Govert E. Bijwaard, Mikko Myrskylä, Per Tynelius, Finn Rasmussen

Educational gains in cause-specific mortality: accounting for cognitive ability and family-level confounders using propensity score weighting

Article (Published version) (Refereed)

Original citation:
Bijwaard, Govert E. and Myrskylä, Mikko and Tynelius, Per and Rasmussen, Finn (2017) *Educational gains in cause-specific mortality: accounting for cognitive ability and family-level confounders using propensity score weighting.* Social Science and Medicine, 184. pp. 49-56. ISSN 0277-9536

DOI: 10.1016/j.socscimed.2017.05.019

Social Science & Medicine 184 (2017) 49–56

Contents lists available at ScienceDirect

Social Science & Medicine

journal homepage: www.elsevier.com/locate/socscimed

Educational gains in cause-specific mortality: Accounting for cognitive ability and family-level confounders using propensity score weighting

Govert E. Bijwaard [a, *], Mikko Myrskylä [b, e, f], Per Tynelius [c], Finn Rasmussen [d]

[a] Netherlands Interdisciplinary Demographic Institute (NIDI-KNAW/University of Groningen), PO Box 11650, 2502 AR, The Hague, The Netherlands
[b] Max Planck Institute for Demographic Research, Rostock, Germany
[c] Department of Public Health Sciences, Karolinska Institute, Tomtebodavägen 18A; Widerströmska huset, 171 77 Stockholm, Sweden
[d] Department of Health Sciences, Lund University, Box 117, 221 00 Lund, Sweden
[e] Department of Social Policy, London School of Economics and Political Science, London, UK
[f] Department of Social Research, University of Helsinki, Helsinki, Finland

ARTICLE INFO

Article history:
Received 16 February 2017
Received in revised form
5 May 2017
Accepted 7 May 2017
Available online 8 May 2017

Keywords:
Sweden
Cause-specific mortality
Causal effect of education
Months lost analysis
Inverse probability weighting
Fixed effects

ABSTRACT

A negative educational gradient has been found for many causes of death. This association may be partly explained by confounding factors that affect both educational attainment and mortality. We correct the cause-specific educational gradient for observed individual background and unobserved family factors using an innovative method based on months lost due to a specific cause of death re-weighted by the probability of attaining a higher educational level. We use data on men with brothers from the Swedish Military Conscription Registry (1951–1983), linked to administrative registers. This dataset of some 700,000 men allows us to distinguish between five education levels and many causes of death. The empirical results reveal that raising the educational level from primary to tertiary would result in an additional 20 months of survival between ages 18 and 63. This improvement in mortality is mainly attributable to fewer deaths from external causes. The highly educated gain more than nine months due to the reduction in deaths from external causes, but gain only two months due to the reduction in cancer mortality and four months due to the reduction in cardiovascular mortality. Ignoring confounding would lead to an underestimation of the gains by educational attainment, especially for the less educated. Our results imply that if the education distribution of 50,000 Swedish men from the 1951 cohort were replaced with that of the corresponding 1983 cohort, 22% of the person-years that were lost to death between ages 18 and 63 would have been saved for this cohort.

1. Introduction

The fact that individuals with lower socioeconomic status have higher mortality rates has been well established in social science research (Hummer and Lariscy, 2011). Of the various socioeconomic measures that are commonly used to investigate this relationship, the educational gradient has been shown to be particularly robust (Cutler et al., 2011). This educational gradient is apparent not only in less-developed countries, but in the US (Masters et al., 2012) and in other western countries with advanced health care systems (Huisman et al., 2005). For example, in the year 2010, life expectancy at age 30 was 53 years for Swedish men with a university degree and only 48 years for Swedish men with less than secondary education. Differences by educational attainment have been reported for most causes of death, and particularly for cardiovascular diseases and some types of cancer (Huisman et al., 2005). Many of these causes of death are theoretically avoidable through prevention or treatment. According to the "fundamental cause theory" proposed by Link and Phelan (1995), the educational gradient is steeper for such causes of death because personal resources and social context can be used to acquire health-related knowledge that prevents these causes or facilitates recovery from these diseases (Link and Phelan, 1995). According to this theory, highly educated individuals find it easier than their less educated counterparts to handle complex treatment regimens; whereas the educational gradient for non-preventable diseases, which are less under human control, is smaller.

While it is commonly assumed that this educational gradient in mortality has a causal interpretation, this assumption has been challenged in the literature. The association may be partly explained by confounding factors that affect both educational

* Corresponding author.
E-mail addresses: bijwaard@nidi.nl (G.E. Bijwaard), Myrskyla@demogr.mpg.de (M. Myrskylä), Per.Tynelius@ki.se (P. Tynelius), finn.evald.rasmussen@gmail.com (F. Rasmussen).

http://dx.doi.org/10.1016/j.socscimed.2017.05.019

attainment and mortality (Grossman, 2015). If such common traits exist, the association may be spurious. This would have important policy implications. Yet most of the studies investigating differences by educational attainment in cause-specific mortality account only partially for such factors.

We account for confounding factors using a propensity score method. Propensity score weighting methods for hazard models, such as the mortality hazard, that account for censoring, truncation and dynamic selection issues have recently been introduced (Cole and Hernán, 2004; Austin, 2014). The critical assumption in propensity score weighting is that there is no selection on unobserved factors. However, twin studies have emphasized that even after accounting for an individual's socioeconomic background, the association between education and mortality may be a reflection of other factors shared within a family. In addition to sharing 50% of their genes, siblings typically share an environment during childhood. To account for the impact of these unobserved family characteristics on the educational gradient in mortality, we apply a family fixed effects model that compares only siblings (brothers).

