An IPW estimator for mediation effects in hazard models: with an application to schooling, cognitive ability and mortality[*]

Govert E. Bijwaard[†]
NIDI

Andrew M. Jones[‡]
University of York

February 9, 2018

Abstract

Large differences in mortality rates across those with different levels of education are a well-established fact. Cognitive ability may be affected by education so that it becomes a mediating factor in the causal chain. In this paper we estimate the impact of education on mortality using inverse probability weighted (IPW) estimators. We develop an IPW estimator to analyse the mediating effect in the context of survival models. Our estimates are based on administrative data, on men born between 1944-1947 who were examined for military service in the Netherlands between 1961-1965, linked to national death records. For these men we distinguish four education levels and we make pairwise comparisons. The results show that levels of education have hardly any impact on the mortality rate. Using the mediation method we only find a significant effect of education on mortality running through cognitive ability, for the lowest education group that amounts to a 15% reduction in the mortality rate. For the highest education group we find a significant effect of education on mortality through other pathways of 12%.

JEL classification: C41, I14, I24.

Keywords: Education; Mortality; Inverse probability weighting; Mediators; Mixed proportional hazard

[†]Netherlands Interdisciplinary Demographic Institute (NIDI-KNAW/University of Groningen), PO Box 11650, 2502 AR The Hague, the Netherlands and IZA, Bonn, +31 70 3565224, `bijwaard@nidi.nl`

[‡]Department of Economics and Related Studies, University of York, YO10 5DD, United Kingdom; Centre for Health Economics, Monash University, Australia `andrew.jones@york.ac.uk`

1 Introduction

Traditionally, causal mediation analysis has been formulated within the framework of linear structural models (Baron and Kenny 1986). These models are difficult to extend to inherently non-linear duration outcomes such as the mixed proportional hazard model. Recent papers have placed causal mediation analysis within the counterfactual/potential outcomes framework (Imai et al. 2010; Imai et al. 2010; Huber 2014, VanderWeele 2015) all assuming sequential unconfoundedness. Tchetgen Tchetgen (2013) also introduced a weighting method by for mediation analysis in a Cox proportional hazard model. His method implies estimating a regression model for the mediator conditional on the treatment and pre-treatment covariates, while our method is based on estimating the propensity score (with and without the mediator). In general it is more difficult to formulate a suitable model for the mediator than for the propensity score.

Our outcome, the age at death is a duration variable and the mortality hazard rate, the instantaneous probability that an individual dies at a certain age conditional on surviving up to that age, is modelled. Accounting for right-censoring, when the individual is only known to have survived up to the end of the observation window, and left-truncation, when only those individuals are observed who were alive at a certain time, are easy to handle in hazard models (Van den Berg 2001). A common way to accommodate the presence of observed characteristics is to specify a proportional hazard model, in which the hazard is the product of the baseline hazard, the age dependence, and a log-linear function of covariates. Neglecting confounding in inherently non-linear models, such as proportional hazard models, leads to biased inference.

Propensity score methods are increasingly used to take account of confounding in observational studies, e.g. see Caliendo and Kopeinig (2008) for a survey. The advantage of the propensity score is that it enables us to summarize the many possible confounding covariates as a single score (Rosenbaum and Rubin 1983). With a duration outcome, right censoring makes inference of differences in means, as is standard in treatment analysis, unreliable. Propensity score methods for hazard models have been introduced for duration data that account for censoring, truncation and dynamic selection issues (Cole and Hernán 2004; Austin 2014). We apply inverse probability weighting (IPW) methods using the propensity score (Hirano et al. 2003), which belongs to the larger class of marginal structural models that account for time-varying confounders when estimating the effect of time-varying covariates (Robins et al. 2000).

Cognitive ability can be considered a principal source of education selection and, an endowment that determines success at school. Then intelligence precedes education in the causal path to health and mortality. However, cognitive ability, at least as measured by standard IQ-tests, is likely to change with the education attained. Recent research (Falch and Massih 2011; Banks and Mazzonna 2012; Schneeweis et al. 2014; Carlsson et al. 2015; Dahmann 2017) has shown that additional education improves cognitive ability. In that case, cognitive ability is a mediator in the causal path from education to health. Ideally, we would have continuous measurement of the (development) of cognitive ability over the life cycle, to account for both the selection and mediation of cognitive ability in the causal path from education to mortality. However, in our data, we only observe cognitive ability at late adolescence when measured intelligence can be either the result of the attained education or a proxy of early childhood intelligence which influences education choice. When cognitive ability is a medi-

1

ator we can decompose the effect of education on mortality into an effect running through improvement of cognitive ability and an effect through other pathways. An effect of education through improvement of cognitive ability is likely if education raises cognitive ability that aids disease management and in seeking appropriate treatment where necessary. Other possible pathways from education to mortality emerge if higher education leads to improvement in socioeconomic status later in life, such as labour market signals, non-cognitive skills and peer effects, which influence health and mortality.

In our empirical analyses we use administrative data on Dutch men who were examined for military service in the Netherlands between 1961-1965 after completing their secondary schooling. We followed 39,803 men selected from the national birth cohorts 1944-1947. These examinations are based on yearly listings of all Dutch male citizens aged 18 years in the national population registers. The sampled examination records were linked by Statistics Netherlands to recent national death records (up till the end of 2015). The records include a standardized recording of demographic and socioeconomic characteristics such as education, father's occupation, religion, family size, and birth order, along with a standardized psychometric test battery. The educational level was classified in four categories: primary school, lower vocational education, lower secondary education, and intermediate vocational education, general secondary education, higher non-university and university education.

Under the assumption that cognitive ability is a mediator of the education effect on mortality we also extend the IPW methods to mediation analysis for a (mixed) proportional hazard (MPH) model, the common model for econometric duration analysis. The main methodological contribution of this paper is that we disentangle the total effect of a treatment on a duration into an effect that runs through the mediator and an effect through other pathways. We derive and implement an IPW estimator for such a decomposition of the total effects in MPH models. The estimator identifies causal mechanisms given that a sequential unconfoundedness condition holds. This is a strong assumption and nonrefutable. We therefore carry out a set of sensitivity analyses to quantify the robustness of our empirical findings to violation of the sequential ignorability assumption. We focus, in particular, on how the possibility of selection into education based on cognitive ability may influence our results.

The empirical results show that improving education has hardly any impact on the mortality rate when accounting for cognitive ability. Using the mediation method we only find a significant effect of education on mortality running through cognitive ability, for the lowest education group that amounts to a 15% reduction in the mortality rate. For the highest education group we find a significant effect of education on mortality through other pathways of 12%.

2 Methods

2.1 The mortality hazard rate

We seek to find the impact of education level on the mortality risk for the men in our sample of conscripts. However, mortality may be influenced by factors that also determine the education choice. This may render education a selective choice and makes it endogenous to mortality later in life. We follow a propensity score method to account for selection on observed characteristics and estimate the effect of education on the mortality rate. Figure 1 provides a graphical illustration of the relationship between cognitive ability, education and mortality later in life using a directed acyclic graph, where each arrow represents a causal path (Pearl 2000; Pearl 2012). It states that early childhood characteristics X, such as parental background and family size, influence the education choice D, the unmeasured childhood (pre-age 18) factors, U_0, and the cognitive ability at age 18, Q_{18}. The latter is also influenced by other childhood factors, which may include early life cognitive ability, and the education followed up to age 18. In our data we do not observe these childhood factors (U_0).
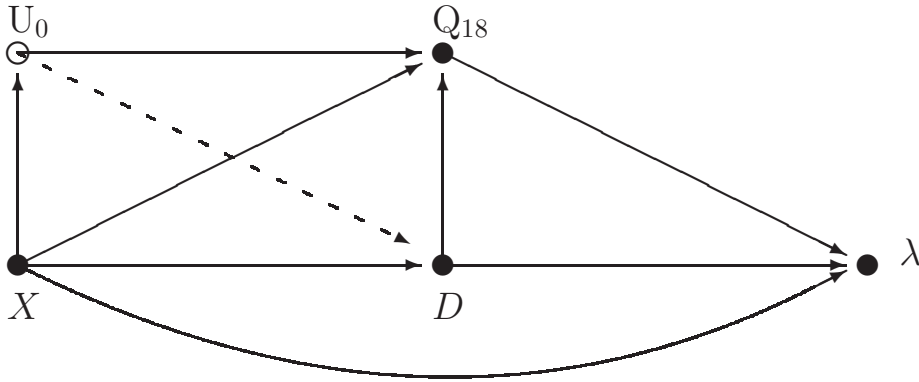

Figure 1: Directed acyclic graph of mediation through Q_{18} conditional on X

We define the treatment effect, of moving up one education level, in terms of a proportional change in the (mortality) hazard rate. First, we discuss the assumptions, common in the potential outcomes literature that uses propensity score methods, to identify the impact of education on the mortality risk. In Section 2.2 we extend this to decompose the effect of education on the mortality rate into an effect running through improvement in cognitive ability and an effect running through other pathways. The main difference with standard propensity score methods is that we use potential hazard rates, the hazard rate that would be observed if the individual was untreated, $\lambda(t|0)$, or treated $\lambda(t|1)$. Let $D_i = 1$ be the treatment, moving up one education level. We observe pre-treatment (educational level) covariates X that influence the education choice.

Assumption 1. *Unconfoundedness:* $\lambda(t|d) \perp D|X$ for $d = 0, 1$

where $\perp$ denotes independence. The unconfoundedness assumption (Rubin 1974, Rosenbaum and Rubin 1983) asserts that, conditional on covariates X, treatment assignment (education level) is independent of the potential outcomes. This assumption requires that all variables that affect both the mortality and the education choice are observed. Note that this does not imply that we assume all relevant covariates are observed. Any missing factor is allowed to influence either the outcome or the education choice, not both. We check the robustness of our estimates to this, rather strong, unconfoundedness assumption by assessing to what extent the estimates are robust to violations of this assumption induced by including an additional simulated binary variable to capture unobservables (Nannicini 2007; Ichino et al. 2008).

The overlap, or common support assumption requires that the propensity score, the conditional probability to choose a higher education given covariates X, is bounded away from zero and one. In our data we distinguish four (ordered) education levels in line with the contemporary Dutch education system (see Section 3). By comparing only adjacent education levels we remove the overlap problems.

Rosenbaum and Rubin (1983) show that if the potential outcomes are independent of treatment conditional on covariates X, they are also independent of treatment conditional on the propensity score, $p(x) = \Pr(D = 1 | X = x)$. Hence if unconfoundedness holds, all biases due to observable covariates can be removed by conditioning on the propensity score (Imbens 2004). The average effects can be estimated by matching or weighting on the propensity score. Here we use weighting on the propensity score. Inverse probability weighting based on the propensity score creates a pseudo-population in which the education choice is independent of the measured confounders. The pseudo-population is the result of assigning to each individual a weight that is proportional to the inverse of their propensity score. Inverse probability weighting (IPW) estimation is usually based on normalized weights that add to unity.

$$W_i = \left[\frac{D_i}{\hat{p}(X_i)} \middle/ \sum_{j=1}^{n} \frac{D_j}{\hat{p}(X_j)} \right] + \left[\frac{(1 - D_i)}{1 - \hat{p}(X_i)} \middle/ \sum_{j=1}^{n} \frac{1 - D_j}{1 - \hat{p}(X_j)} \right] \tag{1}$$

In survival analysis it is standard to compare the (non-parametric) Kaplan-Meier curves for the treated and the controls. The unadjusted survival curves may be misleading due to confounding. Cole and Hernán (2004) describe a method to estimate the IPW adjusted survival curves. Biostatisticians usually focus on Cox regression models and Cole and Hernán (2004) describe how Cox proportional hazard models can be weighted by the inverse propensity score to estimate causal effects of treatments. This method is related to the g-computation algorithm of Robins and Rotnitzky (1992) and Robins et al. (2000).

In economics the interest is often also in the duration dependence of the hazard. The Gompertz hazard, which assumes that the hazard increases exponentially with age, $\lambda_0(t) = e^{\alpha_0 + \alpha_1 t}$, is known to provide accurate mortality hazards (Gavrilov and Gavrilova 1991). However, it is hardly ever possible to include all relevant factors, either because the researcher does not know all the relevant factors or because it is not possible to measure then. Ignoring such unobserved heterogeneity or frailty may have a huge impact on inference in proportional hazard models, see e.g. Van den Berg (2001). A common solution is to use a Mixed Proportional Hazard (MPH) model, in which it is assumed that all unmeasured factors and measurement error can be captured in a multiplicative random term V. The hazard rate becomes

$$\lambda(t | D, V) = V \lambda_0(t) \exp(\gamma D), \tag{2}$$

4

The (random) frailty $V > 0$ is time-invariant and independent of the observed characteristics X and treatment D. Note that independence of V and D is crucial, otherwise Assumption 1 would be violated. So, we assume that some factors influencing the mortality rate are not observed and that these factors do not influence the education choice. In the empirical application it is assumed that V has a gamma-distribution; a common assumption used in the empirical literature.

To adjust for confounding we estimate a standard MPH model, that does not include the measured confounders as covariates, using the re-weighted pseudo-population. Fitting a (mixed) proportional hazard model in the pseudo-population is equivalent to fitting a weighted MPH model in the original sample. The parameters of such weighted MPH models can be used to estimate the causal effects of education on mortality in the original sample. The IPW estimator in the (M)PH model is equivalent to solving the weighted derivatives of the log-likelihood:

$$L(\theta) = \sum_{i=1}^{N} W_i \left[\delta_i \frac{\partial \log \lambda(t_i|\cdot)}{\partial \theta} - \frac{\partial \Lambda(t_i|\cdot)}{\partial \theta} \right] \tag{3}$$

where θ is the vector of parameters of the hazard in (2), $\Lambda(t|\cdot) = \int_0^t \lambda(s|\cdot)\,ds$, the integrated hazard and δ indicates whether the duration for individual i is censored $\delta_i = 0$ or not.[1]

2.2 Mediation analysis for the mortality hazard rate

In this section we discuss a model in which cognitive ability measured at age 18 mediates the impact of education on mortality. Mediation analysis aims to unravel the underlying causal mechanism into an effect running through changes of an intermediate variable, the *mediator*, and through other pathways. The counterfactual notation for average treatment effects can be extended to define causal mediation, (see Huber 2014). We are particularly interested in the mediating effect of cognitive ability on mortality. It has been proven that high levels of cognitive ability is positively associated with high education, (Ceci 1991; Hansen et al. 2004). Recent research (Falch and Massih 2011; Banks and Mazzonna 2012; Schneeweis et al. 2014; Carlsson et al. 2015; Dahmann 2017) has shown that one additional year of education improves intelligence up to 0.3 standard deviations, both for the US and for some European countries. We use Q_i to denote the observed cognitive ability (IQ-score), which is measured around age 18 when the men had their military examination and after they had completed secondary schooling. The mediation model we assume is illustrated by the DAG in Figure 1.

