020). The causal effect of education on chronic health conditions in the uk. *Journal of Health Economics*, 70:102252.

- Jones, A. M., Rice, N., and Dias, P. R. (2012). Quality of schooling and inequality of opportunity in health. *Empirical Economics*, 42(2):369--394.
- Jones, A. M., Rice, N., and Rosa Dias, P. (2011). Long-term effects of school quality on health and lifestyle: Evidence from comprehensive schooling reforms in England. *Journal of Human Capital*, 5(3):342--376.
- Jones, A. M., Roemer, J. E., and Dias, P. R. (2014). Equalising opportunities in health through educational policy. *Social Choice and Welfare*, 43(3):521--545.
- Jonsson, J. O. (1991). *School reforms, educational expansion, and educational attainment: trends towards equality in Sweden*. Univ., Institutet för social forskning.
- Jürges, H., Reinhold, S., and Salm, M. (2011). Does schooling affect health behavior? evidence from the educational expansion in western Germany. *Economics of Education Review*, 30(5):862--872.
- Kirdar, M. G., Dayioğlu, M., and Koc, I. (2015). *Does longer compulsory education equalize schooling by gender and rural/urban residence?* The World Bank.
- Lager, A., Seblova, D., Falkstedt, D., and Lövdén, M. (2016). Cognitive and emotional outcomes after prolonged education: a quasi-experiment on 320 182 Swedish boys. *International Journal of Epidemiology*, 46(1):303--311.
- Lager, A. C. J. and Torssander, J. (2012). Causal effect of education on mortality in a quasi-experiment on 1.2 million Swedes. *Proceedings of the National Academy of Sciences*, 109(22):8461--8466.
- Larsson, E. (2011). *Utbildning och social klass*. In Larsson, E. & Westberg, J. (Ed.) *Utbildningshistoria*, Studentlitteratur. Lund.
- Lindahl, M. (2005). Estimating the effect of income on health and mortality using lottery prizes as an exogenous source of variation in income. *Journal of Human Resources*, 40(1):144--168.
- Lindmark, D. (2015). *Hemundervisning och läskunnighet*. Studentlitteratur. Lund.
- Løken, K. V., Mogstad, M., and Wiswall, M. (2012). What linear estimators miss: The effects of family income on child outcomes. *American Economic Journal: Applied Economics*, 4(2):1--35.
- Lundborg, P., Lyttkens, C. H., and Nystedt, P. (2016). The effect of schooling on mortality: New evidence from 50,000 Swedish twins. *Demography*, 53(4):1135--1168.
- Lundborg, P. and Majlesi, K. (2018). Intergenerational transmission of human capital: Is it a one-way street? *Journal of Health Economics*, 57:206--220.
- Lundborg, P., Majlesi, K., E Black, S., and Devereux, P. J. (2018). Learning to take risks?: The effect of education on risk-taking in financial markets. *Review of Finance*.
- Lundborg, P., Nilsson, A., and Rooth, D.-O. (2014). Parental education and offspring outcomes: evidence from the Swedish compulsory school reform. *American Economic Journal: Applied Economics*, 6(1):253--278.

- Marklund, S. (1982). Från reform till reform: Skolsverige 1950--1975, del 2 försöksverksamheten.
- Marklund, S. (1989). *Skolsverige 1950-1975: Rullande reform*. Liber/Utbildningsförl.
- Mazumder, B. (2012). The effects of education on health and mortality. *Nordic Economic Policy Review*, 1(2012):261--301.
- Meghir, C. and Palme, M. (2005). Educational reform, ability, and family background. *American Economic Review*, pages 414--424.
- Meghir, C., Palme, M., and Schnabel, M. (2012). The effect of education policy on crime: an intergenerational perspective. Technical report, National Bureau of Economic Research.
- Meghir, C., Palme, M., and Simeonova, E. (2018). Education and mortality: Evidence from a social experiment. *American Economic Journal: Applied Economics*, 10(2):234--56.
- Moffitt, T. E., Arseneault, L., Belsky, D., Dickson, N., Hancox, R. J., Harrington, H., Houts, R., Poulton, R., Roberts, B. W., Ross, S., et al. (2011). A gradient of childhood self-control predicts health, wealth, and public safety. *Proceedings of the National Academy of Sciences*, 108(7):2693--2698.
- Morawski, J. (2010). *Mellan frihet och kontroll: Om läroplanskonstruktioner i svensk skola*. PhD thesis, Örebro universitet.
- Orring, J., Read, A., et al. (1962). *Comprehensive school and continuation schools in Sweden: a summary of the principal recommendations of the 1957 School Commission*. Kungl. ecklesiastikdepartementet.
- Palme, M. and Simeonova, E. (2015). Does women's education affect breast cancer risk and survival? evidence from a population based social experiment in education. *Journal of Health Economics*, 42:115--124.
- Paulsson, E. (1946). *Om folkskoleväsendets tillstånd och utveckling i Sverige under 1920-och 1930-talen (till omkring år 1938)*. Länstryckeriaktiebolaget.
- Ravesteijn, B., van Kippersluis, H., Avendano, M., Martikainen, P., Vessari, H., and Van Doorslaer, E. (2017). The impact of later tracking on mortality by parental income in finland.
- Richardson, G. (1978). *Svensk skolpolitik 1940-1945*. Liber förlag. Stockholm.
- Richardson, G. (1992). *Ett folk börjar skolan: folkskolån 150 [hundrafemtio] år 1842-1992*. Allmänna förlaget.
- Robalino, J. D. and Macy, M. (2018). Peer effects on adolescent smoking: Are popular teens more influential? *PloS one*, 13(7):e0189360.
- Sacerdote, B. et al. (2011). Peer effects in education: How might they work, how big are they and how much do we know thus far. *Handbook of the Economics of Education*, 3(3):249--277.