The most common approach used to analyze the impact of education on cause-specific mortality has been to estimate a Cox proportional hazard for each cause of death separately. However, when there are competing risks, and education influences each cause-specific hazard, the interpretation of the proportional hazard results of education on cause-specific mortality is unclear. Even the sign of the effect of education is unclear, because both the total survival and the cause-specific cumulative incidence functions depend not only on the cause-specific hazard, but on the hazards of all of the other causes. Another problem is that the contribution of a given cause to the gain by educational attainment cannot be derived. A direct approach that can be used to account for the importance of a specific cause is to estimate the impact of education on the months lost due to a specific cause. To account for the observed confounding factors, we estimate these month-lost models using inverse probability weighting (IPW) based on the propensity score. Drawing upon this synthetic sample, we estimate linear month-lost models with family fixed effects. The family fixed effects account for the unobserved family confounding factors. Although the family fixed effect approach has been used in previous analyses of cause-specific mortality and education or of other adult characteristics, these analyses have been conducted within the Cox proportional hazard framework (Barclay et al., 2016; Elo et al., 2014).

The Swedish Military Conscription Register for men born in 1951–1983 that are linked to Swedish administrative registers provides us with the opportunity to investigate these questions. These data include information on the men's demographic and socioeconomic characteristics, such as education, parents (both father and mother), socioeconomic status, parents' education, and area of residence. The data also include the results of anthropometric measures, and an intelligence test. We include in our analysis all men with a known conscription date and at least one brother ($N = 700{,}043$). This large sample allows us to distinguish 22 causes of death and five educational levels.

The empirical results show that improving education by one level would lead to three to 10 additional months of life between ages 18 and 63. Most of these gains are attributable to the reduction in mortality due to external causes. Men with primary education would gain the most from the reduction in mortality from suicide (1.3 months) and from other external causes (2.4 months). The reduction in CVDs and cancers with improvements in education is rather small. Ignoring confounding would underestimate the gains by educational level for the men with primary education.

2. Previous research and conceptual framework

2.1. Cause-specific mortality

For most causes of death, large mortality differences by socioeconomic status have been observed in both North America and Europe. These inequalities are persistent and several theories seek to explain them (Mackenbach, 2012); we review only the most prominent ones. According to the “fundamental cause theory” (Link and Phelan, 1995), the educational gradient should be steepest for the causes of death that are more preventable and/or curable. This theory posits that these health differences are caused by variation across socioeconomic groups in access to resources that can be used to avoid these diseases, and to slow down the development of these diseases after they have been contracted, regardless of the current level of exposure to risk factors. The “life course perspective” emphasizes the importance of unfavorable early life circumstances in explaining the pathways to both health and social disadvantages in adulthood. Though widely observed, these negative education-mortality associations may not necessarily reflect the beneficial effects of education on mortality, as other individual factors may influence both education and mortality.

Since the evidence suggests that education has varying associations with different diseases, there is also a range of educational gradients in cause-specific mortality (Galobardes et al., 2004). For example, the educational gradient in mortality from cardiovascular diseases (CVD) appears to be steeper than the educational gradient in mortality due to other causes of death (Kulhánová et al., 2014). A potential explanation for this finding is that having a low level of education tends to be associated with having cardiovascular risk factors, such as being a smoker, having hypertension, and being overweight. For cancers, the educational gradient varies by cancer type (Galobardes et al., 2004; Kulhánová et al., 2014). The relatively high levels of mortality from lung cancer found among the less educated are related to smoking. The relationship between educational level and mortality rates is more complex for other cancers, though it may be attributable to lifestyle factors, such as differences in levels of physical activity. However, a recent study has found little evidence that education affects cancer mortality (Leuven et al., 2016), except mortality from lung and prostate cancer among men.

Several studies have investigated the relationship between education and death from external causes, including suicides and traffic accidents (Borrell et al., 2005; Lorant et al., 2005). Such causes make up a large share of mortality among young adults. Socioeconomic inequalities in suicide rates among men have been observed in many countries. Differences by educational attainment in the incidence of mental illness, which is more prevalent among the less educated, may explain this educational gradient. Meanwhile, the differences by educational level in deaths from traffic accidents can be explained by differences in exposure, such as differences in the use of protective devices and in susceptibility.

2.2. Causal inference

Most existing studies address confounding by using a multivariate regression framework with the putative observed confounders as control variables. We seek to gain a better understanding of the causal impact of education level on cause-specific mortality. In the literature, three distinct approaches have been employed to estimate the causal effects of education on mortality. The first approach exploits changes in compulsory schooling policies as instrumental variables for schooling attainment in order to control for possible confounders. The estimates based on these studies suggest that the causal effect of education

on health outcomes is either small (Lleras-Muney, 2005; Van Kippersluis et al., 2011; Meghir et al., 2013) or absent (Arendt, 2005; Albouy and Lequien, 2009; Clark and Royer, 2013). A limitation of using changes in compulsory schooling to detect educational effects on mortality is that in many contexts, only a relatively small share of the population is affected by these laws (Mazumder, 2008; Fletcher, 2015).

A second approach is to control for unobserved genetic and environmental factors by examining variation in education between siblings. These studies generally obtain estimates of the effects of the differences in schooling by examining the health differences in a pair of siblings, sometimes twins, at various levels of schooling. The results indicate that a portion of the differences by educational attainment in cause-specific mortality disappears when shared family background is accounted for (Lundborg, 2013; Amin et al., 2015; Elo et al., 2014). These studies suffer from issues related to data limitations and generalization, as they only use data on siblings, even though families with multiple children might not represent the general population.