Traditionally, causal mediation analysis has been formulated with the framework of linear structural models (Baron and Kenny 1986). Recent papers have placed causal mediation analysis within the counterfactual/potential outcomes framework (Imai et al. 2010; Imai et al. 2010; Huber 2014). In the previous section the potential outcome was solely a function of the treatment, e.g. education choice, but in mediation analysis the potential outcomes also depend on the mediator. Because cognitive ability can be affected by the education attained[2], there exist two potential values, $Q_i(1)$ and $Q_i(0)$, only one of which will be observed, i.e. $Q_i = D_i \cdot Q_i(1) + (1 - D_i) \cdot Q_i(0)$. For example, if individual i actually attained education

---

[1]In Appendix A we provide a counting process interpretation and prove consistency.

[2]For example, Jones et al. (2011) discuss how performance in IQ tests could be influenced by coaching received by primary school pupils to prepare them for entrance tests for secondary school.

level 1, we would observe $Q_i(1)$ but not $Q_i(0)$. Next we use $\lambda_i\big(t|d, q(d)\big)$ to denote the potential mortality hazard that would result from education equals d and cognitive ability equals q. For example, in the conscription data, $\lambda_i\big(t|1, 110\big)$ represents the mortality hazard that would have been observed if individual i had education level 1 and a measured IQ-score of 110. As before, we only observe one of the multiple hazards $\lambda_i = \lambda_i\big(t|D_i, Q_i(D_i)\big)$.

Because we base our treatment effect on (mixed) proportional hazard models, it is again natural to define the mediator effects proportionally. Abbring and van den Berg (2003) also define, in a different setting with a dynamic treatment, a proportional treatment effect for a duration outcome. In other non-linear settings, such as count data regression, a proportional treatment effect has been defined (Lee and Kobayashi 2001). We define the average effect of other pathways, depending on treatment status d:

Assumption 2. *Proportional decomposition*

$$\theta(d) = \frac{\mathrm{E}\Big[\lambda\big(t|1, Q(d)\big)\Big]}{\mathrm{E}\Big[\lambda\big(t|0, Q(d)\big)\Big]} \tag{4}$$

This framework enables us to disentangle the underlying causal pathway from education to mortality into an effect of education through improvement of cognitive ability and an effect through other pathways. We assume conditional independence (given X) of the treatment and the mediator:

Assumption 3. *Sequential ignorablility:* $\{\lambda(t|d', q), Q(d)\} \perp D|X$ and $\lambda(t|d', q) \perp Q|D = d, X$, $\forall d, d' = 0, 1$ and q in the support of Q.

The first condition of Assumption 3 implies that, conditional on observed covariates X, no unobserved confounder exists that jointly affects the education choice, the cognitive ability and the mortality. The second condition implies that, conditional on observed covariates X and the education attained, no unobserved confounder exists that jointly affects cognitive ability and mortality. This would imply that X explains all the variation in U_0 or that U_0 does not (directly) affect education, the dashed line in Figure 1 Huber (2014) and Imai et al. (2010) make the same assumptions for identification of the direct and indirect effects in a linear model. Assumption 3 is a strong assumption and nonrefutable. We therefore carry out a set of sensitivity analyses to quantify the robustness of our empirical findings to violation of the sequential ignorability assumption based on an extension of the sensitivity analyses of Nannicini (2007) and Ichino et al. (2008). We focus, in particular, on how the possibility of selection into education based on cognitive ability may influence our results. We also have a common support restriction for the propensity score including the mediator.

In addition we assume independent censoring[3] and a proportional mediator effect $\theta(d)$:

Assumption 4. *Independent censoring:*
Censoring is, conditional on the treatment D, independent of the covariates X, the outcome

---

[3]In principle it is possible to extend the method to the assumption that censoring is independent of the outcome conditional on the treatment, the covariates and the mediator using a similar weighting for the censoring.

T and the mediator Q.

Assumption 5. *Proportional mediator effect:* $\lambda\big(t|1, Q(d)\big) = e^{\theta(d)}\lambda\big(t|0, Q(d)\big).$

This is equivalent to assuming that the effect of the treatment, D, is not moderated by the value of the mediator. Thus, we assume no interaction effect, $D \cdot Q$, in the hazard. Note that Assumption 5 does not rule out an MPH model. It only assumes that the unobserved heterogeneity is independent of the treatment D (as before) and the mediator Q. This leads to the following identification theorem for the effect of a treatment on the hazard running through other pathways (holding the mediator constant):

Theorem 1: Identification of other pathways effect $\theta(d)$.
Under Assumptions 1 to 5 the other pathways effect is identified through a weighted MPH regression with weights:

$$W(d) = \frac{\Pr(D = d|Q, X)}{\Pr(D = d|X)}\left(\frac{D}{\Pr(D = 1|Q, X)} + \frac{1 - D}{\Pr(D = 0|Q, X)}\right) \tag{5}$$

with weight $W(d)$ for $\theta(d)$, for $d = 0, 1$.
(See Appendix A for the proof.)

The 'total effect' of education on the mortality rate, from an IPW estimation in which the mediator is excluded from the propensity score, can be decomposed into an effect of education running through the mediator $\eta(\cdot)$ and an effect of education running through other pathways $\theta(\cdot)$ using assumption 2:

$$\frac{\lambda\big(t|D = 1, Q(1)\big)}{\lambda\big(t|D = 0, Q(0)\big)} = \frac{\lambda\big(t|D = 1, Q(1)\big)}{\lambda\big(t|D = 0, Q(1)\big)} \cdot \frac{\lambda\big(t|D = 0, Q(1)\big)}{\lambda\big(t|D = 0, Q(0)\big)} = \exp\Big(\theta(1) + \eta(0)\Big) \tag{6}$$

$$= \frac{\lambda\big(t|D = 1, Q(1)\big)}{\lambda\big(t|D = 1, Q(0)\big)} \cdot \frac{\lambda\big(t|D = 1, Q(0)\big)}{\lambda\big(t|D = 0, Q(0)\big)} = \exp\Big(\eta(1) + \theta(0)\Big) \tag{7}$$

The effect running through other pathways (holding the mediator constant) can be estimated solving (3), using $W(d)$ from (5) as weights. The effect running through the mediator can be obtained from the log-difference of the estimated total and the estimated effect running through other pathways, using (6) or (7). The first effect represents the effect of education on the mortality hazard while holding cognitive ability constant at the level that would have been realized for chosen education level d. The second effect represents the effect of education on mortality if one changes cognitive ability from the value that would have been realized for education level 0 to the value that would have been observed for education level 1, while holding the education level at level d.

For estimation we use normalized versions of the sample implied by the weights in (5), such that the weights in either treatment or control groups add up to unity, as advocated earlier. We estimate the additional propensity scores conditional on the pre-treatment covariates and the mediator, $\Pr(D = 1|X_i, Q_i)$, by probit specifications.

A nice feature of Theorem 1 is that it is straightforward to implement, and only involves estimation of two propensity scores and plugging them into standard mixed proportional

hazard estimation. No parametric restriction is imposed on the model of the mediator. Tchetgen Tchetgen (2013) also defines mediation analysis in (Cox) proportional hazard models. His method, which is also based on proportional decomposition, sequential ignorability, independent censoring and a proportional mediator effect, implies estimating a regression model for the mediator conditional on the treatment and pre-treatment covariates $f(Q|D, X)$, while our method is based on estimating the propensity score (with and without the mediator). In general it is more difficult to formulate a suitable model for the mediator than for the propensity score. VanderWeele (2011) also derived a mediation estimator for the Cox proportional hazards model. Although his method does not need assumption 5, a proportional mediator effect, it requires an additional assumption that the outcome is rare over the entire follow-up period.

3 Data

Data from a large sample from the nationwide Dutch Military Service Conscription Register for the years 1961-1965 and male birth cohorts 1944-1947 are analysed. All men, except those living in psychiatric institutions or in nursing institutes for the blind or for the deaf-mute, were called to a military service induction exam. The majority attended the conscription examination around age 18.[4] We have information from the military examinations for 45,037 men. The data were described elsewhere, (Ekamper et al. 2014), here we provide the main characteristics. These data were linked to the Dutch death register through to the end of 2015 using unique personal identification numbers. Follow-up status was incomplete (due to emigration and other right-censoring events) for 1,316 (2.9%) and entirely unknown for 2,626 (8.3%) men.[5] The latter were removed from the data. These data allow us to follow a large group of men from age 18 until age 68–72 or until death. At the military examination a standardized recording of demographic and socioeconomic characteristics such as education, father's occupation, religion, family size, region of birth, and birth order is recorded. We exploit the information on education attained at age 18 and the age at death to investigate the mortality difference while accounting for other factors that influence both educational level and mortality.

The educational level is classified in four categories[6], (Doornbos and Kromhout 1990): primary school (age 6-12 years); lower vocational education (two years post primary school); lower secondary education (four years post primary school); and higher education (intermediate vocational education, general secondary education, higher non-university and university education , i.e. at least six years post primary school). For this study, we excluded partly institutionalized conscripts who had attended special schools for those with disabilities or learning difficulties and conscripts who had not completed 6 years of schooling. After exclusion of these 2,608 conscripts, 39,803 men remain for analysis.

---

[4]Many men who continued to higher education were examined in their 20s.

[5]Table B.1 in the Appendix shows that some of differences between the sample we used and those that were removed is significant. We address this issue in Section 3.3, see Table B.4.

[6]Education in the Netherlands is characterized by years of education and by school level. There are two parallel streams in the educational system: general academic and vocational. Streaming choices are made at the end of primary school. Students in the vocational stream cannot directly enter university. Students with more than twelve years of education will nearly always be in the academic stream (Schröder and Ganzeboom 2014; Vrooman and Dronkers 1986).

A standardized psychometric test battery is included: comprising Raven Progressive Matrices, a nonverbal untimed test that requires inductive reasoning about perceptual patterns, the Bennett Mechanical Comprehension test, and tests for Clerical Aptitude, Language Comprehension, Arithmetic and a Global comprehensive score, that combines all five tests. All tests were administered to over 95% of the population who were examined at induction. Scores for all tests were grouped in six levels from 1 (highest) to 6 (lowest). The test scores are highly correlated with Pearson's r values in the range of .63 to .76. Here, we only focus on the scores of the comprehensive test.

Selected demographic and socioeconomic characteristics at the time of military examinations by education level are given in Table 1. First born conscripts tend to have higher education. Father's occupation was classified into five categories: professional and managerial workers; clerical, self-employed and skilled workers; farmers; semi-skilled workers including operators, process workers and shop assistants; and labourers and miners. Fathers with unknown occupations were classified separately. Education level is also strongly related to father's occupation; men with the highest education tend to have fathers in professional or managerial occupations. Religion was classified into five categories. The place of birth was categorized in six regions. The combined cognition measure is the Global comprehensive score. Not surprisingly, men with the highest education tend to do best on the comprehensive IQ test. Our principal measure of health is mortality with ages of death ranging from 18 up to 68–72. The lowest education group has a 70% higher mortality.

The Kaplan-Meier survival curves for the four education categories are shown in Figure 2 and reflect these mortality differences. Survival increases with the education level and the differences between the education levels increase with age. The curves differ significantly ($\chi^2 = 180.76$ for a log-rank test with 3 degrees of freedom). In subgroup analyses, survival differences comparing adjacent education levels are also statistically significant ($\chi^2 = 54.79, 9.97, 29.80$). This mortality difference by education is not necessarily due to education per se. It could be that the higher cognitive ability of higher educated people causes the difference. For example, understanding a doctor's advice and adhering to complex treatments may be driven by cognitive ability rather than education. From Table 1 we have seen already that education and IQ are highly correlated. Figure 3 shows that survival also increases with IQ and the differences are statistically significant ($\chi^2 = 277.72$ for a log-rank test with 5 degrees of freedom). For all, except the two lowest, adjacent IQ levels the differences in the Kaplan-Meier survival curves are significant. Within each education level the Kaplan-Meier curves also differ significantly by IQ-level (not shown here).

Next we investigate the relationship between IQ and educational attainment. The IQ scores are measured on a six-point ordinal scale. Comparing individuals on the extremes of the education level is not helpful as these individuals differ too much in many respects. We focus on adjacent education levels only and estimate separate ordered probit models for the IQ-score in relation to the highest education level in each pair and other observed individual characteristics. The results of ordered probit analyses reveal a strong association between education and IQ.[7]

---

[7]The results are available upon request.