- Sargan, J. D. (1958). The estimation of economic relationships using instrumental variables. *Econometrica: Journal of the Econometric Society*, pages 393--415.
- Skatteverket (2007). Sveriges församlingar genom tiderna. Technical report, <https://www.skatteverket.se/folkbokforing/sverigesforsamlingargenomtiderna.4.18e1b10334ebe88bc80>
- Skolöverstyrelsen (1955). *Skolan och de stora årskullarna. Förslag av Skolöverstyrelsens planeringskommitté för de stora årskullarna*. Nordstedts.
- Skott, P. (2011). *Utbildningspolitik och läroplanshistoria*. In Larsson, E. & Westberg, J.(Ed.) *Utbildningshistoria*, Studentlitteratur. Lund.
- SOU (1944). Utredning och förslag. vidgade möjligheter till högre undervisning för landsbygdens ungdom. *Statens Offentliga Utredningar*, 1944:22.
- SOU (1945). Skolplikts tidens skolformer. *Statens offentliga utredningar*, 1945:60.
- SOU (1961). 1957 års skolutredning. *Statens Offentliga Utredningar*, 1961:31.
- SOU 1963:41 (1963). *1960 års gymnasieutredning 3 Specialutredningar om gymnasiet*. Ecklesiastikdep., Stockholm.
- Spasojevic, J. (2010). Effects of education on adult health in sweden: Results from a natural experiment. *Current Issues in Health Economics (Contributions to Economic Analysis, Volume 290)*, Emerald Group Publishing Limited, pages 179--199.
- Statistics Sweden (1945). Riksdagsmannavalen. åren 1941-1944. Stockholm.
- Statistics Sweden (1948). Årsbok för sveriges kommuner. årg. 31 (1948). Stockholm.
- Statistics Sweden (1952). Folkräkningen den 31 december 1950 = [census of the population in 1950]. Stockholm.
- Statistics Sweden (1961). Folkräkningen den 1 november 1960 = census of the population. Stockholm.
- Statistics Sweden (1965). Folk- och bostadsräkningen den 1 november 1960 : redogörelse för folk- och bostadsräkningens uppläggning och utförande ... Stockholm.
- Statistics Sweden (1974). Folk- och bostadsräkningen 1970. d. 12, redogörelse för folk- och bostadsräkningens uppläggning och utförande = [report on the planning and ... Stockholm.
- Statistics Sweden (2008). Undersökning av levnadsförhållanden (ulf). *SCB (www.scb.se)*.
- Stock, J. H., Wright, J. H., and Yogo, M. (2002). A survey of weak instruments and weak identification in generalized method of moments. *Journal of Business & Economic Statistics*, 20(4):518--529.
- Strezhnev, A. (2018). Semiparametric weighting estimators for multi-period difference-in-differences designs. In *Annual Conference of the American Political Science Association, August*, volume 30.

- Sweden, S. (1974). Elever i obligatoriska skolor 1847--1962. *Stockholm: SCB*.
- Sweden, S. (1977). Elever i icke-obligatoriska skolor 1864--1970. *Stockholm: SCB*.
- Van Doorslaer, E. and Gerdtham, U.-G. (2003). Does inequality in self-assessed health predict inequality in survival by income? evidence from swedish data. *Social Science & Medicine*, 57(9):1621--1629.
- van Kippersluis, H., O'Donnell, O., and Van Doorslaer, E. (2011). Long-run returns to education does schooling lead to an extended old age? *Journal of Human Resources*, 46(4):695--721.
- Waldow, F. (2013). Utbildningspolitik, ekonomi och internationella utbildningstrender i sverige 1930--2000. *Liber förlag*.

Appendix A Background to the Swedish school reforms

This appendix provides background information on the Swedish school system, the two school reforms used in this paper and their interrelation.

A.1 *The Swedish School System*

Long before compulsory schooling was introduced by law on a nationwide level in Sweden, a large fraction of the population had basic reading and writing capabilities, and a notable share of all parishes had introduced some kind of primary school on a voluntary basis (Lindmark, 2015). Regulations announced in the mid to late 19th century came to imply both the right to cost-free primary schooling and an obligation to take part in the schooling offered (Fredriksson, 1971). Specifically the 1882 legal statute of *Folkskolan* stated that every parish had to offer primary schooling by an approved teacher, that school attendance was compulsory for all children, and that children should start primary school the year they turn seven years old (Edgren, 2011).

The country was divided into school districts (generally corresponding to a parish, and later a municipality) and the local school board was responsible for the organization of elementary education. To overcome differences in content and format across school districts, a national central education plan was introduced in 1919 (Paulsson, 1946). These guidelines were published by the Ministry of Ecclesiastical Affairs and included time tables and syllabuses for compulsory schooling. The ministry also appointed school inspectors responsible for yearly evaluations of a number of school districts (Fredriksson, 1971). Completion rates were high and more than 90 per cent of all pupils finished compulsory schooling with full curriculum (Fredriksson, 1950).

In the 1920s elementary schooling in Sweden was compulsory for six years, but the central education plan provided curricula also for seven years of schooling and in 1920 a clause was introduced in the primary school code (paragraph 47 mom. 4) that stated a seventh school year could be made compulsory in a school district (Fredriksson, 1950). At the time Sweden applied a tracking system, where good performing students (defined by an assessment) could select to switch to an academic educational track and study at a four or three year long junior secondary school (*Realskola*) after the fourth or the sixth year of elementary education, respectively.³⁷ The alternative was to continue and finish basic compulsory schooling. Attending junior secondary school allowed students to continue to higher secondary school (*Gymnasium*) which was a prerequisite for University. During the first half of the twentieth century, the Swedish school system was highly selective and the vast majority of people only completed compulsory education Björklund et al. (2004).³⁸

³⁷The tracks starting after 4th or 6th grade led to the same degree (*realexamen*) provided the a student passed a written and oral final examination (Marklund, 1989).

³⁸See e.g. Sweden (1977) for yearly numbers of students matriculating to lower and higher secondary schooling. For the cohorts of interest in this study matriculation to lower secondary schooling increased over time (16 percent of cohort 1938 and 30 percent of cohort 1951). In 1942, when the first cohort in our analysis sample had the opportunity to enroll in lower secondary education, 12 per cent of all municipalities had a realskola. But attendance was not restricted only to individuals who lived in a

In 1936 the national Government decided that a seventh year of schooling should be compulsory and seven years of schooling had to be implemented across the whole country over a twelve-year period, before 1949 (Fischer et al., 2013). The reform was seen as a new epoch, especially among teachers, because previously *Folkskola* had remained the same since 1882, offering six years of compulsory education (Folkskolläraryförbund, 1949).³⁹ With the bill of 1936, school districts were also allowed to introduce an eight year of compulsory schooling, but for this they had to send in a formal application to the Ministry of Ecclesiastical Affairs and get an approval.