A third approach that is used to account for confounding factors is to incorporate them directly into the model. A structural framework developed by Conti et al. (2010) explicitly models the interdependence of education, health, and the confounding factor(s). The results of such models for mortality have shown that at least half of the health disparities across educational groups are due to the selection of healthier, more capable individuals into higher education (Bijwaard et al., 2015a,b). In a recent paper, this framework was extended to analyze the educational gains in cause-specific mortality (Bijwaard et al., 2016). The findings indicated that the largest educational gains could be achieved by men with low levels of education, and that such gains would in turn lead to a reduction in mortality from external causes among this group. However, these models impose a rather stringent structure on the relationship between education, mortality, and the influence of confounding factors. Another limitation of these structural models is that estimating them can be very computer intensive if many different causes of death are included.

The propensity score method we employ accounts for possible confounding factors, but without making structural assumptions about the relationship between the confounding factors and (cause-specific) mortality. Our approach is based on the assumption that all of the variables that affect both mortality and education attainment are observed. This is a stringent assumption, but our data contain information on a range of important factors, including detailed information on each individual's family socioeconomic background (including the father's and the mother's socioeconomic status at birth and their educational levels), cognitive skills (IQ test), and non-cognitive skills (psychological test). We further relax this assumption by comparing each individual with his brother(s), and use a propensity score method with fixed effects that takes into account the impact of unobserved family characteristics on the educational gradient in mortality.

3. Methodology

The standard approach to analyzing cause-specific mortality is to formulate a competing risks model with an independent Cox proportional hazards model for each cause-specific outcome. The cause-specific hazard gives the mortality rate due to a particular cause, conditional on not having previously died from any other cause. Caution is needed when interpreting these results because a particular covariate can appear in several competing hazards (Thomas, 1996). In such a case, the coefficients (or the hazard ratios) convey little information about the effect of the covariate on the probability of dying from a specific cause, because that probability depends not only on the hazard of dying from the specific cause, but also on the hazard of dying from all of the other causes. In addition, the Cox competing risk models provide information only on the magnitude of the educational disparity, not on the importance of the specific cause.

To allow for a comparison between our results and those of the previous literature, we also estimate Cox models. However, in using the Cox models, we deviate from the standard literature: instead of pooling the data and using educational dummies, we estimate four separate models using the data of individuals in two adjacent educational levels (for each cause of death) with only one dummy variable for the higher educational level. We thereby relax the assumption of common age dependence by educational level that is implicit in the Cox model.

3.1. Months lost due to a specific cause of death

A quantity with a more natural interpretation that avoids the issues of interdependence that can arise when using a competing risks proportional hazards model is the number of *months lost* due to a specific cause of death (Andersen, 2013; Andersen and Canudas-Romo, 2013). The number of months lost can be defined over the whole age distribution; e.g., as the number of months lost before age 63 (the number we will use). An attractive feature of the months-lost model is that the sum over all alternative causes of death is equal to the total number of months lost. This figure can be calculated using non-parametric methods; and can be based on the estimated hazard coefficients, the implied total survival, and the cumulative incidence functions. Non-parametric estimation of the number of months lost is straightforward because estimating the survival (Kaplan-Meier) and the cumulative incidence functions is straightforward; see the supplementary appendix.

The advantage of using a months-lost model is that the total gain from moving up the whole educational ladder (from primary education to higher education) is the sum of each level of educational progress. This model allows us to directly derive the educational gains associated with the major causes of death (grouped), cancers, CVDs, other diseases, and external causes; and the educational gains made when improving education over the whole educational ladder, from primary education to higher education.

Another advantage of relying on numbers of months lost is that they can be modelled using standard (uncensored) linear models. However, an issue that often arises when using mortality data is that not all of the individuals are followed until they die, but only until the end of the observation period. This implies that the age at death is (heavily) censored; in our case, at age 63. This censoring complicates our regression analyses. For cases in which some of the observations are censored, it is possible to carry out regression analyses for the numbers of months lost for each specific cause of death by using constructed pseudo-observations; see Andersen (2013) and the supplementary appendix.

3.2. Accounting for confounding

We follow a propensity score method in accounting for endogeneity of education. Propensity score methods are increasingly used to estimate causal effects in observational studies. These methods aim to adjust for confounding factors between the treatment groups; in our case, different educational levels. The advantage of using the propensity score is that it enables us to summarize the many possible confounding covariates as a single score. Propensity score methods include matching, stratification on the propensity score, and inverse probability weighting (IPW) using the propensity score. The methods we apply are based on IPW methods.

The propensity score method to identify "treatment effects" relies on the *unconfoundedness* and the *common support* assumptions. The *unconfoundedness* assumption states that, conditional on observed individual characteristics, educational attainment is independent of the *potential outcomes* (months lost). This implies that (conditional on observed characteristics) the difference in the *potential outcome* if the individual had attained a low level of education and the *potential outcome* if the individual had attained a high level of education is only caused by education. This assumption requires that all of the variables that affect both mortality and education attainment are observed. Although this assumption is not testable and is clearly strong, it may be a reasonable approximation. The overlap (or common support) assumption requires that the propensity score, the conditional probability of attaining higher education given observed individual characteristics, is bounded away from zero and one. This assumption is, in principle, testable. If there are values of the included covariates for which the probability of observing a higher educational level is zero or one, we cannot compare these individuals between a high and a low educational level. In that case, we have to limit comparisons to sets of values in which there is sufficient control in the propensity score among individuals with both low and high levels of education. To calculate the propensity score, we could, in principle, estimate an ordered probit or ordered logit propensity score for our five ordered educational levels; see Imai and van Dyk (2004) and Feng et al. (2012). However, the men in the lowest and the highest educational groups differ too much in their observed background characteristics, which causes severe overlap problems. To address this overlap problem, we estimate separate propensity scores of attaining a higher educational level through pairwise comparisons (Lechner, 2002) of adjacent educational levels.