Figure 2: Kaplan-Meier survival curves, by education level
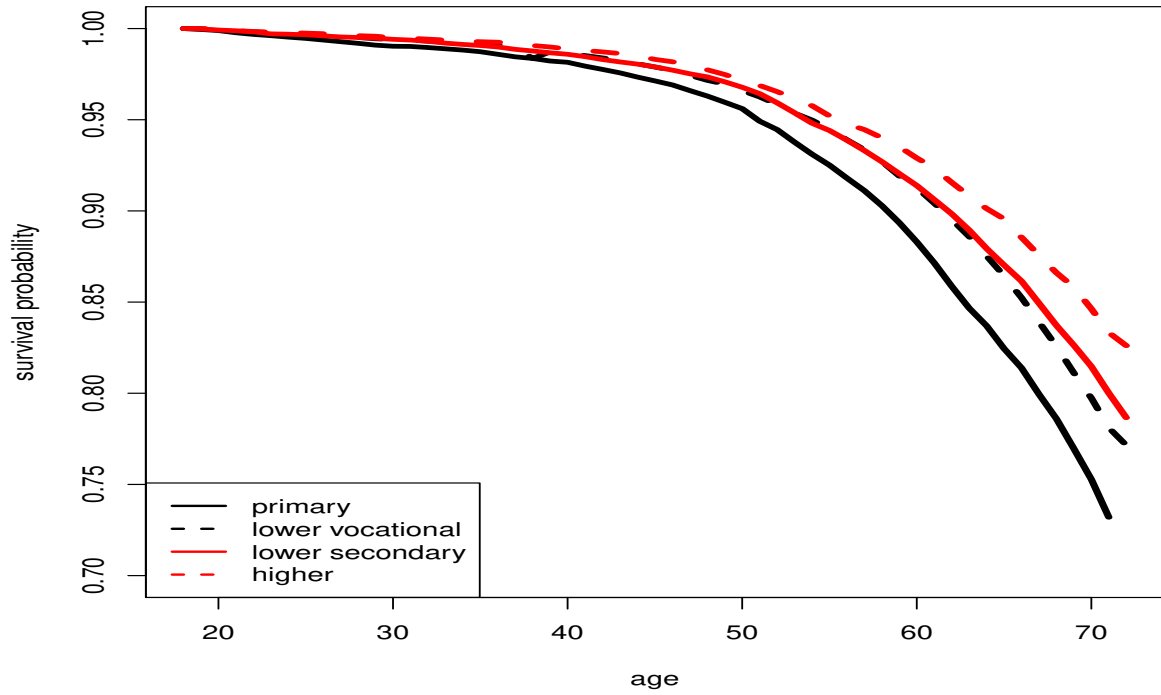
Figure 3: Kaplan-Meier survival curves, by IQ level (overall level)
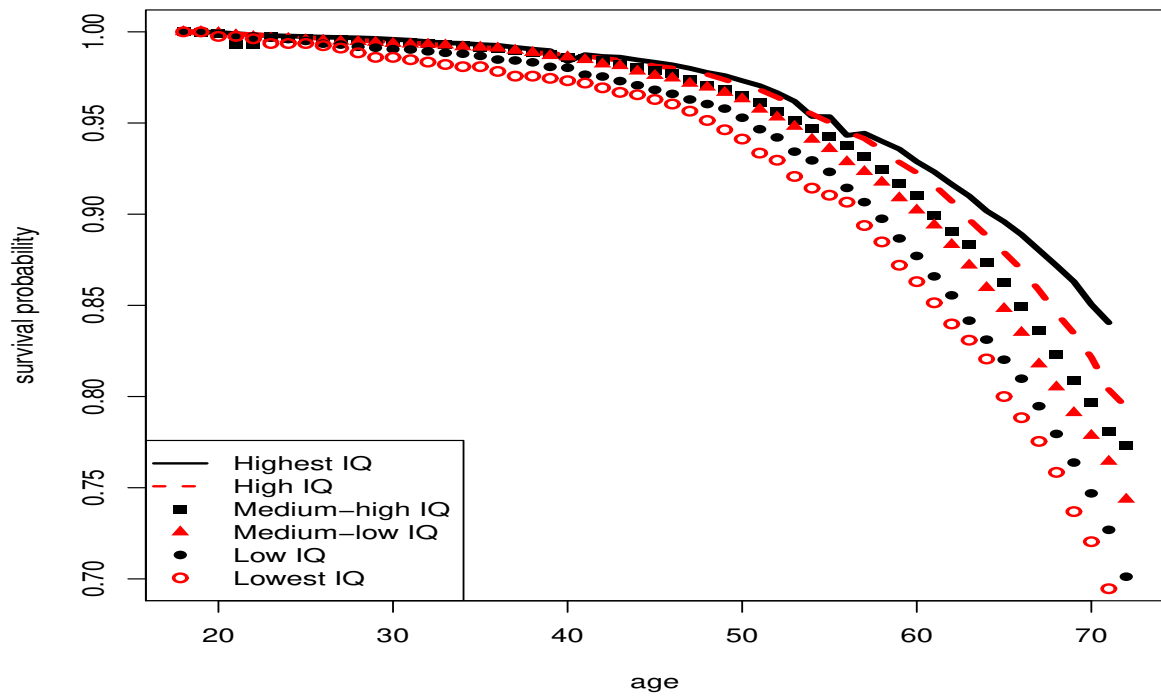
Table 1: Sample distribution by education level

	Primary education	Lower vocational	Lower secondary	Higher education	All levels
Birth order:					
1	27.8	32.1	39.3	42.6	35.5
2	27.1	30.3	30.7	29.9	29.9
3	18.7	18.4	16.3	15.4	17.3
4	11.3	9.2	6.9	7.0	8.4
≥ 5	14.9	10.0	6.7	5.1	8.8
Region of birth:					
North	2.9	4.2	3.2	2.3	3.4
South	8.3	7.2	4.9	5.0	6.4
East	4.8	6.0	3.8	3.6	4.7
North-Holland	35.2	31.8	35.6	38.2	34.2
South-Holland	38.2	43.5	44.7	42.0	43.0
Utrecht	10.7	7.4	8.0	9.0	8.4
Religion:					
Catholic	40.3	32.5	30.3	31.4	32.7
Dutch Reformed	25.5	31.2	31.3	30.2	30.2
Calvin	3.6	7.5	8.6	9.3	7.3
Other religion	0.6	0.5	0.8	1.0	0.8
No religion	30.1	28.2	29.0	28.1	28.8
Father's occupation:					
Professional	8.7	10.2	17.2	39.0	17.0
White collar	19.7	29.7	42.8	42.9	34.8
Farm owner	3.0	5.7	2.2	1.7	3.5
Skilled	38.4	33.3	23.1	9.2	26.7
Unskilled	22.5	14.9	9.4	3.4	12.3
Unknown	7.7	6.2	5.3	3.9	5.7
Global comprehensive IQ score:					
1 (highest)	0.1	6.3	19.8	54.6	17.6
2	3.8	27.5	47.9	37.7	32.5
3	13.7	30.3	20.9	4.0	20.6
4	28.3	22.7	7.2	0.6	14.9
5	39.5	10.6	1.7	0.1	10.1
6 (lowest)	11.5	0.8	0.1	0.02	2.0
missing	3.1	1.7	2.4	3.0	2.4
Total # of deaths	1,404	2,918	2,403	953	7,678
% died	25.2	20.5	18.8	15.4	19.8
Sample size	5,713	14,574	13,125	6,391	39,803

11

3.1 Results

3.2 Hazard models and mediation analysis

Table 2 presents the estimated effect on the mortality hazard of moving up one educational level and its decomposition. We conclude from these analyses that for the lower educated, with only primary education, and for the lower secondary educated obtaining more education reduces their mortality rate (around 25%). Moving from lower vocational education to lower secondary education only reduces the mortality rate by 9%.[8]

Table 2: Impact of education levels on the mortality rate using a Gompertz-gamma MPH model and its decomposition

	total effect		other pathways		cognitive ability	
	Unadjusted	IPW	$\theta(1)$	$\theta(0)$	$\eta(0)$	$\eta(1)$
Primary to	-0.250^{**}	-0.222^{**}	-0.060	-0.093^{+}	-0.162^{+}	-0.128^{+}
lower vocational	(0.038)	(0.034)	(0.067)	(0.045)	(0.075)	(0.056)
Lower vocational to	-0.089^{**}	-0.086^{**}	0.006	0.014	-0.092^{+}	-0.100^{+}
lower secondary	(0.029)	(0.029)	(0.033)	(0.039)	(0.044)	(0.048)
Lower secondary to	-0.229^{**}	-0.206^{**}	-0.127^{+}	-0.097	-0.079	-0.109
higher	(0.044)	(0.048)	(0.053)	(0.070)	(0.071)	(0.085)

$^{+}p < 0.05$ and $^{**}p < 0.01$

The last four columns in Table 2 present the decomposition of the effects of education on the mortality rate. The effect of education through other pathways is only significant for the highest education group while holding cognitive ability at the level of those with high education and for the lowest education group while holding cognitive ability at the level of those with primary education. About two-thirds of the mortality reduction for men moving from lower secondary to higher education runs through other pathways, such as, for example, an increase in income. For the lowest education groups the impact of education on mortality mainly runs through the increase in cognitive ability induced by the additional education. For these men 90% of the reduction in mortality is explained by the effect running through cognitive ability.

3.3 Robustness checks

Throughout we have assumed that the propensity scores are estimated consistently. Misspecification of the propensity score will generally produce bias. An approach to improve the robustness of the proposed methodology can be obtained using a doubly robust estimator which also includes a regression adjustment. Rotnitzky and Robins (1995) point out that if either the regression adjustment or the propensity score is correctly specified the resulting estimator will be consistent. Thus we also estimate doubly robust estimators of the models, including the observed characteristics and the IQ-test both in the propensity score and in the hazard regression, see Table 3. Including regression covariates hardly changes the IPW estimates (compare column 2 of Table 3 and of Table 2). Not surprisingly, including the

---

[8]The estimates of the probit propensity score used to calculate the weights can be found in Table B.2 and B.3 in the Appendix.

covariates does change the "unadjusted" results a little (compare column 1 of Table 3 and of Table 2).

Table 3: Double robust estimation of the total effect of education on the mortality rate and its decomposition using an IPW Gompertz-gamma MPH

	total effect		other pathways		cognitive ability	
	Unadjusted	IPW	$\theta(1)$	$\theta(0)$	$\eta(0)$	$\eta(1)$
Primary to	-0.227^{**}	-0.247^{**}	-0.061	-0.093^{+}	-0.166^{+}	-0.133^{+}
lower vocational	(0.038)	(0.039)	(0.068)	(0.045)	(0.077)	(0.059)
Lower vocational to	-0.086^{**}	-0.090^{**}	0.007	0.014	-0.093^{+}	-0.100^{+}
lower secondary	(0.029)	(0.029)	(0.033)	(0.039)	(0.044)	(0.049)
Lower secondary to	-0.204^{**}	-0.200^{**}	-0.128^{+}	-0.096	-0.077	-0.108
higher	(0.047)	(0.045)	(0.053)	(0.071)	(0.071)	(0.085)

The unadjusted robust estimator includes all the variables used for the propensity score as control variables in the Gompertz-gamma MPH model. $^{+}p < 0.05$ and $^{**}p < 0.01$

The individuals who were removed from the analysis, because their survival status is unknown, may be a selective sample, see Table B.1 in the Appendix. To account for possible sample selection bias we estimated the propensity score of an individual being removed, using a probit model for each level of education separately.[9] Based on this probability of removal we impose additional weighting of all observations in our estimation sample using the inverse of the probability of inclusion in the sample and we re-estimate the total effect and its decomposition. The results after imposing this additional weighting show very little difference from the original results, see Table B.4 in the Appendix.

Another issue is that childhood health problems may influence both education choice and mortality later in life. We perform a robustness analysis by adding health indicators to the educational propensity score. Our data are limited and only include health measurements at the military examination so these can only be used to proxy childhood health. We used indicators for height < 170cm; height > 185cm; overweight(bmi > 25), poor general health; poor hearing; poor sight and poor psychological assessment, and re-estimated the propensity scores, both without IQ, to estimate the total effect of education and, with IQ measurements to decompose the total effect into an effect running through changes in cognitive ability and an effect running through other pathways. The estimated impact of education on mortality changes slightly when accounting for health problems, see Table 4, but only for the lowest education group.

---

[9]The estimation results are available upon request.

Table 4: Impact of education on the mortality rate and its decomposition using an IPW Gompertz-gamma MPH, including health at age 18 indicators in the propensity score

	Total	other pathways		cognitive ability	
		$\theta(1)$	$\theta(0)$	$\eta(0)$	$\eta(1)$
Primary to	-0.194^{**}	-0.044	-0.079	-0.151^{+}	-0.115^{+}
lower vocational	(0.034)	(0.064)	(0.045)	(0.073)	(0.057)
Lower vocational to	-0.089^{**}	-0.005	0.005	-0.085	-0.094
lower secondary	(0.029)	(0.033)	(0.039)	(0.043)	(0.048)
Lower secondary to	-0.213^{**}	-0.140^{**}	-0.108	-0.073	-0.105
higher	(0.049)	(0.054)	(0.072)	(0.073)	(0.087)

$^{+}p < 0.05$ and $^{**}p < 0.01$

14

3.4 Sensitivity analyses

The critical assumption in propensity score weighting is that of no selection on unobservables. To test the sensitivity of the estimates to the unconfoundedness assumption we build on the sensitivity analyses of Nannicini (2007) and Ichino et al. (2008). We extend these analyses to the mixed proportional hazard model. The Ichino et al. (2008) sensitivity analysis assumes that the possible unobserved confounding factors can be summarised in a binary variable, U, and that the unconfoundedness assumption holds conditional on X and U, i.e. $\lambda(t|0) \perp D|X,U$. Given the values of the probabilities that characterize the distribution of U we can simulate a value of the unobserved confounding factor for each individual and re-estimate the IPW-MPH. The probabilities of the distribution of U depend on the value of the treatment and the outcome. The Ichino et al. (2008) sensitivity analysis assumes that the potential outcomes are binary, but Nannicini (2007) shows how to extend this to continuous outcomes by imposing a binary transformation. In survival analysis we have a natural binary transformation, the censoring indicator $\delta_i = 1$ if individual i is still alive at the end of the observation period. Then, the distribution of the unobserved binary confounding factor U can be characterised by specifying the probabilities in each of the four groups.

$$p_{ij} = \Pr(U = 1|D = i, \delta = j, X) = \Pr(U = 1|D = i, \delta = j) \tag{8}$$

for $i, j = 0, 1$.