A.2 *The 8-Year Reform*

With the start of the World War II, the Swedish political debate came to place a large focus on how to best foster democratic members of society. More education was seen as one of the main components for fulfilling this goal (Edgren, 2011). Thus, despite the on-going national implementation of the seventh year of compulsory schooling, a reform work was initiated and assigned to a new expert commission (*Skolutredningen* later replaced by *Skolkommissionen*) in 1940. This was the first governmental commission with a real mission to investigate primary and secondary education together, and with an aim to replace the tracking system with a unified comprehensive school system tying compulsory schooling closer to secondary education.⁴⁰

Between 1940-1948 the commission continuously released reports evaluating the current school system and developing proposals and guiding principles for the future compulsory school (Marklund, 1982). Although the main focus of the commission's work was to postpone tracking decisions to higher grades, thereby improving equality of opportunity, and despite the on-going implementation of the seven-year compulsory schooling, there were also continued efforts to further extend compulsory schooling within the old system. The commission's work considered an eighth year extension of *Folkskola*, and in 1945 the Minister of Ecclesiastical Affairs proposed a bill introducing a compulsory eighth year (without changing tracking options), but no action was taken by the Government (Fredriksson, 1971).

As stated above one of the main arguments for extending compulsory education was *democratic fostering*. This motive was not new. The 7 year reform was also motivated by fostering democracy including universal suffrage which was argued to place great demands on members of society, wherefore a solid education is necessary (Ecklesiastikdepartementet, 1935). The on-going war made this argument even more important in the debate. Specifically an eight-year extension was believed to improve student performance with respect to elementary skills in reading, writing and math,

municipality with a *realskola*. As Fischer et al. (2019) show, more than 50 per cent of high-ability students from rural areas attended *realskola*. In practice many of these students moved away from home. A survey from 1941 show that 17-15 per cent of all students in the most northern counties choose this option (SOU, 1944).

³⁹For a detailed review of the background and the implementation of the seventh compulsory year, see Fischer et al. (2013).

⁴⁰Since the 1890's there had been a quite heated debate about the rational of the the so-called parallel system where student took different tracks. The main argument in the debate for that all students should have to complete the very same basic education before continuing to secondary schooling, was that it created inequalities (Morawski, 2010).

but also other subjects. An extension would also allow for the introduction of foreign languages (English) as a subject. In addition to theoretical arguments an extension was further justified by social and ethical arguments and that an eight-year could fill a supportive and nurturing role for young people that were not established on the labour market (SOU, 1945).

A second argument for extending compulsory education was induced by international *benchmarking*—that Sweden was lagging behind (Waldow, 2013). Compared to other countries few students matriculated to higher levels of education, and the time spent in compulsory education was still quite modest. For example, compulsory education in the US endured at least until age 16, in Germany there were *Volkschule* or *Hauptschule* until the age of 15 and in the UK students generally had nine years of compulsory schooling in the late 1930s.

A third argument for extending compulsory schooling was the *increasing specialization of the labour market* and the *increased complexity* of society and societal life, implying a need to significantly increase educational goals of *Folkskolan*. Finally, the economic and societal *duality* that existed between urban and rural areas was brought forward to motivate a general 8-year reform or a compulsory school reform. Specifically with respect to education, the rural areas of the country were falling behind, e.g. smaller shares of students matriculating to junior secondary education in rural compared to urban areas (Sweden, 1977). With a general implementation of the 8-year reform such differences could decrease.

The main arguments of the proponents of the 8-year reform to why the realization of an eight-year extension was seen as preferred compared to a comprehensive school reform was that there was (i) no large demand from students nor from the parties of the labour market for a 9 year comprehensive reform, and that (ii) the supply of teachers was too limited for a comprehensive reform, but also that the teachers generally had too limited education (SOU, 1945).

Likely spurred by the political debate some municipalities applied and got approval and took the opportunity to implement a mandatory eight-year of *Folkskola* (*Folkskolläraryrörbundet*, 1943).⁴¹ The first two municipalities to implement an eight mandatory year were Kävlinge and Mariestad in the school year of 1941/42. The number of municipalities offering an eighth year gradually increases in the following decade: In 1946/47 there were 33 and in 1958/59 207 municipalities, respectively. A characteristic of the municipalities introducing a mandatory eighth year in this time period is that they were urban and most of the larger cities of Sweden were early birds in this development. Consequently a quite large share of all students in the country had eight years of compulsory schooling: in the school year 1948/49 this was 16 per cent and in the school year 1951/52 this was 25 per cent (*Folkskolläraryrörbundet*, 1952).

All municipalities introducing the eighth year followed the *main form* curriculum requiring full time reading and a teacher with an appropriate teacher degree.⁴² Normative

⁴¹Only in a few cases a municipality did not get the permission to implement the extension. The reason was that the request was to do an isolated change and only introduce the change in a separate school in a municipality (Fredriksson, 1971).

⁴²The alternative to the main form were *exception forms*, characterised by half time reading or that

and binding curricula regarding the eight-year were missing in the early period, but the curriculum and hourly plan presented in the proposal of the expert commission in 1946 generally became the norm for the school districts that implemented an eight-year of *Folkskola*. The mandatory subjects in the eight grade were the same as in seventh grade, but local preferences could to some extent be met (Fredriksson, 1971).

A.3 The 9-Year Reform

In 1948, the expert commission proposed to replace the compulsory primary and the junior secondary school with a nine-year compulsory comprehensive school. The expert commission however wanted to evaluate the new school form before introducing it to all schools across the country. The reform was therefore introduced during an assessment period where the 9 year comprehensive school was introduced in different locations at different points in time. Starting from 1949/1950 the 9-year reform was rolled out at the municipality level.⁴³ In the first year of the roll-out of the reform 14 municipalities were selected to participate in the assessment.⁴⁴ The evaluation period was not run as a random experiment, but the National School Board chose the areas from a group of applicants to form a representative set based on observable municipality characteristics. Municipalities participating in the early assessment period were compensated with earmarked money from the central government for the increased costs following the expansion of mandatory education (Holmlund, 2008). After the assessment period, the national parliament decides to permanently introduce the 9-year reform to all schools the country in 1962. Seven years later, by 1969, all municipalities were obliged to have the new comprehensive school running (Marklund, 1982) and *Folkskolan* was fully discontinued.

The reform reshaped the school system and compared to the old tracking system students were kept in the same school type for nine years. Besides extending compulsory education from seven or eight to nine years and postponing tracking, the educational reform also came with a small change in the national curriculum implying English became a compulsory subject from 5th grade, but there were no major changes to the total number of hours or the distribution of hours taught in different subjects (Richardson, 1992). In 1953, English was made compulsory even for the old primary school system so all cohorts born after 1945, untreated or treated by the 9-year reform, had the same curriculum (Fredriksson, 1971).

The educational reform also had pedagogical aims. The commission proposal of 1948 was very clear on that the traditional school and its working methods were obsolete. Specifically whole-group teaching and questions-response methods should be replaced by more individualized and activating elements, pandering students drive and

the teacher did not have an appropriate teachers degree. In the early 1940's more than 90 percent of all pupils in Sweden went to a school that were assigned to the main forms (Fredriksson, 1950).