If *unconfoundedness* holds, all of the biases due to observable covariates can be removed by conditioning on the propensity score. The average educational gains can be estimated by weighting on the propensity score.

Inverse probability weighting based on the propensity score creates a synthetic sample in which the educational attainment is independent of the included covariates. The synthetic sample is the result of assigning to each individual a weight that is proportional to the inverse of his propensity score.

We implement an IPW version of the procedure for estimating the number of months lost based on pseudo-observations. The estimation of the IPW months-lost model is straightforward, and is based on an extension of the Stata procedure stplost of Overgaard et al. (2015). After calculating the weights based on the estimated propensity scores, we construct a pseudo-population. We then estimate for this pseudo-population the total survival function using a Kaplan-Meier estimate, and the cumulative incidence function using the Aalen-Johansen method. Next, we use these two quantities to estimate the base number of months lost for each cause of death. Finally, we regress these IPW pseudo-observations on the education indicator to obtain our estimate of the impact of education on the number of months lost.

One of the weaknesses of previous analyses is that the estimated educational effects may be biased by unobserved genetic or social factors that influence both the educational level and mortality. To eliminate this bias, we also include family fixed effects in the month-lost models.

3.3. Data

The data were retrieved from several Swedish population-wide registers, which were linked using a unique individual identification number assigned to all Swedish citizens. The Swedish Military Conscription Register (SMCR) includes demographic information (birthdate, area of residence), as well as the results of intelligence tests, a psychological assessment, and anthropometric and health measures (height, weight, blood pressure, and muscular strength). This SMCR was linked to the National Population and Housing Censuses (1960–1990), which contains information on the socioeconomic status and educational levels of the parents. Information on each conscript's own education was obtained from the Longitudinal Integration Database for Health Insurance and Labour Market Studies (LISA) for the 1990–2012 period, and information on cause-specific mortality as the underlying cause of death was obtained from the Cause of Death Register for the period up to 2012.

The study population consists of men born between 1950 and 1983 who were identified in the Multi-Generation Register, and who were conscripted into the military between 1969 and 2001; usually when they were aged 18–20. At that time in Sweden, military service was mandatory for men only. We selected only those men for whom at least one parent was known. We also removed men without a known conscription date. Finally, because we wanted to account for shared family influences, we only included men with (observed) brothers. We aggregated the observed education into five classes: (*i*) primary school (less than 10 years of primary school), (*ii*) secondary education (at most two years), (*iii*) full secondary education (2–3 years), (*iv*) post-secondary education (less than three years), and (*v*) higher education (university and PhD). We identified 700,043 men with at least one brother in 369,526 families.

Selected demographic and socioeconomic characteristics of the men at the time of their military examinations by educational level are given in Table 1. We see a clear relationship between parental socioeconomic status (SES), parental education, and the level of education attained by the conscript. The higher the social status and the educational level of the parent, the higher the educational level of the conscript. The results of intelligence and psychological fitness tests are also clearly related to the individual's attained education.

4. Results

Before we turn to the estimation results of the month-lost models, we discuss the results of the standard Cox proportional hazards models; i.e., the model that is commonly applied in the literature to analyze cause-specific competing risks (Elo et al., 2014; Kulhánová et al., 2014).

Table 2 reports the estimated hazard ratios of education (moving up one level) for each cause-specific outcome. We find that for most causes of death—with the main exception being some cancers—an improvement in education reduces the risk of dying from that cause. In the comparison of the men with the two lowest educational levels, we can see that only the risks of dying from certain cancers, stroke, and infectious diseases are not significantly lower among men with some secondary education. When we compare men with two and three years of secondary education, we see no differences in the cancer mortality rates. When we compare men with three years of secondary education and men post-secondary education, we find that digestive cancer is the only cancer for which mortality rates differ. Finally, when we compare men in the two highest educational levels, post-secondary and higher education; we find significant differences for two types of cancer, for a few other diseases (stroke, other CVD, and other natural causes), and for mortality due to external causes.

Next, we estimate the impact of education on the number of months lost due a specific cause of death. We focus on the number of months of life lost from age 18 (the age at which men undergo the military examination) to age 63 (the maximum age observed). The columns denoted by (1) in Table 3 provide the estimated

Table 1
Descriptive statistics.

	Primary	Secondary (2 years)	Education (3 years)	Post-secondary (<3 years)	Higher
SES mother at birth					
Non-manual (high)	1%	1%	2%	4%	8%
Non-manual (intermediate)	3%	4%	9%	12%	19%
Non-manual (low)	13%	18%	20%	28%	28%
Farmers	13%	9%	5%	6%	4%
Skilled workers	39%	38%	16%	20%	11%
Unskilled workers	15%	16%	18%	13%	10%
Not classified	16%	13%	30%	16%	20%
Unknown	1%	1%	1%	1%	1%
Education mother					
Primary (<9 yrs)	54%	45%	22%	23%	13%
Primary (9–10 yrs)	12%	11%	16%	11%	9%
Secondary education (2 yrs)	25%	32%	38%	33%	28%
Secondary education (3 yrs)	2%	3%	6%	7%	8%
Post-secondary	3%	5%	10%	13%	18%
Higher	2%	3%	7%	12%	23%
Unknown	2%	1%	1%	1%	1%
SES father at birth					
Non-manual (high)	2%	2%	5%	9%	18%
Non-manual (intermediate)	7%	11%	17%	25%	29%
Non-manual (low)	7%	8%	11%	11%	11%
Farmers	9%	7%	7%	5%	4%
Skilled workers	28%	30%	27%	21%	15%
Unskilled workers	40%	37%	28%	22%	15%
Not classified	6%	5%	5%	6%	7%
Unknown	1%	1%	1%	1%	1%
Education father					
Primary (<9 yrs)	58%	51%	29%	27%	16%
Primary (9–10 yrs)	7%	6%	12%	6%	5%
Secondary education (2 yrs)	17%	21%	26%	21%	17%
Secondary education (3 yrs)	7%	11%	16%	20%	18%
Post-secondary	2%	4%	7%	10%	11%
Higher	2%	3%	8%	15%	31%
Unknown	6%	3%	2%	2%	2%
mother < 20 at birth	10%	9%	5%	5%	3%
father > 40 at birth	8%	7%	5%	6%	5%
birth order	3.1	2.9	2.8	2.6	2.5
global IQ[a]	3.7	4.4	5.0	6.1	6.6
Psychological assessment[a]	4.1	4.7	5.1	5.6	5.8
# of individuals	91,588	219,372	138,653	101,883	148,547