A measure of how the different configurations of p_{ij}, chosen to simulate U, translate into associations of U with the outcome is ω, the coefficient of U in a MPH model for the control group ($D = 0$) using U and X as covariates. Ichino et al. (2008) call this (exponentiated) coefficient the 'outcome effect'. A measure of the effect of U on the relative probability to be assigned to the treatment is ξ, with ξ the coefficient of U in a logit model on the treatment assignment ($D = 1$) using U and X as covariates. Ichino et al. (2008) call this (exponentiated) coefficient the 'selection effect'.

For identification of the mediation effects we also impose sequential ignorability (Assumption 2). We therefore also assume that conditional on the binary (unobserved) factor the following two conditions hold (i) $\{\lambda(t|d',m), Q(d)\} \perp D|X, U$ and (ii) $\lambda(t|d',q) \perp Q|D = d, X, U$ for $\forall d, d' = 0, 1$ and q in the support of Q. A new measure, the mediator-effect, is ψ, the coefficient of U in an ordered logit model on the IQ-test values for the control group using U and X as covariates.

The probability values of the distribution for U are chosen so that they mimic the distribution for each included binary variable. For example, consider the probability that an individual in the lowest education group (primary and lower vocational education) is catholic. Then, p_{00} is this probability for catholics with primary education who died before the end of the observation period, p_{01} is the probability for catholics with primary education who survived till the end, p_{10} is the probability for catholics with lower vocational education who died before the end, and p_{11} is the probability for catholics with lower vocational education who survived till the end. For each probability configuration of U we repeat the simulation of U, the estimation of the outcome effect, the selection effect and the IPW-MPH treatment effects $M = 100$ times and obtain the average of these 100 simulations. The total variance of these averages can be estimated from (see Ichino et al. (2008)):

$$\text{Var}_f = \frac{1}{M} \sum_{m=1}^{M} s_m^2 + \frac{M-1}{M(M-1)} \sum_{m=1}^{M} (\hat{f}_m - \bar{f})^2 \tag{9}$$

with $f \in \{\omega, \xi\}$ of each pairwise education comparison, $\hat{f}_m$ is the estimated f in each simulation sample m and s_m^2 is its estimated variance.

Next we re-estimate the total effect of education on mortality using an IPW Gompertz-gamma MPH model including U in the propensity score and the decomposition of the effect using an IPW Gompertz-gamma MPH including U and the IQ-measurements in the propensity score.

An issue with our empirical application is that early childhood IQ (one of the possible factors of U_0 in Figure 1) might be a selection variable, explaining selection into education (rather than a mediation variable).[10] We, therefore, focus on the results of the sensitivity analysis when assuming U mimics the observed distribution of the IQ-measurements, i.e. the observed education choice and censoring probability are equal to the observed education choice and censoring prevalence for individuals with a given IQ level. We find the largest outcome, selection and mediation effects when the distribution of U mimics the impact of IQ on education and censoring.[11] Table 5 reports the simulated total effect and its decomposition into an effect running through cognitive ability and an effect running through other pathways including U in the IPW that mimics the distribution of the education choice and mortality for each IQ-level.[12] We find the largest changes in our IPW estimates when U mimics the education-mortality distribution of those with the highest IQ-level. These differences are, however, not statistically significant.

---

[10]We estimated the selection version of the model (despite the date of measurement on IQ) and that we got similar results for total effect of IQ and education suggesting that selection (only) is another plausible hypothesis, see Bijwaard and Jones (2016).

[11]The results can be found in Table B.5 in Appendix B.

[12]The results when U is based on the distribution of eduction choice and mortality for the other included variables can be found in Tables B.6 and B.7 in Appendix B.

Table 5: Sensitivity analysis : effect running through cognitive ability and running through other pathways (U based on IQ-levels)

Total effect

	Primary to lower vocational	Lower vocational to lower secondary	Lower secondary to higher
original	-0.222^{**}	-0.086^{**}	-0.206^{**}
	(0.034)	(0.029)	(0.048)
IQ			
1 (highest)	-0.222^{**}	-0.053	-0.124^{+}
	(0.140)	(0.030)	(0.057)
2	-0.160^{**}	-0.068^{+}	-0.196^{**}
	(0.058)	(0.030)	(0.049)
4	-0.225^{**}	-0.056	-0.204^{**}
	(0.035)	(0.031)	(0.067)
5	-0.179^{**}	-0.055	-0.207^{**}
	(0.041)	(0.033)	(0.053)
6 (lowest)	-0.198^{**}	-0.081^{**}	-0.206^{**}
	(0.039)	(0.029)	(0.048)
missing	-0.220^{**}	-0.086^{**}	-0.208^{**}
	(0.035)	(0.029)	(0.048)

other pathways

	Primary to lower vocational $\theta(1)$	$\theta(0)$	Lower vocational to lower secondary $\theta(1)$	$\theta(0)$	Lower secondary to higher $\theta(1)$	$\theta(0)$
original	-0.060	-0.093^{+}	0.006	0.014	-0.127^{+}	-0.097
	(0.067)	(0.045)	(0.033)	(0.039)	(0.053)	(0.070)
IQ						
1 (highest)	0.061	-0.087	0.040	0.049	-0.044	-0.009
	0.379)	(0.130)	(0.035)	(0.041)	(0.062)	(0.099)
2	0.085	-0.028	0.023	0.032	-0.117^{+}	-0.086
	0.260)	(0.063)	(0.035)	(0.041)	(0.054)	(0.074)
4	-0.064	-0.097^{+}	0.037	0.049	-0.125	-0.082
	0.068)	(0.045)	(0.036)	(0.046)	(0.072)	(0.202)
5	-0.010	-0.047	0.038	0.052	-0.128^{+}	-0.095
	0.093)	(0.053)	(0.037)	(0.053)	(0.059)	(0.113)
6 (lowest)	-0.033	-0.062	0.011	0.021	-0.127^{+}	-0.097
	0.074)	(0.067)	(0.033)	(0.050)	(0.053)	(0.070)
missing	-0.058	-0.091^{+}	0.006	0.014	-0.129^{+}	-0.099
	0.067)	(0.045)	(0.033)	(0.039)	(0.053)	(0.070)

cognitive ability

	Primary to lower vocational $\eta(0)$	$\eta(1)$	Lower vocational to lower secondary $\eta(0)$	$\eta(1)$	Lower secondary to higher $\eta(0)$	$\eta(1)$
original	-0.162^{+}	-0.128^{+}	-0.092^{+}	-0.100^{+}	-0.079	-0.109
	(0.075)	(0.056)	(0.044)	(0.048)	(0.071)	(0.085)
IQ						
1 (highest)	-0.283	-0.134	-0.092^{+}	-0.102^{+}	-0.081	-0.115
	(0.405)	(0.191)	(0.046)	(0.051)	(0.084)	(0.114)
2	-0.246	-0.132	-0.091^{+}	-0.101^{+}	-0.079	-0.110
	(0.267)	(0.086)	(0.046)	(0.051)	(0.073)	(0.088)
4	-0.161^{+}	-0.129^{+}	-0.093	-0.105	-0.079	-0.122
	(0.076)	(0.057)	(0.047)	(0.055)	(0.098)	(0.213)
5	-0.169	-0.132	-0.093	-0.107	-0.079	-0.111
	(0.101)	(0.067)	(0.049)	(0.062)	(0.079)	(0.125)
6 (lowest)	-0.166^{+}	-0.137	-0.092^{+}	-0.102	-0.079	-0.109
	(0.083)	(0.077)	(0.044)	(0.058)	(0.071)	(0.085)
missing	-0.161^{+}	-0.129^{+}	-0.092^{+}	-0.100^{+}	-0.079	-0.108
	(0.076)	(0.057)	(0.044)	(0.048)	(0.071)	(0.085)

Based on adding U to propensity score with probabilities of U from observed probabilities for each IQ value.
$^{+}p < 0.05$ and $^{**}p < 0.01$

Next we search for the existence of 'killer'-confounders, i.e. the existence of a set of probabilities p_{ij} such that if U were observed the estimated effects would be driven to zero. The reason for doing this is to assess the plausibility of the resulting configuration of U and how comparable this is to the distribution of observed confounders. In order to reduce the dimensionality of the characterisation of the 'killer'-confounders we follow the suggestion of Nannicini (2007) and fix the probability of $\Pr(U = 1)$ to 0.4 and the difference $p_{11} - p_{10}$ to zero. Now the simulated confounders U can be fully described by two differences $d = p_{01} - p_{00}$ and $s = p_{1.} - p_{0.}$, with $p_{i.} = \Pr(U = 1|D = i) = p_{i1} \cdot \Pr(\delta_1 = 1|D = i) + p_{i0} \cdot \Pr(\delta_1 = 0|D = i)$ for $i = 0, 1$, the fraction of individuals with $U = 1$ by education level. Nannicini (2007) argues that d is an (inconsistent) measure of the effect of U on the outcome (mortality, censoring probability) for the untreated (lower education level), while s is an (inconsistent) measure of the selection into treatment (higher education level). Both d and s are inconsistent measures because they do not account for the association between U and W, while our outcome effect, Ω, selection effects, ξ and mediation effects ψ, account for this.

Table 6 reports the simulated total effect and its decomposition when the distribution of U is defined by d, s with $d, s = 0.1, \ldots, 0.5$.[13] Indeed, by using these killer-confounders we do find some large deviations from the original results for the impact of moving from primary to lower vocational education, while the estimates for higher levels of education remain remarkably stable. However these differences apply for combinations of d and s that lie well away from the values implied by our observed confounders. Note that the largest values for d and s we found when the distribution of U mimics the education-censoring distribution of the observed variables was $d = 0.03$ and $s = 0.06$ when using the education-censoring distribution of the highest IQ-level.

---

[13]The simulated outcome-, selection and mediation effects can be found in Table B.8 in Appendix B.

Table 6: Sensitivity analysis characterizing 'killer' confounders (mediator): effect running through cognitive ability and running through other pathways

	Primary to lower vocational			Lower vocational to lower secondary			Lower secondary to higher		
	Total effect	other pathways $\theta(1)$	$\theta(0)$	Total effect	other pathways $\theta(1)$	$\theta(0)$	Total effect	other pathways $\theta(1)$	$\theta(0)$
original	-0.222^{**}	-0.060	-0.093^{+}	-0.086^{**}	0.006	0.014	-0.206^{**}	-0.127^{+}	-0.097
	(0.034)	(0.067)	(0.045)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.070)
$d = 0.1$ & $s = 0.1$	-0.207^{**}	-0.043	-0.078	-0.086^{**}	0.006	0.014	-0.206^{**}	-0.127^{+}	-0.097
	(0.035)	(0.068)	(0.045)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.070)
$d = 0.1$ & $s = 0.2$	-0.175^{**}	-0.003	-0.046	-0.086^{**}	0.006	0.014	-0.206^{**}	-0.127^{+}	-0.097
	(0.037)	(0.076)	(0.046)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.071)
$d = 0.1$ & $s = 0.3$	-0.128^{**}	0.057	0.001	-0.086^{**}	0.006	0.014	-0.206^{**}	-0.127^{+}	-0.097
	(0.040)	(0.098)	(0.050)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.071)
$d = 0.1$ & $s = 0.4$	-0.039	0.195	0.091	-0.086^{**}	0.006	0.014	-0.206^{**}	-0.126^{+}	-0.096
	(0.046)	(0.142)	(0.055)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.072)
$d = 0.1$ & $s = 0.5$	0.228^{**}	0.637^{**}	0.360^{**}	-0.086^{**}	0.005	0.014	-0.206^{**}	-0.127^{+}	-0.097
	(0.052)	(0.212)	(0.062)	(0.029)	(0.033)	(0.040)	(0.048)	(0.054)	(0.073)
$d = 0.2$ & $s = 0.1$	-0.183^{**}	-0.018	-0.056	-0.082^{**}	0.010	0.018	-0.208^{**}	-0.129^{+}	-0.100
	(0.035)	(0.068)	(0.045)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.071)
$d = 0.2$ & $s = 0.2$	-0.159^{**}	0.011	-0.031	-0.086^{**}	0.006	0.014	-0.206^{**}	-0.127^{+}	-0.097
	(0.037)	(0.071)	(0.046)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.071)
$d = 0.2$ & $s = 0.3$	-0.082^{+}	0.108	0.047	-0.386^{**}	0.006	0.014	-0.206^{**}	-0.127^{+}	-0.097
	(0.038)	(0.083)	(0.048)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.071)
$d = 0.2$ & $s = 0.4$	0.048	0.286^{**}	0.177^{**}	-0.086^{**}	0.006	0.014	-0.206^{**}	-0.126^{+}	-0.096
	(0.040)	(0.104)	(0.050)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.072)
$d = 0.2$ & $s = 0.5$	0.253^{**}	0.586^{**}	0.382^{**}	-0.086^{**}	0.005	0.014	-0.206^{**}	-0.127^{+}	-0.097
	(0.040)	(0.119)	(0.051)	(0.029)	(0.033)	(0.040)	(0.048)	(0.054)	(0.073)
$d = 0.3$ & $s = 0.1$	-0.149^{**}	0.016	-0.025	-0.077^{**}	0.014	0.023	-0.221^{**}	-0.142^{**}	-0.113
	(0.036)	(0.069)	(0.046)	(0.029)	(0.033)	(0.039)	(0.049)	(0.054)	(0.073)
$d = 0.3$ & $s = 0.2$	-0.117^{**}	0.056	0.009	-0.079^{**}	0.012	0.021	-0.207^{**}	-0.128^{+}	-0.098
	(0.036)	(0.071)	(0.046)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.070)
$d = 0.3$ & $s = 0.3$	-0.069	0.117	0.059	-0.086^{**}	0.006	0.014	-0.206^{**}	-0.127^{+}	-0.097
	(0.036)	(0.075)	(0.047)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.071)
$d = 0.3$ & $s = 0.4$	0.084^{+}	0.314^{**}	0.211^{**}	-0.086^{**}	0.006	0.014	-0.206^{**}	-0.126^{+}	-0.096
	(0.037)	(0.082)	(0.048)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.072)
$d = 0.3$ & $s = 0.5$	0.207^{**}	0.488^{**}	0.335^{**}	-0.086^{**}	0.005	0.014	-0.206^{**}	-0.127^{+}	-0.097
	(0.038)	(0.094)	(0.049)	(0.029)	(0.033)	(0.040)	(0.048)	(0.054)	(0.073)
$d = 0.4$ & $s = 0.1$	-0.106^{**}	0.059	0.014	-0.071^{+}	0.018	0.027	-0.245^{**}	-0.167^{**}	-0.139
	(0.036)	(0.070)	(0.046)	(0.029)	(0.033)	(0.039)	(0.050)	(0.055)	(0.077)
$d = 0.4$ & $s = 0.2$	-0.061	0.113	0.061	-0.070^{+}	0.021	0.030	-0.216^{**}	-0.137^{+}	-0.108
	(0.036)	(0.072)	(0.047)	(0.029)	(0.033)	(0.039)	(0.048)	(0.054)	(0.071)
$d = 0.4$ & $s = 0.3$	-0.010	0.181	0.115^{+}	-0.077^{**}	0.014	0.023	-0.205^{**}	-0.126^{+}	-0.096
	(0.036)	(0.074)	(0.047)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.070)
$d = 0.4$ & $s = 0.4$	0.066	0.277^{**}	0.193^{**}	-0.086^{**}	0.006	0.014	-0.206^{**}	-0.126^{+}	-0.096
	(0.036)	(0.074)	(0.047)	(0.029)	(0.033)	(0.039)	(0.048)	(0.05;)	(0.072)
$d = 0.4$ & $s = 0.5$	0.160^{**}	0.408^{**}	0.288^{**}	-0.086^{**}	0.005	0.014	-0.206^{**}	-0.127^{+}	-0.097
	(0.037)	(0.082)	(0.048)	(0.029)	(0.033)	(0.040)	(0.048)	(0.054)	(0.073)
$d = 0.5$ & $s = 0.1$	-0.048	0.116	0.066	-0.063^{+}	0.024	0.033	-0.285^{**}	-0.207^{**}	-0.181^{+}
	(0.037)	(0.071)	(0.047)	(0.030)	(0.034)	(0.040)	(0.052)	(0.058)	(0.082)
$d = 0.5$ & $s = 0.2$	0.007	0.185^{+}	0.125^{**}	-0.058^{+}	0.031	0.041	-0.239^{**}	-0.160^{**}	-0.131
	(0.037)	(0.072)	(0.047)	(0.029)	(0.034)	(0.040)	(0.049)	(0.055)	(0.074)
$d = 0.5$ & $s = 0.3$	0.069	0.263^{**}	0.190^{**}	-0.063^{+}	0.028	0.038	-0.212^{**}	-0.133^{+}	-0.104
	(0.037)	(0.073)	(0.047)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.071)
$d = 0.5$ & $s = 0.4$	0.097^{**}	0.303^{**}	0.221^{**}	-0.077^{**}	0.015	0.024	-0.205^{**}	-0.126^{+}	-0.095
	(0.036)	(0.073)	(0.047)	(0.029)	(0.033)	(0.039)	(0.048)	(0.053)	(0.071)
$d = 0.5$ & $s = 0.5$	0.110^{**}	0.332^{**}	0.238^{**}	-0.086^{**}	0.005	0.014	-0.206^{**}	-0.127^{+}	-0.097
	(0.036)	(0.076)	(0.047)	(0.029)	(0.033)	(0.040)	(0.048)	(0.054)	(0.073)