⁴³The comprehensive school system is introduced throughout the whole municipality, or in certain schools within a municipality. At the time there were 1037 municipalities in Sweden. The first treated cohort in a municipality was the fifth graders (c.f. Holmlund (2008)).

⁴⁴Municipalities had to show interest in the reform and also report on various issues, such as e.g. population growth, local demand for education, tax revenues and school situation, and all municipalities that took part in the first year of assessment were required to have eight year comprehensive schooling. The 14 first-movers were selected out of 144 municipalities.

independence (Marklund, 1982).⁴⁵

Based on the principles of the final report of the expert commission a new educational plan for schools to follow is released 1962 (Lgr 62). The pedagogical key concepts of the plan are individualization and activity learning (Larsson, 2011). The pedagogical fundament on the special position of the individual and that the school should foster independent individuals did not meet any major objections (Marklund, 1989). However the first reform municipalities experienced difficulties in getting accurate work material and text books (Marklund, 1982).

A.4 Comparing the Two Reforms

Based on the above it is evident that Sweden experienced a continuous roll-out of extending the compulsory amount of schooling from 6 to 9 years over a period of 40 years, and that the 8-year reform and the 9-year reform were implemented across overlapping cohorts. The roll-out of the two reforms is illustrated in Figure A.1. On average the 8-year and the 9-year reforms were 7 years apart in a municipality.

Both reforms introduced change in the extent of compulsory schooling. However there are important differences between the reforms. Treated students of the 8-year reform faced no significant school system changes, nor any changes in working methods in class. Thus, any effects from the 8-year reform should mainly be driven by changes in the amount of time spent in education.⁴⁶ With the abolishment of the tracking system the 9-year reform implied a fundamental change of the complete school system and the reform also came with new pedagogic methods and and curricula. Effects of the 9-year reform can thus be driven by changes in the amount of time spent in education, by de-tracking, and/or by working methods and pedagogics or curricula change. Given that the curricula for students in *Folkskola* were streamlined with the curricula of the comprehensive school already in 1953, curricula are not a major difference between the two reforms.

The two reforms were gradually implemented across municipalities. The timing of implementation in an individual municipality was based on a mixture of local and national decisions. As regards the wider institutional context, we are unaware of any reforms that might have coincided with the 8-year or the 9-year school reform at the local level. During the assessment period of the 9-year reform it was only municipalities that showed interest in the reform that could be selected implying reform implementation was not random. Previous studies suggest that 9-year reform was implemented earlier in municipalities with higher incomes and with higher average education, see e.g. Lundborg et al. (2014). Regarding the 8-year reform the early birds tended to be more urban and most of the larger cities implemented a mandatory

⁴⁵The emphasis on the importance of the need of new working methods can also be assigned to the aim that education should foster democratic societal members. As discussed by Richardson (1978) there is also a change regarding the view of the individual in the late 1940's. The development of the individual now matters more than the societal development. An essential feature of the report by the commission is that that the school should be more pupil centred and less subject-matter oriented. Another novel perspective is the view that parents, not the school, are responsible for the pupil.

⁴⁶See e.g. discussion by Orring et al. (1962) on that all earlier reforms than the 9-year reform more or less left the fundamental work of schools unaffected.

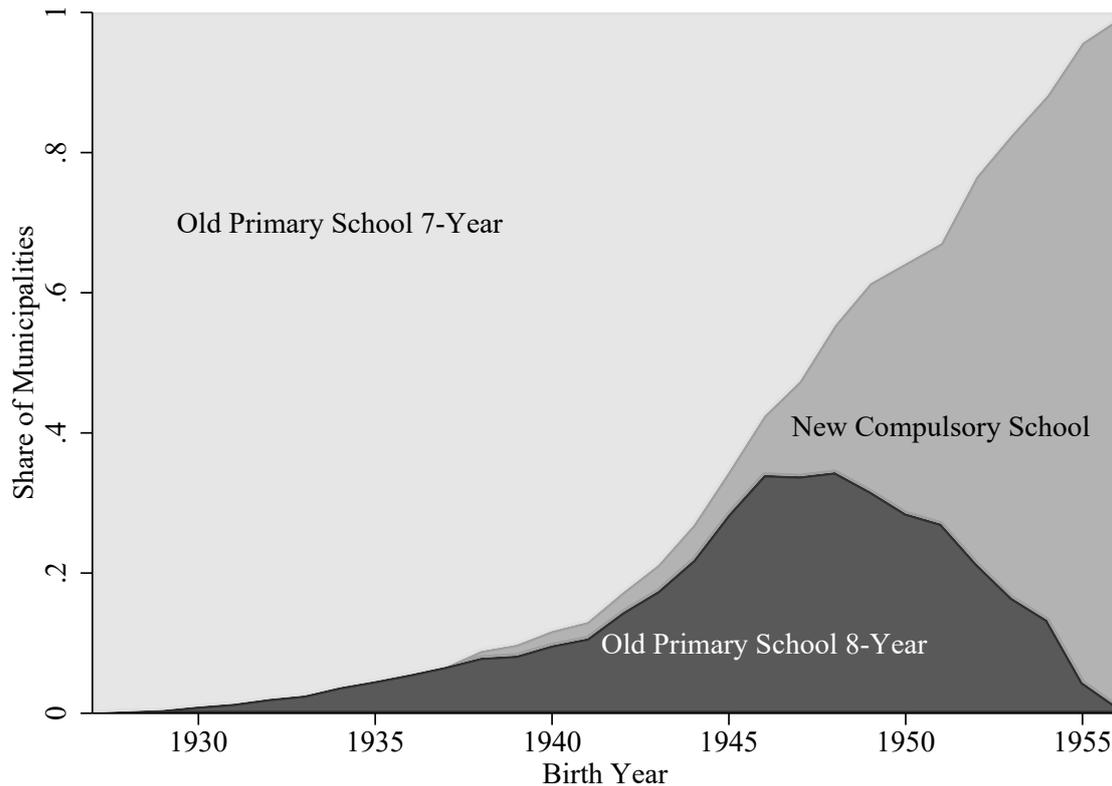


Fig. A.1: Share of municipalities by length of compulsory education

Notes: This figure shows the proportion of municipalities in Sweden who have the 7 year old primary school system, the 8 year old primary school system and the new 9 year compulsory school system by birth cohort.

eighth year. Smaller municipalities followed path and in the end more than half of all municipalities had introduced a mandatory 8th grade before implementing the comprehensive 9-year school reform. In Table A.1 we provide a more systematic assessment of the various determinants of the implementation of the two reforms. As regressors we use a number of municipality characteristics from the 1944-50 period, i.e. observed before or at the very beginning of the reform period.⁴⁷ The variables included are life expectancy and manufacturing employment in 1950 (Statistics Sweden, 1952); some outcomes in the 1944 general election (Statistics Sweden, 1945), the timing of two previous schooling reforms (Fischer et al., 2019), and local economic indicators for 1946 (Statistics Sweden, 1948).