[a] Stanine score 1–9, running from low to high.

results. The "primary" column reports the estimated months of life lost for the men with a primary education only. The next columns report the reduction in the number of months lost with each additional educational level.

According to the results, men with a primary education only lose a total of 23 months of life between ages 18 and 63. Among this group, obtaining two years of secondary education reduces the number of months lost by 8.5 months; getting another year of secondary education reduces the number of months lost by an additional 2.9 months; obtaining post-secondary education reduces the number of months lost by a further 3.5 months; and, finally, obtaining higher education reduces the number of months lost by another 2.5 months Only the months-lost analysis shows that the three external causes of death are far more relevant (for mortality between ages 18 and 63) than other causes of death for explaining the effects of educational gains (and for explaining total mortality). For example, when moving from the primary to the secondary level of education (2 years), the number of months lost due to suicide decreases by 1.4 months, whereas the number of months lost declines by just 0.6 months for all cancers and by 1.3 months for all cardiovascular diseases. The relative importance of external causes in the total gain in the number of months lost is above 50\% for all educational levels. Although the highest educational group only gains 1.8 months due to external causes, members of this group receive 71\% of the gains from education. Other causes of death that are strongly affected by educational level are ischemic heart diseases, psychiatric diseases (except among the highest educational level), and other CVDs.

4.1. Adjusting for confounding and family fixed effects

To adjust for confounding, we re-estimate the months-lost models for each specific cause of death using a re-weighted pseudo-population based on inverse propensity score weighting (see the supplementary appendix). We include variables that influence both the probability of obtaining a higher educational level and mortality. The control variables in the propensity score include parental (mothers and fathers separately) socioeconomic status (SES) at birth, parental education, whether the mother was young or the father was old at birth, IQ level, psychological assessment results, birth order, county dummies, and year-of-birth dummies.

The table with the coefficients of the four estimated logit models can be found in the supplementary appendix (Table B4). We tested whether the propensity score is able to balance the distribution of all of the included variables in each of the two adjacent educational groups using standardized bias calculations. We also checked for common overlap issues (see Table B5 in the supplementary appendix).

Table 2
Cox hazard estimates of the impact of education on cause-specific mortality (HR), for men with brothers.

	Secondary (2 years)	Education (3 years)	Post-secondary (<3 years)	Higher
Cancers				
Bowel cancer	0.951	0.824	0.898	1.171
Digestive cancer	0.859	0.963	0.661**	0.734
Lung cancer	0.733	0.932	0.724	0.551+
Leukemia	0.444**	1.025	1.262	0.497+
Lymphoma	1.113	0.841	1.226	0.829
Other cancer	0.903	0.918	0.925	0.867
CVD				
Ischemic heart disease	0.656**	0.794**	0.631**	0.836
Stroke	0.906	0.647**	0.799	0.569+
Other CVD	0.774**	0.908	0.702**	0.460**
Other diseases				
Infectious	0.795	0.389**	0.855	1.058
Endocrine	0.671	0.800	1.455	0.607
Diabetes	0.618**	0.628+	0.862	0.598
Psychiatric	0.491**	0.555**	0.261**	0.615
Nervous system	0.514**	0.995	0.759	0.652
Respiratory disease	0.570**	0.397**	0.439+	2.024
Digestive disease	0.490**	0.643	0.886	0.593
Liver	0.702**	0.520**	0.427**	0.767
Abnormalities	0.569**	0.900	0.685	0.988
Other natural causes	0.463**	1.295	0.583	0.227**
External causes				
Traffic accidents	0.776**	0.698**	0.634**	0.431**
Suicide	0.700**	0.723**	0.763**	0.592**
Other external causes	0.572**	0.652**	0.670**	0.444**

Separate models are estimated for each two adjacent education levels. +p < 0.05,**p < 0.01.

Based on the logit estimates for attaining a higher educational level, we calculate the propensity scores and weight the men with higher education by the inverse of the propensity score; and the men with lower education by the inverse of one, minus the propensity score (within each pair of educational levels in the four educational groups). However, only accounting for observed confounders is not sufficient, as there may still be some unobserved genetic or social factors that influence both educational levels and

Table 3
Months lost due to a specific cause at ages 18–63 by educational level, for men with brothers. (1) unadjusted model; (2) IPW model with family fixed effects.