Based on adding U to propensity score under the assumption that $\Pr(U = 1) = 0.4$ and $p_{11} - p_{10} = 0$, the differences $d = p_{01} - p_{00}$ and $s = p_{1.} - p_{0.}$. $^{+}p < 0.05$ and $^{**}p < 0.01$

Table 6: Sensitivity analysis characterizing 'killer' confounders (continued)

	Primary to lower vocational cognitive ability		Lower vocational to lower secondary cognitive ability		Lower secondary to higher cognitive ability	
	$\eta(0)$	$\eta(1)$	$\eta(0)$	$\eta(1)$	$\eta(0)$	$\eta(1)$
original	-0.162^+	-0.128^+	-0.092^+	-0.100^+	-0.079	-0.109
	(0.075)	(0.056)	(0.044)	(0.048)	(0.071)	(0.085)
$d = 0.1$ & $s = 0.1$	-0.164^+	-0.128^+	-0.092^+	-0.100^+	-0.079	-0.109
	(0.076)	(0.057)	(0.044)	(0.048)	(0.071)	(0.085)
$d = 0.1$ & $s = 0.2$	-0.172^+	-0.129^+	-0.092^+	-0.100^+	-0.079	-0.109
	(0.084)	(0.059)	(0.044)	(0.048)	(0.071)	(0.085)
$d = 0.1$ & $s = 0.3$	-0.185	-0.129^+	-0.092^+	-0.100^+	-0.079	-0.109
	(0.106)	(0.064)	(0.044)	(0.048)	(0.072)	(0.086)
$d = 0.1$ & $s = 0.4$	-0.233	-0.130	-0.092^+	-0.100^+	-0.079	-0.110
	(0.150)	(0.071)	(0.044)	(0.049)	(0.072)	(0.086)
$d = 0.1$ & $s = 0.5$	-0.409	-0.132	-0.092^+	-0.100^+	-0.079	-0.109
	(0.219)	(0.081)	(0.044)	(0.049)	(0.072)	(0.087)
$d = 0.2$ & $s = 0.1$	-0.165^+	-0.127^+	-0.091^+	-0.100^+	-0.079	-0.109
	(0.077)	(0.057)	(0.044)	(0.048)	(0.071)	(0.085)
$d = 0.2$ & $s = 0.2$	-0.170^+	-0.128^+	-0.092^+	-0.100^+	-0.079	-0.109
	(0.080)	(0.059)	(0.044)	(0.048)	(0.071)	(0.085)
$d = 0.2$ & $s = 0.3$	-0.190^+	-0.128^+	-0.092^+	-0.100^+	-0.079	-0.109
	(0.091)	(0.061)	(0.044)	(0.048)	(0.072)	(0.086)
$d = 0.2$ & $s = 0.4$	-0.238^+	-0.128^+	-0.092^+	-0.100^+	-0.079	-0.110
	(0.111)	(0.064)	(0.044)	(0.049)	(0.072)	(0.086)
$d = 0.2$ & $s = 0.5$	-0.333^{**}	-0.129^+	-0.092^+	-0.100^+	-0.079	-0.109
	(0.125)	(0.065)	(0.044)	(0.049)	(0.072)	(0.087)
$d = 0.3$ & $s = 0.1$	-0.166^+	-0.124^+	-0.091^+	-0.099^+	-0.079	-0.108
	(0.078)	(0.058)	(0.044)	(0.049)	(0.073)	(0.087)
$d = 0.3$ & $s = 0.2$	-0.172^+	-0.126^+	-0.091^+	-0.100^+	-0.079	-0.109
	(0.080)	(0.058)	(0.044)	(0.048)	(0.071)	(0.085)
$d = 0.3$ & $s = 0.3$	-0.186^+	-0.128^+	-0.092^+	-0.100^+	-0.079	-0.109
	(0.083)	(0.059)	(0.044)	(0.048)	(0.072)	(0.086)
$d = 0.3$ & $s = 0.4$	-0.231^{**}	-0.128^+	-0.092^+	-0.100^+	-0.079	-0.110
	(0.090)	(0.051)	(0.044)	(0.049)	(0.072)	(0.086)
$d = 0.3$ & $s = 0.5$	-0.281^{**}	-0.128^+	-0.092^+	-0.100^+	-0.079	-0.109
	(0.101)	(0.062)	(0.044)	(0.049)	(0.072)	(0.087)
$d = 0.4$ & $s = 0.1$	-0.165^+	-0.120^+	-0.089^+	-0.098^+	-0.079	-0.109
	(0.079)	(0.059)	(0.044)	(0.049)	(0.075)	(0.091)
$d = 0.4$ & $s = 0.2$	-0.174^+	-0.123^+	-0.091^+	-0.100^+	-0.079	-0.108
	(0.081)	(0.059)	(0.044)	(0.049)	(0.072)	(0.086)
$d = 0.4$ & $s = 0.3$	-0.191^+	-0.125^+	-0.091^+	-0.100^+	-0.079	-0.109
	(0.082)	(0.059)	(0.044)	(0.048)	(0.071)	(0.085)
$d = 0.4$ & $s = 0.4$	-0.212^+	-0.127^+	-0.092^+	-0.100^+	-0.079	-0.110
	(0.083)	(0.059)	(0.044)	(0.049)	(0.072)	(0.086)
$d = 0.4$ & $s = 0.5$	-0.248^{**}	-0.127^+	-0.092^+	-0.100^+	-0.079	-0.109
	(0.090)	(0.061)	(0.044)	(0.049)	(0.072)	(0.087)
$d = 0.5$ & $s = 0.1$	-0.154	-0.114	-0.087	-0.096^+	-0.078	-0.104
	(0.080)	(0.060)	(0.045)	(0.049)	(0.078)	(0.097)
$d = 0.5$ & $s = 0.2$	-0.178^+	-0.118^+	-0.089^+	-0.099^+	-0.079	-0.107
	(0.081)	(0.060)	(0.045)	(0.049)	(0.074)	(0.089)
$d = 0.5$ & $s = 0.3$	-0.194^+	-0.121^+	-0.091^+	-0.101^+	-0.079	-0.108
	(0.082)	(0.060)	(0.044)	(0.049)	(0.072)	(0.086)
$d = 0.5$ & $s = 0.4$	-0.206^+	-0.124^+	-0.092^+	-0.100^+	-0.079	-0.109
	(0.082)	(0.059)	(0.044)	(0.049)	(0.072)	(0.086)
$d = 0.5$ & $s = 0.5$	-0.222^{**}	-0.127^+	-0.092^+	-0.100^+	-0.079	-0.109
	(0.084)	(0.050)	(0.044)	(0.049)	(0.072)	(0.087)

Based on adding U to propensity score under the assumption that $\Pr(U = 1) = 0.4$ and $p_{11} - p_{10} = 0$, the differences $d = p_{01} - p_{00}$ and $s = p_{1.} - p_{0.}$. $^+p < 0.05$ and $^{**}p < 0.01$

3.5 Implied gain in life-expectancy

From the Gompertz-hazards we can estimate the median survival age of the recruits and their post 18 life expectancy. The median survival age is the age at which half of the people have died (conditional on survival up to age 18). Assuming that the estimated Gompertz hazard holds, the life expectancy at age $t_0 = 18$ can be very well approximated by (see Lenart (2014)):

$$\text{LE}(t_0) = -\exp\left(e^{\alpha_0 + \alpha_1 t_0}\right)(\alpha_0 - \ln(\alpha_1) + \alpha_1 t_0 + 0.5772)/\alpha_1 \tag{10}$$

where 0.5772 is the Euler constant. For the unadjusted Gompertz model the estimated remaining life expectancies are 59.8 (primary); 62.6 (lower vocational); 63.7 (lower secondary) (64.2 based on last two education groups) and 66.7 (higher). Leading to educational gains of 2.8, 1.0 and 2.5 in life expectancy. The median survival ages are 80.1 (primary); 82.9 (lower vocational); 84.1 (lower secondary) (84.6) and 87.1 (higher). Thus leading to the same educational gains.

In Table 7 we report the gains in life expectancy. The lower panel of Table 7 reports the gains in life expectancy based on the mediation analysis and decomposes the effects of education into an effect running through cognitive ability and an effect running through other pathways. Based on the IPW estimates we can conclude that if an individual had improved his education from primary to lower vocational he would have gained 2.5 additional years (and his median age also would have improved by 2.5 years), of which 1.8 years are attributable to cognitive ability and 0.7 years to other changes induced by other pathways. If an individual had improved from lower vocational to lower secondary the gain in life expectancy is 1.0 year (1.1 attributable to cognitive ability and an negative impact of other pathways). The gain in life expectancy if an individual had improved his education from lower secondary to higher education is 2.2 years. For those who attained higher education this gain in life expectancy is mainly attributable to the other pathways (1.2 years), while for those with lower secondary education the effect running through cognitive ability is larger (1.3 years) than effect running through other pathways.

Table 7: Gain in life expectancy

	Primary to lower vocational	Lower vocational to lower secondary	Lower secondary to higher
Unadjusted	2.8	1.0	2.5
IPW mediation			
total	2.5	1.0	2.2
Other pathways			
$\theta(1)$	0.7	-0.1	1.3
$\theta(0)$	1.1	-0.2	1.0
Cognitive ability			
$\eta(0)$	1.8	1.1	0.9
$\eta(1)$	1.5	1.1	1.2

4   Discussion

A large literature documents that higher levels of education are positively associated with a longer life. Possible mechanisms include occupational risks, health behavior, the ability to process information and cognitive ability (Cutler and Lleras-Muney 2008). It is commonly acknowledged that education and cognitive ability are correlated. Cognitive ability may cause differences in educational outcomes or education may cause differences in cognitive ability. Most of the economics literature on the causal effect of education on health focuses on accounting for endogenous selection into education due to confounding factors, such as cognitive ability, either by exploiting natural experiments in education due to changes in compulsory schooling laws (Mazumder 2012) or by defining a structural model (Conti et al. 2010; Bijwaard et al. 2015). The estimates based on natural experiments find little to no effect of education on health, while the studies based on structural models find that around half of the difference in health by education is due to selection. An alternative perspective is that cognitive ability is part of the causal pathway from education to mortality. For instance, it has been proven that high scores on intelligence tests are positively associated with schooling level, (Ceci 1991; Hansen et al. 2004; Carlsson et al. 2015).

We assume that IQ measured at age 18 is affected by educational attainment and has a mediating effect on the mortality difference across education groups. We developed an inverse probability weighting (IPW) method for hazard models to estimate the impact of education on the mortality rate. We use conscription data of Dutch men born between 1944-1947 who were examined for military service between 1961-1965, and linked to national death records, in which we identified four education groups. Using the IPW methods we estimate, for each adjacent education group, the impact of improving education on the mortality risk. We decompose the impact of education into an effect running through cognitive ability and an effect running through other pathways.