⁴⁷These characteristics are thus observed before the municipality reform implemented in 1952, which reduced the number of municipalities from 2,500 to around 1,000. All variables have been aggregated to the new level using population weights.

Table A.1: Predicting Reform Year

	8-year reform			9-year reform		
	All	Rural	Urban	All	Rural	Urban
Life expectancy 1950	-0.0807** (0.036)	-0.0538 (0.045)	-0.1034 (0.081)	-0.0487* (0.025)	-0.0682** (0.030)	-0.0245 (0.043)
Manufacturing	-10.6429 (7.404)	-10.4366 (8.261)	-14.7155 (20.175)	-11.7006** (4.749)	-8.9365** (4.309)	-16.4158 (10.063)
Labour vote	2.0354 (2.555)	2.2780 (2.454)	1.0778 (7.526)	1.4912 (1.553)	2.0063 (1.318)	-3.0779 (4.963)
Agrarian vote	6.8075** (2.771)	7.4819** (2.841)	7.1625 (7.462)	2.5962 (1.578)	2.9858* (1.624)	-1.5047 (4.654)
Turnout	-3.3989* (1.816)	-4.1512** (1.713)	4.6304 (9.559)	-1.7548 (1.320)	-2.6152** (1.151)	6.0440 (7.085)
Reform year 7	0.0651 (0.039)	-0.0009 (0.051)	0.2033*** (0.070)	-0.0281** (0.013)	-0.0341* (0.019)	0.0234 (0.031)
Reform year term ext	0.0046 (0.040)	0.0300 (0.068)	-0.0285 (0.038)	0.0409* (0.020)	0.0328 (0.048)	0.0499*** (0.015)
Property values	-0.0002 (0.000)	-0.0002 (0.000)	-0.0001 (0.000)	-0.0003** (0.000)	-0.0003*** (0.000)	-0.0000 (0.000)
Mean earnings	-0.0017 (0.001)	-0.0011 (0.001)	-0.0021 (0.002)	-0.0009 (0.001)	-0.0015* (0.001)	-0.0003 (0.001)
City	-2.1518** (1.022)		0.1090 (1.208)	1.2852** (0.617)		0.5799 (0.873)
Town	-1.5677* (0.847)			0.1308 (0.664)		
Mean	1961.5	1962.5	1957.5	1965.4	1965.6	1964.9
N. of cases	985	786	199	985	786	199

Notes: The dependent variable in the columns headed “8-year reform” is the first year in which compulsory schooling exceeded 7 years. For the 9-year reform, the dependent variable is the year in which the new comprehensive school system was implemented. Life expectancy was calculated based on age-specific mortality rates in the municipality averaged over the 1948-52 period using the universe of deaths in these years (cf. Fischer et al., 2013). ‘Manufacturing’ represents the proportion of the adult male population working in manufacturing in 1950 (Statistics Sweden, 1952). ‘Labour vote’ represents the share of valid votes for the labour party in the 1944 general election; ‘Agrarian vote’ represents the same statistic for the agrarian party (*Bondeförbundet*) and ‘Turnout’ is the voter turnout in the same election (Statistics Sweden, 1945). ‘Reform year 7’ and ‘Reform term ext’ are the reform years of the two schooling reforms studied in Fischer et al. (2019); in merged municipalities, we always used the earliest year. Property values per capita and mean earnings refer to 1946 and are taken from municipality yearbooks (Statistics Sweden, 1948). Each specification controls for county fixed effects and standard errors are clustered at the county level: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

The estimates in Table A.1 show that the 8-year reform is implemented earlier in urban municipalities: cities are 2 years earlier on average. Also municipalities with a high turnout in the 1944 election reform early, whereas large support for the agrarian party tends to delay the reform. In addition, municipalities with a greater life expectancy in 1950 tended to implement the reform earlier. The 9-year reform is different in some important respects: cities implement it later on average, whereas employment in manufacturing and high property values predict earlier reform year. Also, the two previous school reforms—the term length extension and the 7-year reforms implemented in the 1936-50 period—predict the 9-year reform in different directions. However, also for the 9-year reform, a greater life expectancy in 1950 predicts earlier reform year.

As reported in the main text (cf. Section 3.3), there were no significant changes in the national funding formula for schools during the reform period. Thus it can be assumed that available resources per pupil-year were largely unaffected by both reforms. In Table A.2 we provide some evidence to corroborate this claim. Based on census information, we approximated pupil/teacher ratios for the years 1946, 1950, 1960 and 1970. We calculate the pupil-teacher ratio in municipality m in year t as

$$R_{mt} = \frac{\sum_{a \in \Theta_{mt}} P_{mat}}{T_{mt}} \quad (17)$$

where $\Theta_{mt} \subset \{7, 8, \dots, 16\}$ represents the cohorts of compulsory schooling age in municipality m at time t , P_{mat} is the municipality population of age a in year t , and T_{mt} are the number of elementary and lower secondary teachers residing in municipality m in year t . The coefficients in Table A.2 are from DiD specification similar to the main analysis using $\ln R_{mt}$ as the outcome variable. The increase in number of pupils is larger than the increase in number of teachers in both reforms, causing increases in the pupils/teacher ratios. Whereas the increase in number of pupils is somewhat similar across the reforms, the increase in teachers is smaller for the 8-year reform. Thus, the increase in pupils/teacher ratio is larger for the 8-year reform compared to the 9-year reform. It should be noted that our measure R_{mt} only represents a proxy for the actual pupil-teacher ratios applying in a municipality. First, we include pupils who go to school in another municipality in the measure. Second, we assume that teachers residing in a municipality work as teachers in that municipality. The actual ratios will thus deviate from this proxy.