	Educational gain									
	Primary		Secondary education				Post-secondary		Higher	
			(2 years)		(3 years)		(<3 years)			
	(1)[a]	(2)[a]	(1)	(2)	(1)	(2)	(1)	(2)	(1)	(2)
Cancers	3.28**	3.49**	0.61**	0.98+	0.20	0.45	0.28+	0.57	0.35**	0.44
Bowel cancer	0.34**	0.35**	0.02	0.04	0.05	0.07	0.02	0.02	−0.03	−0.02
Digestive cancer	0.72**	0.70**	0.14	0.16	0.03	0.12	0.18+	0.22	0.05+	0.01
Lung cancer	0.45**	0.42**	0.16+	0.14	0.01	−0.16	0.08	0.15	0.05+	0.11
Leukemia	0.28**	0.26**	0.18**	0.13	−0.00	−0.00	−0.03	−0.03	0.07+	0.04
Lymphoma	0.33**	0.34**	−0.05	−0.04	0.03	0.02	−0.04	−0.08	0.08	−0.03
Other cancer	1.16**	1.43**	0.16	0.55+	0.07	0.40	0.07	0.29	0.12	0.32
CVD	3.67**	4.06**	1.33**	2.07**	0.38**	0.74	0.63**	0.90+	0.31**	0.72
Ischemic heart disease	2.00**	2.22**	0.88**	1.30**	0.18+	0.37	0.33**	0.47+	−0.01	0.29
Stroke	0.44**	0.51**	0.07	1.18	0.12**	0.26	0.07	0.12	0.05	0.16
Other CVD	1.23**	1.34**	0.38**	0.59**	0.06	0.11	0.21	0.31	0.26**	0.27
Other diseases	4.98**	5.07**	2.53**	2.80**	0.67**	0.47	0.75**	0.59	0.11	0.25
Infectious	0.30**	0.29**	0.11+	0.11	0.10**	−0.02	0.03	0.06	0.00	0.10
Endocrine	0.15**	0.14**	0.06+	0.04	0.01	0.04	−0.03	0.09+	0.04	0.08
Diabetes	0.38**	0.33**	0.18**	0.11	0.07+	−0.01	0.03	0.01	0.03	0.07
Psychiatric	1.27**	1.45**	0.74**	1.01**	0.23**	0.32	0.23**	0.09	−0.01	0.01
Nervous system	0.61**	0.66**	0.33**	0.41**	−0.02	0.05	0.08	0.08	0.06+	0.11
Respiratory disease	0.60**	0.60**	0.34**	0.35+	0.13**	0.21	0.09+	0.03	−0.07	−0.11
Digestive disease	0.21**	0.20**	0.13**	0.12	0.02	0.04	0.02	0.04	0.01	0.01
Liver	0.67**	0.51**	0.24**	0.07	0.17**	−0.04	0.14**	−0.01	−0.02	−0.01
Abnormalities	0.59**	0.68**	0.29**	0.44****	−0.01	−0.12	0.10	0.15	0.02	−0.04
Other natural causes	0.20**	0.22**	0.11+	0.14	−0.03	0.00	0.06+	0.06	0.04+	0.03
External causes	10.70**	11.21**	4.10**	4.15****	1.71**	1.87+	1.86**	1.61	1.81**	1.60**
Traffic accidents	2.24**	2.37**	0.61**	0.49	0.36**	0.43	0.59**	0.12	0.54**	0.34+
Suicide	4.02**	4.04**	1.41**	1.27****	0.62**	0.59	0.60**	0.67	0.65**	0.71**
Other external causes	4.44**	4.80**	2.04**	2.39**	0.71**	0.85+	0.65**	0.83+	0.61**	0.55+
Total	22.63	23.84	8.53**	9.99**	2.88**	3.52**	3.53**	3.67**	2.54**	3.00**

+p < 0.05, **p < 0.01.
[a] (1) Unadjusted months lost model; (2) IPW with family fixed effects months lost model.

mortality. To control for this bias, we also account for family fixed effects in the weighted months lost models.

The results of these models are presented in the columns denoted by (2) in Table 3. Accounting for both observed confounders through the IPW method and for unobserved shared family factors through fixed effects increases the educational gains for all of the educational groups; especially for men with primary education only. For all of the educational groups, the gains in months lost due to CVDs increase after accounting for confounding factors. Still, the largest gains in months lost by educational level are due to external causes.

5. Conclusion and discussion

Many studies have found a large positive association between education and longevity. Previous evidence suggests that the impact of education on various diseases differs, and that these differences are reflected in educational gradients that diverge by cause of death. In line with the "fundamental cause theory," which posits that gains in education should lead to reductions in preventable diseases, differences by education in cardiovascular diseases have been reported. A major limitation of most of these studies is that they only reported associations, and ignored the possibility that part of the association between education and cause-specific mortality may be driven by confounding factors; i.e., by factors that influence both educational attainment and cause-specific mortality. It is important to note that using multivariate regression techniques is not sufficient to control for confounding, as those methods do not account for selective education attainment. However, after accounting for both pre-educational individual and shared family factors, we find rather small differences by educational attainment in cardiovascular mortality and death due to other preventable diseases. We find the largest educational gains in the reduction in mortality due to external causes.

In contrast to most other studies on this topic, we analyzed the months lost due to a specific cause of death instead of the cause-specific (proportional) hazards. The interpretation of the estimated education coefficients in standard Cox models can be difficult. Moreover, it is not possible to judge the contribution of the cause to total mortality from Cox models only. Using months lost due to a specific cause of death avoids these problems. Another advantage of this approach is that the month-lost results are, in contrast to hazard ratios, additive, both by education and by causes of death.

The inverse probability weighting (based upon the propensity score) method we employ accounts for the observed confounding factors and the unobserved shared family factors without making any assumptions regarding the relationship between these factors and mortality. Thus, an advantage of our method is that we do not need to make any functional form assumptions about the impact on mortality of the included control variables. Using a propensity score method accounts for the observed factors that influence both education and mortality. By focusing on men with brothers, we were also able to adjust for the unobserved family factors that remain the same across different births within a family; i.e. the so-called family fixed effects.