The results show that controlling for cognitive ability, as a mediating factor, leaves only limited evidence of an educational gain in mortality. When accounting for cognitive ability as a mediator in the causal pathway from education to mortality we find that cognitive ability plays an important role in explaining the educational gradient. For men with primary school we find that the effect of education running through an increase in cognitive ability significantly reduces the mortality risk (about 15% reduction in the mortality rate), which is equivalent to 1.6 years longer life expectancy. For the middle group, with lower vocational education, cognitive ability explains most of the educational gradient of a 10% reduction in the mortality rate. For the highest education group, only the effect of education running through other pathways, such as income effects of education, is significant (about 13% reduction in the mortality rate), leading to 1.3 additional years of life.

A limitation of our data, based on military entrance examination, is that we only observe men and no information on women is available. Bijwaard et al. (2015) found that educational gains for women appear to be higher than for men, in spite of the higher survival difference of women with lower versus higher education. These findings are based on much smaller numbers than the current study however and therefore need to be interpreted with caution. Another issue is that in the 1960s a major change occurred in the education system in the Netherlands and some of the specific education strata in this study no longer exist. In addition, the percentage of people with more than six years of post-primary school education is currently much higher compared to the past. These changes are not likely to affect our gene-

ral conclusion that increased education only has a small effect on survival, but further long term studies will be needed to quantify these effects for contemporary school types. The issue of reverse causality that early childhood health affects educational attainment might distort our analyses (Case et al. 2005; Currie 2009). We have limited information about childhood health status. However, adding health measurements from the military examination to the propensity score did not change our conclusion that education has only limited impact on mortality.

References

Aalen, O. O., O. Borgan, and H. K. Gjessing (2009). *Survival and Event History Analysis*. New–York: Springer–Verlag.

Abbring, J. H. and G. J. van den Berg (2003). The non–parametric identification of treatment effects in duration models. *Econometrica 71*, 1491–1517.

Andersen, P. K. and O. Borgan (1985). Counting process models for life history data: A review. *Scandinavian Journal of Statistics 12*, 97–158.

Andersen, P. K., O. Borgan, R. D. Gill, and N. Keiding (1993). *Statistical Models Based on Counting Processes*. New York: Springer–Verlag.

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. *Statistics in Medicine 33*, 1242–1258.

Banks, J. and F. Mazzonna (2012). The effect of education on old age cognitive abilities: Evidence from a regression discontinuity design. *The Economic Journal 122*(560), 418–448.

Baron, R. M. and D. A. Kenny (1986). The moderator–mediator variable distinction in social psychological research: Conceptual, strategic, and statistical considerations. *Journal of Personlaity and Social Psychology 51*, 1173–1182.

Bijwaard, G. E. and A. M. Jones (2016). Cognitive ability and the mortality gradient by education: Selection or mediation? Discussion Paper 9798, IZA.

Bijwaard, G. E., H. van Kippersluis, and J. Veenman (2015). Education and health: the role of cognitive ability. *Journal of Health Economics 42*, 29–43.

Caliendo, M. and S. Kopeinig (2008). Some practical guidance for the implementation of propensity score matching. *Journal of Economic Surveys 22*, 31–72.

Carlsson, M., G. B. Dahl, B. Öckert, and D.-O. Rooth (2015). The effect of schooling on cognitive skills. *The Review of Economics and Statistics 97*, 533–547.

Case, A., A. Fertig, and C. Paxson (2005). The lasting impact of childhood health and circumstance. *Journal of Health Economics 24*, 365–389.

Ceci, S. J. (1991). How much does schooling influence general intelligence and its cognitive components? A reassessment of the evidence. *Developmental Psychology 27*, 703–722.

Cole, S. R. and M. A. Hernán (2004). Adjusted survival curves with inverse probability weights. *Computer Methods and Programs in Biomedicine 75*, 45–49.

Conti, G., J. J. Heckman, and S. Urzua (2010). The education-health gradient. *American Economic Review 100*, 234–238.

Currie, J. (2009). Healthy, wealthy, and wise: Socioeconomic status, poor health in childhood and human captial development. *Journal of Economic Literature 47*, 87–122.

Cutler, D. and A. Lleras-Muney (2008). Education and health: Evaluating theories and evidence. In J. S. House, R. F. Schoeni, G. A. Kaplan, and H. Pollack (Eds.), *Making Americans Healthier: Social and Economic Policy as Health Policy*. New-York: Russell Sage Foundation.

Dahmann, S. C. (2017). How does education improve cognitive skills? Instructional time versus timing of instruction. *Labour Economics 47*, 35–47.

Doob, L. J. (1953). *Stochastic Processes*. New-York: Wiley.

Doornbos, G. and D. Kromhout (1990). Educational level and mortality in a 32-year follow-up study of 18-year-old men in the Netherlands. *International Journal of Epidemiology 19*, 374–379.

Ekamper, P., F. van Poppel, A. D. Stein, and L. H. Lumey (2014). Independent and additive association of prenatal famine exposure and intermediary life conditions with adult mortality between age 18–63 years. *Social Science & Medicine 119*, 232–239.

Falch, T. and S. S. Massih (2011). The effect of education on cognitive ability. *Economic Inquiry 49*, 838–856.

Gavrilov, L. A. and N. S. Gavrilova (1991). *The Biology of Life Span: A Quantitative Approach*. New York: Harwood Academic Publisher.

Hansen, K. T., J. J. Heckman, and K. J. Mullen (2004). The effect of schooling and ability on achievement test scores. *Journal of Econometrics 121*, 39–98.

Hirano, K., G. W. Imbens, and G. Ridder (2003). Efficient estimation of average treatment effects using the estimated propensity score. *Econometrica 71*, 1161–1189.

Huber, M. (2014). Identifying causal mechanisms (primarily) based on inverse probability weighting. *Journal of Applied Econometrics 29*, 920–943.

Ichino, A., F. Mealli, and T. Nannicini (2008). From temporary help jobs to permanent employment: What can we learn from matching estimators and their sensitivity? *Journal of Applied Econometrics 23*, 305–327.

Imai, K., L. Keele, and D. Tingley (2010). A general approach to causal mediation analysis. *Psychological Methods 15*, 309–334.

Imai, K., L. Keele, and T. Yamamoto (2010). Identification, inference and sensitivity analysis for causal mediation effects. *Statistical Science 25*, 51–71.

Imbens, G. W. (2004). Nonparametric estimation of average treatment effects under exogeneity. *The Review of Economics and Statistics 86*, 4–29.

Jones, A. M., N. Rice, and P. Rosa Dias (2011). Long-term effects of school quality on health and lifestyle: Evidence from comprehensive schooling reforms in England. *Journal of Human Capital 5*, 342–376.

Klein, J. P. and M. L. Moeschberger (1997). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer–Verlag.

Lee, M.-J. and S. Kobayashi (2001). Proportional treatment effects for count response panel data: Effects of binary excercise on health care demand. *Health Economics 10*, 411–428.

Lenart, A. (2014). The moments of the Gompertz distribution and the maximum likelihood of its parameters. *Scandinavian Actuarial Journal of the Institute of Actuaries 3*, 255–277.

Mazumder, B. (2012). The effects of education on health and mortality. *Nordic Economic Policy Review 1*, 261–301.

Meyer, P. A. (1963). Decomposition of supermartingales: The uniqueness theorem. *Illinois Journal of Mathematics 7*, 1–17.

Nannicini, T. (2007). A simulation-based sensitivity analysis for matching estimators. *STATA Journal 7*, 334–350.

Pearl, J. (2000). *Causality: Models, Reasoning and Inference*. New York: Cambridge University Press.

Pearl, J. (2012). The mediation formula: A guide to the assessment of causal pathways in nonlinear models. In C. Berzuini, P. Dawid, and L. Bernardinelli (Eds.), *Causality: Statistical Perspectives and Applications*, pp. 151–179. Chichester: John Wiley.

Robins, J. M., M. A. Hernán, and B. Brumback (2000). Marginal structural models and causal inference in epidemiology. *Epidemiology 11*, 550–560.

Robins, J. M. and A. Rotnitzky (1992). Recovery of information and adjustment for dependent censoring using surrogate markers. In N. Jewell, K. Dietz, and V. Farewell (Eds.), *AIDS Epidemiology– Methodological Issues*, pp. 297–331. Boston: Birkhäuser.

Rosenbaum, P. and D. B. Rubin (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika 70*, 41–55.

Rotnitzky, A. and J. M. Robins (1995). Semiparametric regression estimation in the presence of dependent censoring. *Biometrika 82*, 805–820.

Rubin, D. B. (1974). Estimating causal effects of treatments in randomized and non–randomized studies. *Journal of Educational Psychology 66*, 688–701.

Schneeweis, N., V. Skirbekk, and R. Winter-Ebmer (2014). Does education improve cognitive performance four decades after school completion? *Demography 51*(2), 619–643.

Schröder, H. and H. B. G. Ganzeboom (2014). Measuring and modelling level of education in European societies. *European Sociological Review 47*(9), 119–136.

Tchetgen Tchetgen, E. J. (2013). Inverse odds ratio-weighted estimation for causal mediation analysis. *Statistics in Medicine 32*, 4567–4580.

Therneau, T. and P. Grambsch (2000). *Modeling Survival Data: Extending the Cox Model*. New–York: Springer–Verlag.

Van den Berg, G. J. (2001). Duration models: Specification, identification, and multiple duration. In J. Heckman and E. Leamer (Eds.), *Handbook of Econometrics, Volume V*, Chapter 55, pp. 3381–3460. Amsterdam: North–Holland.

VanderWeele, T. J. (2011). Causal mediation analysis with survival data. *Epidemiology 22*, 575581.

VanderWeele, T. J. (2015). *Explanation in Causal Inference: Methods for Mediation and Interaction*. New York: Oxford University Press.

Vrooman, J. C. and J. Dronkers (1986). Changing educational attainment processes: Some evidence from the Netherlands. *Sociology of Education 59*(2), 69–78.

Appendix A   Counting processes and proofs

To prove the consistency and the properties of our estimation strategy we rely on counting process theory for duration models. In a Mixed Proportional Hazard (MPH) model the waiting time to some event T has a conditional distribution given observed X, treatment D, mediator Q and unobserved heterogeneity V with hazard rate

$$\lambda(t|D, X, Q, V) = V\lambda_0(t)\exp(\beta'X + \gamma D), \tag{A.1}$$

The cdf and pdf of the distribution of the duration T can be expressed as functions of the hazard rate. The counting process approach has increasingly become the standard framework for analysing duration data and Andersen et al. (1993) have provided an excellent survey of this literature. Less technical surveys have been given by Klein and Moeschberger (1997), Therneau and Grambsch (2000), and Aalen et al. (2009). The main advantage of this framework is that it allows us to express the duration distribution as a regression model with an error term that is a martingale difference. Regression models with martingale difference errors are the basis for inference in time series models with dependent observations. Hence, it is not surprising that inference is much simplified by using a similar representation in duration models.

To start the discussion, we first introduce some notation. A counting process $\{N(t); t \geq 0\}$ is a stochastic process describing the number of events in the interval $[0, t]$ as time proceeds. The process contains only jumps of size $+1$. For single duration data, the event can only occur once, because the units are observed until the event occurs. Therefore we introduce the observation indicator $Y(t) = I(T \geq t)$ that equal to 1 if the unit is under observation at time t and zero after the event has occurred. The counting process is governed by its random intensity process $Y(t)\lambda(t)$, with $\lambda(t)$ is the hazard in (A.1). If we consider a small interval $(t - dt, t]$ of length dt, then $Y(t)\lambda(t)$ is the conditional probability that the increment $dN(t) = N(t) - N(t - dt)$ jumps in that interval given all that has happened until just before t. By specifying the intensity as the product of this observation indicator and the hazard rate we effectively limit the number of occurrences of the event to one. It is essential that the observation indicator only depends on events up to time t.

Usually, some of the observations are right-censored $\tilde{T} = \min(T, C_r)$. By defining the observation indicator as the product of the indicator $I(t \leq T)$ and, if necessary, an indicator of the observation plan, we capture when a unit is at risk for the event. A related concept is *left-truncation*. Left truncation occurs when individuals are only observed conditional on survival up till some duration C_l, the age of military examination in our application. In the case of right censoring and left-truncation the at-risk indicator: $Y(t) = I(t \leq T)I(t \leq C_r)I(t \geq C_l)$. We assume that C_r, C_l and T are conditionally independent given X. The history up to t, $\overline{Y}(t)$ is assumed to be a left continuous function of t. The history of the whole process also includes the (history of the) the covariates, treatment and mediator. Thus, we have

$$\Pr\big(dN(t) = 1|\overline{Y}(t), D, X, Q, V\big) = Y(t)\lambda(t|X, D, Q, V) \tag{A.2}$$

A fundamental result in the theory of counting processes, the *Doob-Meyer decomposition*[14],

---

[14]Doob (1953) published the Doob decomposition theorem which gives a unique decomposition for certain discrete time martingales. Meyer (1963) proved a continuous time version of the theorem, which became known as the Doob-Meyer decomposition. Both Andersen et al. (1993) and Aalen et al. (2009) provide a thorough discussion of this theorem.

allows us to write

$$dN(t) = Y(t)\lambda(t|\overline{X}(t)D, X, Q)dt + dM(t) \tag{A.3}$$

with $M(t), t \geq 0$ a martingale with conditional mean and variance

$$\mathrm{E}\big(dM(t)|\overline{Y}(t), D, X, Q\big) = 0 \tag{A.4}$$
$$\mathrm{Var}\big(dM(t)|\overline{Y}(t), D, X, Q\big) = Y(t)\lambda(t|D, X, Q, V)dt \tag{A.5}$$

The (conditional) mean and variance of the counting process are equal, so that the disturbances in (A.3) are heteroscedastic. The probability in (A.2) is 0, if the unit is no longer under observation. A counting process can be considered as a sequence of Bernoulli experiments because, if dt is small, (A.4) and (A.5) give the mean and variance of a Bernoulli random variable. The relation between the counting process and the sequence of Bernoulli experiments is given in (A.3), which can be considered as a regression model with an additive error that is a martingale difference. This equation resembles a time-series regression model. The Doob-Meier decomposition is the key to the derivation of the distribution of the estimators, because the asymptotic behavior of partial sums of martingales is well-known.