Table A.2: Compulsory Schooling Reforms' Impact on Pupil-Teacher Ratios

	RATIO		TEACHERS		PUPILS	
	(1)	(2)	(3)	(4)	(5)	(6)
8-year reform	0.0774*** (0.018)	0.1035*** (0.024)	0.0677*** (0.019)	0.0411 (0.023)	0.1451*** (0.008)	0.1446*** (0.010)
9-year reform	0.0274 (0.029)	0.0781* (0.031)	0.1380*** (0.030)	0.0729* (0.032)	0.1654*** (0.009)	0.1511*** (0.010)
Mean	13.128	13.128	0.010	0.010	0.118	0.118
N. of cases	3,952	3,952	3,952	3,952	3,952	3,952
Year fixed effects	✓	✓	✓	✓	✓	✓
Municipality fixed effects	✓	✓	✓	✓	✓	✓
Municipality trends		✓		✓		✓

Notes: This table shows the impact of the 8 year and 9 year school reforms on pupil-teacher ratios, teacher densities and pupil densities controlling census year and municipality fixed effects. Robust standard errors clustered at the municipality level are in parentheses. Testing the null of the coefficient: * $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$

Source: Population censuses 1950, 1970 and 1970 (Statistics Sweden, 1952, 1965, 1974). Own calculations.

A.5 Peer Group Composition in the Big Cities

As mentioned in section 4.2, we need to exclude the three big cities Stockholm, Gothenburg and Malmö due to the staggered implementation of the 9-year reform

in those cities. Since 19% of the Swedish population resided in these cities in the reform period, it is of interest to consider whether results will be representative also for them. To the extent that the big cities are more segregated, the implications of the reforms for peer mix may be very different compared to the rest of Sweden. In order to shed some light on this issue, we compare the local probability of taking secondary schooling in the pre-reform period. We thus calculated the cohort-specific probability of taking secondary schooling in Stockholm, Gothenburg and Malmö and compared them to the corresponding numbers pre-reform in other cities. The resulting histograms are provided in Figure A.2. We chose cohorts 1937-42 because in Stockholm, the implementation of the 9-year reform started with cohort 1943; in the two other cities it was implemented a few years later. The left figure shows the distribution at the parish-cohort level. It indicates that even though the three biggest cities have higher secondary schooling attendance on average, there is substantial overlap in the distributions. When we go down to the district level, the overlap between the two histograms increases even further. Importantly, even though the three major cities have higher attendance rates on average, these rates do not appear to be more spread out.

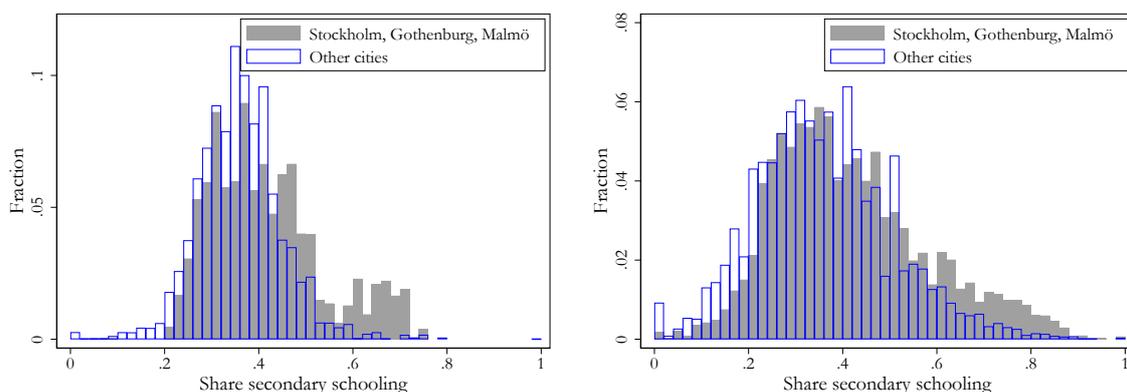


Fig. A.2: Distribution of Secondary School Attendance Within Cities, Cohorts 1937-42.

Notes: The left panel is for the parish level ($N = 275$), the right panel is for the district level ($N = 1,435$). The district level normally corresponds to electoral district, sometimes to taxation district (Statistics Sweden, 1965). Secondary schooling attendance calculated at the cohort level within each administrative unit. *Source:* Population censuses 1950, 1970 and 1970 (Statistics Sweden, 1952, 1965, 1974). Own calculations.

A.6 Reform Data and Validation

The reform data for the 9-year reform was generously shared by Helena Holmlund and we rely on a dataset as used in Hjalmarsson et al. (2015), of which an earlier version is described in detail in Holmlund (2008). The dataset encompasses information on the year a specific school district introduced the new comprehensive school.

While the 9-year reform has previously been used in several economic applications, this paper is the first to use the 8-year reform and the reform data on the timing of the year of introduction of the eight year in each municipality was purposively collected from archives and digitized by the authors. Various official sources provide aggregate information on the development of the implementation on the 8-year reform.

To check the accuracy of the gathered reform data we perform checks to confirm that the collected information conform with aggregate official statistics. For example, information on the share of school districts in the country that had eight years of compulsory schooling in certain years from Skolöverstyrelsen (1955) and from Sweden (1977), respectively, suggest our data conform with aggregate statistics.

The decision to introduce eight years of compulsory schooling was made on the municipal level, and the assumption is that schools within the same district implemented the reform in the same year. Theoretically there could however be discrepancies between schools within municipalities. We believe the assumption is valid since official sources state that the change generally applied to a whole district (see e.g. Fredriksson (1971) and Skolöverstyrelsen (1955)). Moreover, aggregate figures on the share of all students taking on the extra year of compulsory education in certain years (Sweden, 1977) suggest that there should be no major deviations from this rule.

Appendix B Conceptual Model: Extensions

B.1 Spillover Effects

Our theoretical model requires that there are no spillover effects into higher education as a consequence of the two reforms. Our empirical analysis in appendix section C.1 illustrates that spillovers are negligible for the 8-year reform. This simplifies interpretation of both the causal effect of schooling, as well as interpretation of the the comparison of the two reforms. For the 9-year reform, however, Table C.3 suggests there was a small increase in the propensity to take secondary education of 4 percentage points. These spillover effects increased average years of schooling by approximately 0.1 years.⁴⁸ This should be compared to the first-stage estimate of 0.54 years. We demonstrate below that these spillovers likely have little impact on our original interpretation of δ .