Based on standard Cox proportional hazard models, we find, in line with the literature, many significant differences by educational attainment in cause-specific mortality. Using months-lost analyses and accounting for confounding factors reduces the number of causes with a significant educational gradient. In contrast to previous findings, we obtained hardly any significant educational gains for death due to cancers or cardiovascular diseases. Importantly, we found that death due to external causes played a much larger role in explaining the differences by educational attainment. We found the largest educational gains for the lowest educational group.

To illustrate the size of the estimated educational gains, we calculated how many men-years the 50,000 men born in 1951 would gain if the 1983 education distribution had applied to them. Based on the IPW with family fixed effects, we found that the 1951 cohort would gain 423 person-years if they had the 1983 education distribution, which is a 22% reduction in the person-years between ages 18 and 63 that were lost to death. Most of this gain, or 248 person-years (a 28% reduction), is attributable to reduced mortality due to external causes (traffic accidents, 57 person-years; suicides, 88 person-years; and other external causes, 103 person-years). The contribution of the reduction in cancer mortality (28 person-years, or an 8% reduction) and CVD mortality (53 person-years, or an 18% reduction; of which 27 person-years were due to IHD) is rather small. The reduction of death from other diseases leads to 94 person-years (a 28% reduction) saved (30 person-years due to psychiatric diseases, or a 40% reduction).

Ignoring confounding factors would underestimate the educational gains for the less educated men, and especially for the conscripts with primary education only. This implies that at the lower end of the education distribution, men with secondary education are negatively selected (compared to men with primary education only), and thus have higher mortality. This selection is mainly driven by differences in unobserved shared family factors (see the supplementary appendix). This negative selection is most prominent for death due to CVDs. The negative selection is linked to lower levels of education in families with lower mortality. The observed family factors that lead to higher levels of education, such as high levels of parental SES and education, seem to play relatively minor roles in explaining the gains in education.

Our study has four distinct strengths compared to previous research. First, a clear advantage of the study is our very large sample size (700,000), which allows for the estimation of the educational gains for five educational levels and 22 causes of death. Second, the data are population-based and are not prone to self-selection issues, because military conscription was mandatory in Sweden during the study period. Third, the statistical method we use, inverse propensity score weighting with family fixed effects, accounts for confounding effects on cause-specific mortality. This approach enables us to draw more accurate causal conclusions, without encountering the generalization issues inherent in relying on compulsory schooling reforms to account for confounding. Fourth, contrary to the standard literature on causes of death (competing risks) analysis, we define the impact of educational gains on causes of death in terms of the number of months lost due to each specific cause of death, instead of using the hazard ratio. This number can be interpreted more naturally, it provides a direct measure of the importance of the specific cause, and it avoids the issues of dependence in competing risks proportional hazard models. The number of months lost can be defined over a segment of the age distribution, and it is an additive measure. The sum over all alternative causes of death within a single educational level is equal to the total number of months lost (and the educational gain) for that educational level, and the sum of educational gains over all education levels within a cause of death is equal to the total impact of education on that cause of death.

A limitation of our data, which are based on information collected in a military entrance examination, is that we do not have information on women. Another limitation is that we only observe mortality before the age of 63 (or earlier for cohorts from the 1970s and the 1980s). In the future, when these men have been followed for a longer period, we expect to find that mortality due to cardiovascular diseases and cancers plays a larger role, while mortality due to current external causes plays a smaller role. A final drawback

is that although military conscription was mandatory in Sweden, men with severe mental disabilities or severe chronic diseases were exempt from the military examination. Thus, our results only apply to those who did not have severe mental or chronic conditions at age 18.

Our IPW model with family fixed effects controls for unobserved factors at the family level. A drawback of using a family fixed effects model is that it relies on within-brother variation in education only (and other covariates in the IPW) to identify educational gains. Brothers with the same educational levels do not contribute to the estimation, nor do brothers whose educational levels differ by more than one level. If the main reason for the difference between the educational levels of the brothers is a difference in health, the family fixed effects model will overestimate the educational gain. This is especially likely be the case among affluent families at the upper end of the education distribution. Another issue is that there may be unobserved differences between brothers in early life health that could have influenced both schooling and mortality. First, we assume a time-invariant family effect, excluding any impact of a change in family income and wealth for a later-born brother. Second, although we include important individual confounders, such as cognitive ability and psychological fitness, we may have nonetheless excluded individual characteristics that influence both schooling and mortality, such as low birth weight, that we could not observe.

Acknowledgements

MM was supported by the European Research Council Grant 336475 (COSTPOST).

Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.socscimed.2017.05.019.