We begin with a proof of the unbiased of the inverse probability weighted Gompertz PH-model estimator given in equation (3). This applies for the model in which the hazard function does not include a mediator Q.

Proof of equation (3): IPW Gompertz is unbiased

In a parametric PH model the log-likelihood in counting process notation is (Andersen and Borgan 1985):

$$\ln L_i = \int \big[\ln \lambda_0(t_i; \alpha) + \gamma D_i\big] dN(t_i) - \int_0^1 Y_i(s)\lambda_0(s; \alpha)e^{\gamma D_i} ds \tag{A.6}$$

where $\lambda_0(t; \alpha)$ is the baseline hazard with parameters α, e.g. for a Gompertz baseline hazard $\lambda_0(t; \alpha) = e^{\alpha_0 + \alpha_1 t}$. Standard maximum likelihood estimation solves the roots of the derivatives of the log-likelihood:

$$L_\alpha(\theta) = \sum_{i=1}^N \left[\int \frac{\partial \lambda_0(t_i; \alpha)/\partial\alpha}{\lambda_0(t_i; \alpha)} dN(t_i) - \int_0^1 Y_i(s)\frac{\partial\lambda_0(s;\alpha)}{\partial\alpha}e^{\gamma D_i} ds\right] \tag{A.7}$$

$$U_\gamma(\theta) = \sum_{i=1}^N \left[\int D_i dN(t_i) - D_i \int_0^1 Y_i(s)\lambda_0(s; \alpha)e^{\gamma D_i} ds\right] \tag{A.8}$$

with $\theta = (\alpha, \gamma)'$ and $L_\alpha(\theta)$ and $L_\gamma(\theta)$ are the gradients of the log-likelihood w.r.t. α and γ. The IPW version includes the weights W in equation (A.7) and (A.8). Because our main parameter of interest is γ we only focus on $L_\gamma(\theta)$.

First we derive $\mathrm{E}\big[WDdN(t)\big]$. Redefine the propensity score $p(d) = \mathrm{Pr}(D_i = d|X_i)$, with $d = 0, 1$. Note that the integral of the sum is equal to the sum of the integrals. First, we derive $\mathrm{E}[WDdN(t)]$ and $\mathrm{E}[\sum Y(t)\lambda_0(t; \alpha)e^{\gamma D}WD]$.

$$\begin{aligned}
\mathrm{E}\big[WDdN(t)\big] &= \mathrm{E}\big[\lambda(t|D, X)S(t|D, X)WDdt\big] \\
&= \int \sum_d p(d)\mathrm{E}\Big[f(t|D = d, X = x)\frac{d}{p(d)}\Big]dtf_X(x)dx \tag{A.9} \\
&= f(t|D = 1)dt
\end{aligned}$$

$$\mathrm{E}\Big[\sum Y(t)W\lambda_0(t;\alpha)e^{\gamma D}D\Big] \;=\; \mathrm{E}\big[\lambda_0(t;\alpha)e^{\gamma D}S(t|D,X)WD\big] \tag{A.10}$$

$$= \int\sum_d p(d)\mathrm{E}\Big[\lambda_0(t;\alpha)e^{\gamma D}S(t|D=d,X=x)\frac{de^{\gamma d}}{p(d)}\Big]f_X(x)\mathrm{d}x$$

$$= e^{\gamma}\lambda_0(t;\alpha)S(t|D=1)$$

$$= f(t|D=1)$$

From (A.9) we have $\mathrm{E}[WD\mathrm{d}N(t)] = f(t|D=1)\mathrm{d}t$. Thus, if we assume the right parametric model this implies that $L_\gamma(\theta)$ has zero mean.

Proof of equation (3)**: IPW Gamma-Gompertz is unbiased**
In a MPH model with a parametric baseline hazard and a unit-mean Gamma-distributed unobserved heterogeneity with variance σ^2 the (unconditional) hazard is:

$$\lambda(t|D) = \frac{\lambda_0(t;\alpha)e^{\gamma D}}{1+\sigma^2\int_0^t \lambda_0(s;\alpha)e^{\gamma D}\,ds}$$

and the likelihood (in counting process notation) is:

$$L_i = \Big[\frac{\lambda_0(t;\alpha)e^{\gamma D}}{1+\sigma^2\int Y_i(s)\lambda_0(s;\alpha)e^{\gamma D}\,ds}\Big]^{\mathrm{d}N_i(t)}\Big[1+\sigma^2\int Y_i(s)\lambda_0(s;\alpha)e^{\gamma D}\,ds\Big]^{-1/\sigma^2} \tag{A.11}$$

IPW solves the roots of the weighted derivatives of the log-likelihood. The weighted derivative w.r.t. γ is:

$$L_\gamma(\theta) = \sum_{i=1}^N\Big[\int \frac{W_iD_i}{1+\sigma^2\int Y_i(s)\lambda_0(s;\alpha)e^{\gamma D}\,ds}\mathrm{d}N_i(t) - \frac{W_iD_i\int Y_i(t)\lambda_0(t;\alpha)e^{\gamma D}\,dt}{1+\sigma^2\int Y_i(s)\lambda_0(s;\alpha)e^{\gamma D}\,ds}\Big] \tag{A.12}$$

To prove (3) we use similar reasoning as above. First, we derive

$$\mathrm{E}\Big[\frac{WD}{1+\sigma^2\int Y(s)\lambda_0(s;\alpha)e^{\gamma D}\,ds}\mathrm{d}N(t)\Big] \;=\; \mathrm{E}\Big[\frac{WD\lambda_0(t;\alpha)e^{\gamma D}Y(t)dt}{1+\sigma^2\int_0^t \lambda_0(s;\alpha)e^{\gamma D}\,ds}\Big] \tag{A.13}$$

$$= \int\sum_d p(d)\mathrm{E}\Big[\frac{\lambda_0(t;\alpha)e^{\gamma D}S(t|D=d,X=x)}{1+\sigma^2\int_0^t \lambda_0(s;\alpha)e^{\gamma D}\,ds}\frac{de^{\gamma d}}{p(d)}\Big]\mathrm{d}t f_X(x)\mathrm{d}x$$

$$= \frac{\lambda_0(t;\alpha)e^{\gamma D}}{1+\sigma^2\int_0^t \lambda_0(s;\alpha)e^{\gamma D}\,ds}S(t|D=1)\mathrm{d}t$$

$$= f(t|D=1)\mathrm{d}t$$

and

$$\mathrm{E}\Big[\sum \frac{WDY(t)\lambda_0(t;\alpha)e^{\gamma D}}{1+\sigma^2\int Y(s)\lambda_0(s;\alpha)e^{\gamma D}\,ds}\Big] \;=\; \tag{A.14}$$

$$= \int\sum_d p(d)\mathrm{E}\Big[\frac{\lambda_0(t;\alpha)e^{\gamma D}S(t|D=d,X=x)}{1+\sigma^2\int_0^t \lambda_0(s;\alpha)e^{\gamma D}\,ds}\frac{de^{\gamma d}}{p(d)}\Big]f_X(x)\mathrm{d}x$$

$$= f(t|D=1)$$

29

Thus, if we assume the right parametric model for the baseline hazard and a Gamma distribution for the unobserved heterogeneity (A.13) has mean zero.[15]

We now turn our attention to specifications that include a mediator Q and provide a proof for Theorem 1 on the identification of the decomposition.

Proof Theorem 1 and equation (5):

The direct effect $\theta(d)$ solves $\mathrm{E}\big[L(\theta(d))\big] = 0$ with $L(\theta(d))$ as in (A.12) with $W_i = W_i(d)$.

$$
\mathrm{E}\Big[\frac{W(d)D}{1 + \sigma^2 \int Y(s)\lambda_0(s;\alpha)e^{\theta(d)D}\, ds}\mathrm{d}N(t)\Big] = \mathrm{E}\Big[\frac{W(d)D\lambda_0(t;\alpha)e^{\theta(d)D}Y(t)dt}{1 + \sigma^2 \int_0^t \lambda_0(s;\alpha)e^{\theta(d)D}\, ds}\Big]
$$

$$
= \int \mathrm{E}\Big[\frac{\lambda_0(t;\alpha)e^{\theta(d)}S(t|D=1,Q=q,X=x)}{1 + \sigma^2 \int_0^t \lambda_0(s;\alpha)e^{\theta(d)}\, ds}\frac{\Pr(D=d|Q,X)f_Q(q|x)}{\Pr(D=d|X)}\Big]\mathrm{d}q f_X(x)\mathrm{d}x
$$

$$
= \frac{\lambda_0(t;\alpha)e^{\theta(d)}}{1 + \sigma^2 \int_0^t \lambda_0(s;\alpha)e^{\theta(d)}\, ds}S(t|D=1,Q(d))\mathrm{d}t \quad = f(t|D=1Q(d))\mathrm{d}t \quad \text{(A.15)}
$$

and

$$
\mathrm{E}\Big[\sum \frac{W(d)DY(t)\lambda_0(t;\alpha)e^{\theta(d)D}}{1 + \sigma^2 \int Y(s)\lambda_0(s;\alpha)e^{\theta(d)D}\, ds}\Big] =
$$

$$
= \int \mathrm{E}\Big[\frac{\lambda_0(t;\alpha)e^{\theta(d)}S(t|D=1,Q=q,X=x)}{1 + \sigma^2 \int_0^t \lambda_0(s;\alpha)e^{\theta(d)}\, ds} \quad \frac{\Pr(D=d|Q,X)f_Q(q|x)}{\Pr(D=d|X)}\Big]f_X(x)\mathrm{d}x
$$

$$
= f(t|D=1,Q(d)) \quad \text{(A.16)}
$$

---

[15]The proof for any other MPH model with known functional form of the baseline hazard and given distribution of the unobserved heterogeneity is essentially the same.

Appendix B   Additional Tables and Figures

Table B.1: Sample distribution by education level (selection for unknown mortality status)

	mortality status		
	unknown	known	difference
Education level:			
Primary	12.3	14.4	-2.1^+
	(0.7)	(0.2)	(0.7)
Lower vocational	32.5	36.6	-4.2^+
	(1.0)	(0.2)	(1.0)
Lower secondary	35.7	33.0	2.7^+
	(0.7)	(0.2)	(1.0)
Higher	19.6	16.1	3.5^+
	(0.7)	(0.2)	(0.8)
Birth order	2.35	2.41	0.06
	(0.01)	(0.03)	(0.03)
Region of birth:			
North	6.1	3.7	2.4^+
	(0.7)	(0.2)	(0.4)
South	4.6	6.2	-1.6^+
	(0.7)	(0.2)	(0.5)
East	2.9	4.7	-1.8^+
	(0.7)	(0.2)	(0.4)
North-Holland	30.2	34.5	-4.3^+
	(0.7)	(0.2)	(1.0)
South-Holland	34.7	42.8	-8.0^+
	(0.7)	(0.2)	(1.0)
Utrecht	21.6	8.3	13.3^+
	(0.7)	(0.2)	(0.6)

$^+p < 0.05$

	mortality status		
	unknown	known	difference
Religion:			
Catholic	34.4	32.8	1.6
	(0.7)	(0.2)	(1.0)
Dutch Reformed	30.5	30.2	0.3
	(0.7)	(0.2)	(0.9)
Calvin	6.5	7.6	−1.0
	(0.7)	(0.2)	(0.5)
Other religion	1.8	0.7	1.1[+]
	(0.7)	(0.2)	(0.2)
No religion	26.7	28.7	−2.0
	(0.7)	(0.2)	(1.2)
Father's occupation:			
Professional	21.2	16.7	4.5[+]
	(0.7)	(0.2)	(0.8)
White collar	28.5	28.3	0.2
	(0.7)	(0.2)	(0.5)
Self-employed	6.1	6.5	−0.4
	(0.7)	(0.2)	(0.9)
Skilled	25.0	26.9	−2.8[+]
	(0.7)	(0.2)	(0.9)
Unskilled	9.6	12.4	−2.8[+]
	(0.7)	(0.2)	(0.7)
Unknown	10.5	9.1	1.4
	(0.7)	(0.2)	(0.8)
Global comprehensive IQ score:			
1 (highest)	21.3	17.6	3.6[+]
	(0.7)	(0.2)	(0.8)
2	33.4	32.5	0.9
	(0.7)	(0.2)	(1.0)
3	20.2	20.6	−0.4
	(0.7)	(0.2)	(0.8)
4	12.1	14.9	−2.8[+]
	(0.7)	(0.2)	(0.7)
5	8.9	10.1	−1.2
	(0.7)	(0.2)	(0.6)
6 (lowest)	1.6	2.0	−0.4
	(0.7)	(0.2)	(0.3)
missing	2.6	2.3	0.3
	(0.7)	(0.2)	(0.3)
Sample size	2,626	39,803	42,303

[+]$p < 0.05$

Table B.2: Probit estimates of propensity scores, pairwise comparisons

	Primary to lower vocational	Lower vocational to lower secondary	Lower secondary to higher
Father's occupation:			
Professional	-0.244^{**}	0.154^{**}	0.463^{**}
Self-employed	-0.367^{**}	-0.040	-0.282^{**}
Clerical	$-$	$-$	$-$
Skilled	-0.405^{**}	-0.392^{**}	-0.584^{**}
Unskilled	-0.590^{**}	-0.447^{**}	-0.620^{**}
Missing	-0.473^{**}	-0.309^{**}	-0.214^{**}
Family size	0.222^{**}	0.186^{**}	0.007
Born in Utrecht	-0.255^{**}	0.048	0.079^{+}
Religion:			
Catholic	-0.077^{**}	-0.053	0.044
Dutch Reformed	0.163^{**}	-0.021	-0.049
Calvinist	0.436^{**}	0.035	0.028
Other religion	0.016	0.213^{+}	0.088
None	$-$	$-$	$-$
Famine exposure:			
postnatal	0.025	0.142^{**}	0.003
third trimester	0.048	0.041	-0.032
second trimester	-0.033	0.068^{**}	-0.054
first trimester	-0.043	0.048	0.042
just before conception	-0.029	0.158^{**}	-0.017