Under the assumption of no spillovers $\beta_{9,2SLS}$ equals:

$$\beta_{9,2SLS} = \frac{(1 - \alpha) [\eta A_0 + \theta P(\mathbf{A}_0^t)] + 5\theta [P(\mathbf{A}^c) - \alpha P(\mathbf{A}_1^t) - (1 - \alpha) P(\mathbf{A}_0^t)]}{(1 - \alpha)} \quad (18)$$

The denominator equals the first stage, which is equal to the share of individuals in the basic track increasing their education by one year $(1 - \alpha)$. The denominator equals the effect of the 9-year reform on schooling. Recall from our exposition in section 3, that $\beta_{8,2SLS} = \eta A_0 + \theta P(\mathbf{A}_0^t)$ and that our suggested measure of the peer effect, δ , equals the difference in human capital due to the change in peer composition $\delta = 5\theta [P(\mathbf{A}^c) - \alpha P(\mathbf{A}_1^t) - (1 - \alpha) P(\mathbf{A}_0^t)]$. Thus we can express $\beta_{9,2SLS}$ as

$$\beta_{9,2SLS} = \beta_{8,2SLS} + \frac{\delta}{(1 - \alpha)} \quad (19)$$

Consider a spillover effect to upper secondary school of ω years. The first stage will now equal the share of individuals in the basic track and this spillover effect: $(1 - \alpha) + \omega$. The effect of the reform now also includes the effects of schooling for individuals continuing to upper secondary school. Using a notation similar to our model, a year of upper secondary schooling yields additional human capital $\eta_u A_1 + \theta P(\mathbf{A}_1^u)$, where we allow the individual effect of upper secondary schooling η_u to be different from η (the effect for lower levels). Note that we assume that these individuals are of the high-ability type, denoting their ability by A_1 , and that in general only high-ability individuals attend school at this level, i.e., the ability profile equals \mathbf{A}_1^u .⁴⁹

⁴⁸3.2 percentage points took 2-years upper secondary, 1.2 percentage points 3-years upper secondary education.

⁴⁹The historical documentation suggest that this is a reasonable assumption. In the case also low ability individuals attend upper secondary school, the exposition in this section would on the one hand be simpler as we would comparing the effect of schooling for the same type of individuals at different margins so that the difference in the effect comes to down to the accuracy of our linearity assumption. On the other hand, low-ability individuals continuing to upper secondary schooling would on the margin change the ability composition at this level of schooling. However, this would not give rise to peer effects, given that there was strong sorting within upper secondary schooling based on ability

In the presence of spillovers, the estimate $\beta_{9,2SLS}$ now equals:

$$\beta_{9,2SLS} = \frac{(1 - \alpha) [\eta A_0 + \theta P(\mathbf{A}_0^t)] + \omega [\eta_u A_1 + \theta P(\mathbf{A}_1^u)]}{(1 - \alpha) + \omega} + \frac{\delta}{(1 - \alpha) + \omega} \quad (20)$$

The first term is a weighted average of marginal effects arising from additional years of schooling at different levels. We rearrange this term to illustrate that it will be approximately similar to $\beta_{8,2SLS}$ given that spillover effect ω is much smaller than the first stage $(1 - \alpha)$:

$$\beta_{8,2SLS} \approx \frac{(1 - \alpha + \omega) [\eta A_0 + \theta P(\mathbf{A}_0^t)] - \omega [\eta A_0 + \theta P(\mathbf{A}_0^t)] + \omega [\eta_u A_1 + \theta P(\mathbf{A}_1^u)]}{(1 - \alpha) + \omega} \quad (21)$$

This approximation will be accurate if either ω is small or if the effect on human capital acquisition is similar for the 9th year of schooling (for a low-ability individual) and for a year of upper secondary schooling for a high-ability individual, i.e., $[\eta A_0 + \theta P(\mathbf{A}_0^t)] \approx [\eta_u A_1 + \theta P(\mathbf{A}_1^u)]$. In part, this relates to the linearity assumption imposed in section 3. Spillover effects are noted at 11 and 12 years of education. Figure C.2 suggests that linearity may apply in this segment as well.

Using the approximation (21), it becomes clear that our estimate $\hat{\delta}$ approximates δ also in the presence of spillovers:

$$\hat{\delta} = (\beta_{9,2SLS} - \beta_{8,2SLS}) (1 - \alpha + \omega) \approx \delta \quad (22)$$

where δ is shorthand for the peer effect associated with de-tracking, defined in section 3.1. However, the approximation error in $\hat{\delta}$ is a function of model parameters:

$$\frac{\hat{\delta} - \delta}{\delta} = \frac{\omega (\beta_{8,2SLS} - [\eta_u A_1 + \theta P(\mathbf{A}_1^u)])}{\delta} = \frac{\omega \lambda \beta_{8,2SLS}}{\delta} \quad (23)$$

where $\lambda = \left(1 - \frac{[\eta_u A_1 + \theta P(\mathbf{A}_1^u)]}{\beta_{8,2SLS}}\right)$ captures the difference in years-of-schooling effects between 8th grade and upper secondary school (for the relevant types). If $\lambda > 0$, $[\eta_u A_1 + \theta P(\mathbf{A}_1^u)] > \beta_{8,2SLS}$ and thus a year of upper secondary schooling has a smaller impact on mortality than the 8th year. We can thus define ranges for λ that assure that the bias term is less than some value c in absolute value:

$$\frac{\hat{\delta} - \delta}{\delta} \in (-c, c) \iff \lambda \in \left(\frac{c}{(1+c)\omega\beta_{8,2SLS}}, -\frac{c}{(1-c)\omega\beta_{8,2SLS}} \right) \quad (24)$$

Since we estimate the parameters $\beta_{8,2SLS}$, ω and $\hat{\delta}$, we are able to calculate these bounds for any level of tolerance c . It follows that the bias will be less than ten per cent ($|\frac{\hat{\delta} - \delta}{\delta}| < 0.1$) whenever $\lambda \in (-0.61, 0.74)$. Therefore, the mortality effect of a year

(SOU 1963:41, 1963).

of upper secondary schooling for a high-ability individual may be 61 per cent larger or 74 per cent smaller than the effect of the 8th year for a low-ability individual and the bias in $\hat{\delta}$ is still below 10 per cent. Hence, we conclude that results are robust to relatively large deviations in the effect of schooling from different levels and ability types.