References

Albouy, V., Lequien, L., 2009. Does compulsory education lower mortality? J. Health Econ. 28 (1), 155–168.
Amin, V., Behrman, J.R., Kohler, H.P., 2015. Schooling has smaller or insignificant effects on adult health in the US than suggested by cross-sectional associations: new estimates using relatively large samples of identical twins. Soc. Sci. Med. 127, 181–189.
Andersen, P.K., 2013. Decomposition of number of life years lost according to causes of death. Stat. Med. 32, 5278–5285.
Andersen, P.K., Canudas-Romo, V., 2013. Cause-specific measures of life years lost. Demogr. Res. 29, 1127–1152.
Arendt, J.N., 2005. Does education cause better health? A panel data analysis using school reforms for identification. Econ. Educ. Rev. 24 (2), 149–160.
Austin, P.C., 2014. A tutorial on the use of propensity score methods with survival or time-to- event outcomes: reporting measures of effect similar to those used in randomized experiments. Stat. Med. 33, 1242–1258.
Barclay, K., Keenan, K., Grundy, E., Kolk, M., Myrskylä, M., 2016. Reproductive history and post- reproductive mortality: a sibling comparison analysis using Swedish register data. Soc. Sci. Med. 155, 82–92.
Bijwaard, G.E., Myrskylä, M., Tynelius, P., Rasmussen, F., 2016. Education, Cognitive Ability and Cause-specific Mortality: a Structural Approach. MPIDR. Working Paper 2016-007.
Bijwaard, G.E., van Kippersluis, H., Veenman, J., 2015a. Education and health: the role of cognitive ability. J. Health Econ. 42, 29–43.
Bijwaard, G.E., van Poppel, F., Ekamper, P., Lumey, L.H., 2015b. Gains in life expectancy associated with higher education in men. PLoS One 10, e0141200.
Borrell, C., Plasència, A., Huisman, M., Costa, G., Kunst, A., Andersen, O., Bopp, M., Borgan, J.-K., Deboosere, P., Glickman, M., Gadeyne, S., Minder, C., Regidor, E., Spadea, T., Valkonen, T., Mackenbach, J.P., 2005. Education level inequalities and transportation injury mortality in the middle aged and elderly in European settings. Inj. Prev. 11 (3), 138–142.
Clark, D., Royer, H., 2013. The effect of education on adult mortality and health: evidence from Britain. Am. Econ. Rev. 103 (6), 2087–2120.
Cole, S.R., Hernán, M.A., 2004. Adjusted survival curves with inverse probability weights. Comput. Methods Progr. Biomed. 75, 45–49.
Conti, G., Heckman, J.J., Urzua, S., 2010. The education-health gradient. Am. Econ. Rev. 100, 234–238.
Cutler, D.M., Lleras-Muney, A., Vogl, T., 2011. Socioeconomic status and health: dimensions and mechanisms. In: Glied, S., Smith, P.C. (Eds.), The Oxford Handbook of Health Economics. Oxford, Oxford UP, pp. 124–163.
Elo, I.T., Martikainen, P., Myrskylä, M., 2014. Socioeconomic status across the life course and all-cause and cause-specific mortality in Finland. Soc. Sci. Med. 119, 198–206.
Feng, P., Zhou, X.-H., Zou, Q.-M., Fan, M.-Y., Li, Z.-S., 2012. Generalized propensity score for estimating the average treatment effect of multiple treatments. Stat. Med. 31, 681–697.
Fletcher, J.M., 2015. New evidence of the effects of education on health in the US: compulsory schooling laws revisited. Soc. Sci. Med. 127, 101–107.
Galobardes, B., Lynch, J.W., Smith, G.D., 2004. Childhood socioeconomic circumstances and cause-specific mortality in adulthood: systematic review and interpretation. Epidemiol. Rev. 26 (1), 7–21.
Grossman, M., 2015. The relationship between health and schooling: what's new? Nordic J. Health Econ. 3 (1), 7–17.
Huisman, M., Kunst, A.E., Bopp, M., Borgan, J.-K., Borrell, C., Costa, G., Deboosere, P., Gadeyne, S., Glickman, M., Marinacci, C., Minder, C., Valkonen, T., Mackenbach, J.P., 2005. Educational inequalities in cause-specific mortality in middle-aged and older men and women in eight western European populations. Lancet 365 (9458), 493–500.
Hummer, R.A., Lariscy, J.T., 2011. Educational attainment and adult mortality. In: Rogers, R.G., Grimmins, E.M. (Eds.), Int. Handb. Adult Mortal., 2. Springer, Dordrecht, The Netherlands, p. 260.
Imai, K., van Dyk, D.A., 2004. Causal inference with general treatment regimes: generalizing the propensity score. J. Am. Stat. Assoc. 99, 854–866.
Kulhánová, I., Hoffmann, R., Eikemo, T.A., Menvielle, G., Mackenbach, J.P.J.P., 2014. Educational inequalities in mortality by cause of death: first national data for The Netherlands. Int. J. public health 59 (5), 687–696.
Lechner, M., 2002. Some practical issues in the evaluation of heterogeneous labour market programmes by matching methods. J. R. Stat. Soc. Ser. A 165, 59–82.
Leuven, E., Plug, E., Rønning, M., 2016. Education and cancer risk. Labour Econ. 43, 106–121.
Link, B.G., Phelan, J.C., 1995. Social conditions as fundamental causes of disease. J. Health Soc. Behav. Extra issue 80–94.
Lleras-Muney, A., 2005. The relationship between education and adult mortality in the United States. Rev. Econ. Stud. 72, 189–221.
Lorant, V., Kunst, A.E., Huisman, M., Costa, G., Mackenbach, J.P., 2005. Socio-economic inequalities in suicide: a European comparative study. Br. J. psychiatry 187 (1), 49–54.
Lundborg, P., 2013. The health returns to schooling: what can we learn from twins? J. Popul. Econ. 26 (2), 673–701.
Mackenbach, J.P., 2012. The persistence of health inequalities in modern welfare states: the explanation of the paradox. Soc. Sci. Med. 75, 761–769.
Masters, R.K., Hummer, R.A., Powers, D.A., 2012. Educational differences in US adults mortality: a cohort perspective. Am. Sociol. Rev. 77 (4), 548–572.
Mazumder, B., 2008. Does education improve health: a reexamination of the evidence from compulsory schooling laws. Econ. Perspect. 33, 2–16.
Meghir, C., Palme, M., Simeonova, E., 2013. Education, Cognition and Health: Evidence from a Social Experiment. NBER. Working paper 19002.
Overgaard, M., Andersen, P.K., Parner, E., 2015. Regression analysis of censored data using pseudo-observations: an update. Stata J. 15, 809–821.
Thomas, J.M., 1996. On the interpretation of covariate estimates in independent competing risks models. Bull. Econ. Res. 48, 27–39.
Van Kippersluis, H., O'Donnell, O., van Doorslaer, E., 2011. Long run returns to education: does schooling lead to an extended old age? J. Hum. Resour. 46 (4), 695–721.