B.2 Relaxing Additivity

In the main treatment in section 3 we assume that peer effects are additive. This assumption is a useful approximation but it is unnecessarily restrictive. We now relax that assumption and show what it implies for our estimates. Some additional notation is necessary. Thus, we denote by the function $g_{k\tau s}$ the human capital (in our case: health) acquired after s years of schooling by an individual of ability type k and tracking type τ . An individual of type τ is defined as an individual who would attend track τ would there be a tracking system in place. Note that this means that can we relax the assumption of perfect separation of ability types in a tracking system, an issue we return to below. The peer group to which an individual is exposed in each year of school, starting with year 5, denoted $\mathbf{A}_{\tau s}^r$ enters the production function as arguments, where $r \in \{t, c\}$ captures how the peer group depends on the school system (t for tracking, c for comprehensive) and tracking type. Hence:

$$g_{k\tau s} = g_{ks}(\mathbf{A}_{\tau 5}^r, \dots, \mathbf{A}_{\tau s}^r) \forall s \geq 5 \quad (25)$$

Working with this set of production functions, we can define the reform effects under much more general assumptions. The impact of the 8-year reform can be defined as

$$\Gamma_8^{k\tau} = \begin{cases} g_{k8}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{08}^t) - g_{k7}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{07}^t) & \text{if } \tau = 0 \\ 0 & \text{if } \tau = 1 \end{cases} \quad (26)$$

Likewise, the impact of the 9-year reform now becomes

$$\Gamma_9^{k\tau} = \begin{cases} g_{k9}(\mathbf{A}_{05}^c, \dots, \mathbf{A}_{09}^c) - g_{k8}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{08}^t) & \text{if } \tau = 0 \\ g_{k9}(\mathbf{A}_{15}^c, \dots, \mathbf{A}_{19}^c) - g_{k9}(\mathbf{A}_{15}^t, \dots, \mathbf{A}_{19}^t) & \text{if } \tau = 1 \end{cases} \quad (27)$$

Comparing equation (26) and (27) for the academic track types ($\tau = 1$) clearly shows that—independently of ability—the difference between the two reform effects represent the effect of being exposed to different peers between years 5 and 9:

$$\Delta_{89}^{k1} = \Gamma_9^{k1} - \Gamma_8^{k1} = g_{k9}(\mathbf{A}_{\tau 5}^c, \dots, \mathbf{A}_{\tau 9}^c) - g_{k9}(\mathbf{A}_{\tau 5}^t, \dots, \mathbf{A}_{\tau 9}^t) \quad (28)$$

For the basic track types, the difference between the two reform effects has two

components:

$$\begin{aligned}
\Delta_{89}^{k0} &= \Gamma_9^{k0} - \Gamma_8^{k0} \\
&= g_{k9}(\mathbf{A}_{05}^c, \dots, \mathbf{A}_{09}^c) - g_{k8}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{08}^t) - [g_{k8}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{08}^t) - g_{k7}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{07}^t)] \\
&= g_{k9}(\mathbf{A}_{05}^c, \dots, \mathbf{A}_{09}^c) - g_{k9}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{09}^t) \\
&\quad + [g_{k9}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{09}^t) - g_{k8}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{08}^t)] - [g_{k8}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{08}^t) - g_{k7}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{07}^t)]
\end{aligned}$$

The first component—in the second row—represents the effect of being exposed to different peers through years 5 to 9. The second component—in the third row—equals the difference in human capital acquired in the 8th and 9th year of schooling. In Section 3 we imposed a linearity assumption, i.e. that an additional year of schooling has the same effect on human capital when schooling is extended to 9 years, as it has when schooling is extended to 8 years. Under that assumption, this second component equals zero. Thus, the main conclusion from section 3 holds also in this more general case, but this exposition makes clear that we identify a “local” effect which applies when the peer group is changed in exactly these age ranges; results may not have external validity in contexts where the peer group is changed at older or younger ages (or for a different duration).

B.3 Separation of types and mis-classifications

Next, we study the consequences of relaxing the assumption of perfect separation of types. Let γ_τ denote the share of high ability individuals of track type τ . Our 2SLS-estimate of the 8-year reform now equals a weighted average of the effect of the *low-ability and basic track* types and the *high-ability and basic track* types of individuals:

$$\beta_{8,2SLS} = \frac{(1 - \alpha) [\gamma_0 \Gamma_8^{00} - (1 - \gamma_0) \Gamma_8^{10}]}{(1 - \alpha)} \quad (29)$$

As before the 2SLS-estimate of the 9-year reform equals

$$\beta_{8,2SLS} = \beta_{8,2SLS} + \frac{\delta}{(1 - \alpha)} \quad (30)$$

where δ is re-weighted difference between the reforms. Now δ equals the weighted average of effect of the change in peer exposure for the four groups (defined by ability \times track):

$$\begin{aligned}
\delta &= \alpha \gamma_1 [g_{19}(\mathbf{A}_{15}^c, \dots, \mathbf{A}_{19}^c) - g_{19}(\mathbf{A}_{15}^t, \dots, \mathbf{A}_{19}^t)] + \dots \\
&\quad \dots + \alpha(1 - \gamma_1) [g_{09}(\mathbf{A}_{15}^c, \dots, \mathbf{A}_{19}^c) - g_{09}(\mathbf{A}_{15}^t, \dots, \mathbf{A}_{19}^t)] + \dots \\
&\quad \dots + (1 - \alpha) \gamma_0 [g_{19}(\mathbf{A}_{05}^c, \dots, \mathbf{A}_{09}^c) - g_{19}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{09}^t)] + \dots \\
&\quad \dots + (1 - \alpha)(1 - \gamma_0) [g_{09}(\mathbf{A}_{05}^c, \dots, \mathbf{A}_{09}^c) - g_{09}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{09}^t)]
\end{aligned}$$

Assuming perfect separation of types is equivalent to $\gamma_1 = 1$ and $\gamma_0 = 0$ and then δ

simplifies to weighted average of the high-ability academic track and the low-ability basic track types as in the main exposition:

$$\delta = \alpha \left[g_{19}(\mathbf{A}_{15}^c, \dots, \mathbf{A}_{19}^c) - g_{19}(\mathbf{A}_{15}^t, \dots, \mathbf{A}_{19}^t) \right] + \dots \\ + \dots (1 - \alpha) \left[g_{09}(\mathbf{A}_{05}^c, \dots, \mathbf{A}_{09}^c) - g_{09}(\mathbf{A}_{05}^t, \dots, \mathbf{A}_{09}^t) \right]$$

This exposition illustrates that independent of whether the tracking system perfectly separate ability types the difference between the two reform effects represent the effect of being exposed to different peers between years 5 and 9. The change in peer groups differ between but not within track type. That is, high and low ability individuals in the same track is exposed to the same change in peers. Of course, the response to this change may be different depending on the person's ability. The change in human capital acquisition for a high-ability basic track student that is now being exposed to more individuals of similar high-ability is not necessarily the same as that of a low ability (basic track) student being exposed to more high ability students. As shown this has little implication for our main result that any difference in the effect of the 8-year and 9-year reform is indicative of the peer effects being present. However, our possibilities to make statements about the type of peer effect at play based on subgroup analysis is reduced if either the degree of misclassification or the peer effects among misclassified individuals are large.