$^{+}p < 0.05$ and $^{**}p < 0.01$

Table B.3: Probit estimates of propensity scores, pairwise comparisons including IQ

	Primary to lower vocational	Lower vocational to lower secondary	Lower secondary to higher
Father's occupation:			
Professional	−0.236**	0.138**	0.477**
Self-employed	−0.328**	0.040	−0.181**
Clerical	−	−	−
Skilled	−0.277**	−0.300**	−0.512**
Unskilled	−0.387**	−0.292**	−0.479**
Missing	−0.277**	−0.205**	−0.139**
Family size	0.166**	0.178**	0.082$^+$
Born in Utrecht	−0.136**	0.117**	0.139**
Religion:			
Catholic	−0.112**	−0.054**	0.068$^+$
Dutch Reformed	0.163**	−0.027	−0.051
Calvinist	0.326**	−0.036	−0.011
Other religion	0.031	0.259**	0.287**
None	−	−	−
Famine exposure:			
postnatal	−0.086**	0.093**	−0.027
third trimester	−0.020	0.004	−0.061$^+$
second trimester	−0.108**	0.019	−0.106**
first trimester	−0.127**	0.009	0.028
just before conception	−0.152**	0.093**	−0.051
Comprehensive IQ:			
1 (highest)	1.316**	0.884**	1.525**
2	0.581**	0.556**	0.779**
3	−	−	−
4	−0.585**	−0.453**	−0.388**
5	−1.263**	−0.838**	−0.472**
6 (lowest)	−2.076**	−0.944**	0.054
missing	−0.866**	0.412**	1.028**

$^+p < 0.05$ and $^{**}p < 0.01$

Table B.4: Impact of education on the mortality rate and its decomposition using an IPW Gompertz-gamma MPH, IPW accounting for selective unknown status

	Total	other pathways		cognitive ability	
		$\theta(1)$	$\theta(0)$	$\eta(0)$	$\eta(1)$
Primary to	-0.223^{**}	-0.061	-0.093^{+}	-0.162^{+}	-0.130^{+}
lower vocational	(0.034)	(0.066)	(0.045)	(0.075)	(0.056)
Lower vocational to	-0.086^{**}	0.006	0.016	-0.091^{+}	-0.102^{+}
lower secondary	(0.029)	(0.033)	(0.039)	(0.044)	(0.048)
Lower secondary to	-0.206^{**}	-0.127^{+}	-0.098	-0.079	-0.107
higher	(0.048)	(0.053)	(0.069)	(0.071)	(0.084)

$^{+}p < 0.05$ and $^{**}p < 0.01$

35

Table B.5: Sensitivity analysis : outcome, selection and mediator effects

	Primary to lower vocational			Lower vocational to lower secondary			Lower secondary to higher		
	ω	ξ	ψ	ω	ξ	ψ	ω	ξ	ψ
Father's occupation:									
Professional	−0.147	0.169$^+$	−0.064	−0.026	0.592**	−0.180**	−0.062	1.106**	−0.547**
	(0.140)	(0.077)	(0.059)	(0.089)	(0.049)	(0.044)	(0.079)	(0.050)	(0.044)
Self-employed	0.106	−0.033	0.015	0.102	0.211**	−0.061	0.022	−0.421**	0.116
	(0.155)	(0.091)	(0.072)	(0.108)	(0.070)	(0.062)	(0.120)	(0.098)	(0.073)
Skilled	−0.010	−0.205**	0.071	−0.088	−0.510**	0.145**	0.001	−1.089**	0.264**
	(0.082)	(0.048)	(0.039)	(0.060)	(0.039)	(0.032)	(0.072)	(0.071)	(0.048)
Unskilled	0.074	−0.523**	0.200**	0.016	−0.531**	0.155**	0.059	−1.050**	0.250**
	(0.097)	(0.058)	(0.047)	(0.076)	(0.054)	(0.044)	(0.100)	(0.110)	(0.070)
Missing	−0.006	−0.229**	0.091	0.290**	−0.212**	0.068	0.149	−0.300**	0.082
	(0.145)	(0.055)	(0.072)	(0.102)	(0.075)	(0.067)	(0.133)	(0.116)	(0.059)
Family size	0.176	0.333**	−0.105	0.178	0.277**	−0.023	0.102	−0.080	0.029
	(0.166)	(0.099)	(0.070)	(0.101)	(0.066)	(0.064)	(0.109)	(0.084)	(0.058)
Born in Utrecht	0.246$^+$	−0.413**	0.164$^+$	0.203$^+$	0.087	−0.125$^+$	0.177	0.143	−0.037
	(0.121)	(0.079)	(0.065)	(0.098)	(0.065)	(0.064)	(0.109)	(0.082)	(0.070)
Religion:									
Catholic	0.011	−0.334**	0.117**	−0.113	−0.110**	0.168**	0.000	0.052	−0.021
	(0.080)	(0.047)	(0.039)	(0.060)	(0.037)	(0.030)	(0.069)	(0.048)	(0.042)
Dutch Reformed	−0.014	0.278**	−0.098$^+$	0.032	0.013	−0.011	0.051	−0.052	0.010
	(0.094)	(0.053)	(0.040)	(0.060)	(0.037)	(0.031)	(0.067)	(0.049)	(0.042)
Calvinist	−0.216	0.808**	−0.244**	−0.012	0.127$^+$	−0.327**	−0.233	0.082	−0.028
	(0.236)	(0.111)	(0.068)	(0.103)	(0.064)	(0.057)	(0.127)	(0.079)	(0.068)
Other religion	−0.050	−0.009	0.019	−0.022	0.381	−0.289	−0.320	0.261	−0.071
	(0.548)	(0.309)	(0.241)	(0.361)	(0.225)	(0.170)	(0.448)	(0.235)	(0.197)
Famine exposure:									
postnatal	0.109	0.053	−0.017	0.142$^+$	0.147**	−0.041	0.157$^+$	−0.026	0.010
	(0.107)	(0.062)	(0.046)	(0.069)	(0.044)	(0.038)	(0.079)	(0.058)	(0.049)
third trimester	0.177	0.089	−0.030	0.097	0.099$^+$	−0.030	0.064	−0.083	0.024
	(0.106)	(0.062)	(0.046)	(0.069)	(0.043)	(0.038)	(0.082)	(0.059)	(0.050)
second trimester	0.115	−0.060	0.023	0.106	0.051	−0.015	0.050	−0.063	0.020
	(0.109)	(0.065)	(0.049)	(0.075)	(0.048)	(0.041)	(0.090)	(0.062)	(0.052)
first trimester	0.108	−0.123	0.046	0.053	0.145**	−0.044	0.006	0.111	−0.032
	(0.125)	(0.076)	(0.061)	(0.091)	(0.058)	(0.050)	(0.104)	(0.072)	(0.061)
pre-conception	−0.046	−0.050	0.015	−0.065	0.208**	−0.067	−0.102	0.050	−0.018
	(0.113)	(0.063)	(0.048)	(0.077)	(0.045)	(0.039)	(0.087)	(0.060)	(0.050)
Comprehensive IQ									
1 (highest)	−1.280	3.962**	−0.508**	−0.393**	1.292**	−0.376**	−0.188$^+$	1.557**	−0.494**
	(4.701)	(0.607)	(0.079)	(0.131)	(0.059)	(0.046)	(0.087)	(0.050)	(0.041)
2	−0.359	2.296**	−0.522**	−0.118$^+$	0.883**	−0.268**	−0.065	−0.419**	0.118**
	(0.250)	(0.103)	(0.042)	(0.063)	(0.036)	(0.031)	(0.065)	(0.047)	(0.037)
4	−0.131	−0.298**	0.107$^+$	0.098	−1.326**	0.354**	0.267$^+$	−2.508**	0.400**
	(0.090)	(0.052)	(0.042)	(0.063)	(0.056)	(0.040)	(0.109)	(0.240)	(0.082)
5	0.151	−1.709**	0.701**	0.179$^+$	−1.959**	0.444**	0.426	−2.804**	0.406$^+$
	(0.080)	(0.055)	(0.047)	(0.084)	(0.103)	(0.061)	(0.216)	(0.586)	(0.176)
6 (lowest)	0.159	−2.772**	1.045**	0.335	−2.190**	0.460$^+$	0.275	−0.913	0.358
	(0.120)	(0.153)	(0.085)	(0.268)	(0.461)	(0.218)	(2.611)	(1.137)	(0.771)
missing	0.206	−0.612**	0.249$^+$	−0.064	0.326**	−0.092	0.143	0.245	−0.064
	(0.216)	(0.136)	(0.125)	(0.230)	(0.123)	(0.106)	(0.201)	(0.136)	(0.127)

Based on adding U to propensity score with probabilities of U from observed probabilities for each covariate.
No effect would give $\omega = 0, \xi = 0$ and $\psi = 0$. $^+p < 0.05$ and $^{**}p < 0.01$

Table B.6: Sensitivity analysis total effect

	Primary to lower vocational	Lower vocational to lower secondary	Lower secondary to higher
original	-0.222^{**}	-0.086^{**}	-0.206^{**}
	(0.034)	(0.029)	(0.048)
Father's occupation:			
Professional	-0.220^{**}	-0.083^{**}	-0.200^{**}
	(0.034)	(0.029)	(0.052)
Self-employed	-0.221^{**}	-0.087^{**}	-0.204^{**}
	(0.034)	(0.029)	(0.045)
Skilled	-0.223^{**}	-0.085^{**}	-0.189^{**}
	(0.035)	(0.029)	(0.054)
Unskilled	-0.218^{**}	-0.082^{**}	-0.195^{**}
	(0.036)	(0.029)	(0.051)
Missing	-0.220^{**}	-0.084^{**}	-0.204^{**}
	(0.034)	(0.029)	(0.048)
Family size	-0.225^{**}	-0.088^{**}	-0.205^{**}
	(0.034)	(0.029)	(0.048)
Born in Utrecht	-0.214^{**}	-0.087^{**}	-0.208^{**}
	(0.035)	(0.029)	(0.048)
Religion:			
Catholic	-0.224^{**}	-0.087^{**}	-0.207^{**}
	(0.036)	(0.029)	(0.048)
Dutch Reformed	-0.222^{**}	-0.086^{**}	-0.206^{**}
	(0.036)	(0.029)	(0.048)
Calvinist	-0.217^{**}	-0.085^{**}	-0.205^{**}
	(0.036)	(0.029)	(0.048)
Other religion	-0.222^{**}	-0.085^{**}	-0.205^{**}
	(0.034)	(0.029)	(0.048)
Famine exposure:			
postnatal	-0.222^{**}	-0.089^{**}	-0.206^{**}
	(0.034)	(0.029)	(0.048)
third trimester	-0.224^{**}	-0.087^{**}	-0.206^{**}
	(0.034)	(0.029)	(0.048)
second trimester	-0.221^{**}	-0.088^{**}	-0.206^{**}
	(0.034)	(0.029)	(0.048)
first trimester	-0.221^{**}	-0.086^{**}	-0.206^{**}
	(0.034)	(0.029)	(0.048)
pre-conception	-0.222^{**}	-0.083^{**}	-0.206^{**}
	(0.034)	(0.029)	(0.048)

Based on adding U to propensity score with probabilities of U from observed probabilities for observed variables. $^{+}p < 0.05$ and $^{**}p < 0.01$

Table B.7: Sensitivity analysis (mediator): effect running through cognitive ability and running through other pathways

| | Primary to lower vocational | | | | Lower vocational to lower secondary | | | | Lower secondary to higher | | | |
| | other pathways | | cognitive ability | | other pathways | | cognitive ability | | other pathways | | cognitive ability | |
	$\theta(1)$	$\theta(0)$	$\eta(0)$	$\eta(1)$	$\theta(1)$	$\theta(0)$	$\eta(0)$	$\eta(1)$	$\theta(1)$	$\theta(0)$	$\eta(0)$	$\eta(1)$
original	-0.060	-0.093^{+}	-0.162^{+}	-0.128^{+}	0.006	0.014	-0.092^{+}	-0.100^{+}	-0.127^{+}	-0.097	-0.079	-0.109
	(0.067)	(0.045)	(0.075)	(0.056)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.070)	(0.071)	(0.085)
Father's occupation:												
Professional	-0.058	-0.092^{+}	-0.163^{+}	-0.128^{+}	0.009	0.017	-0.092^{+}	-0.100^{+}	-0.120^{+}	-0.091	-0.080	-0.109
	(0.067)	(0.045)	(0.075)	(0.056)	(0.033)	(0.039)	(0.044)	(0.049)	(0.057)	(0.081)	(0.077)	(0.096)
Self-employed	-0.059	-0.093^{+}	-0.162^{+}	-0.128^{+}	0.005	0.013	-0.092^{+}	-0.100^{+}	-0.125^{+}	-0.095	-0.079	-0.109
	(0.067)	(0.045)	(0.075)	(0.056)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.071)	(0.072)	(0.086)
Skilled	-0.061	-0.094^{+}	-0.162^{+}	-0.129^{+}	0.007	0.016	-0.092^{+}	-0.100^{+}	-0.110	-0.083	-0.079	-0.106
	(0.068)	(0.045)	(0.076)	(0.057)	(0.033)	(0.040)	(0.044)	(0.049)	(0.059)	(0.094)	(0.080)	(0.108)
Unskilled	-0.056	-0.089	-0.162	-0.129	0.010	0.018	-0.092	-0.100	-0.115^{+}	-0.085	-0.079	-0.110
	(0.070)	(0.046)	(0.078)	(0.058)	(0.033)	(0.039)	(0.044)	(0.049)	(0.056)	(0.084)	(0.075)	(0.098)
Missing	-0.059	-0.092^{+}	-0.162^{+}	-0.128^{+}	0.008	0.016	-0.092^{+}	-0.100^{+}	-0.125^{+}	-0.095	-0.079	-0.109
	(0.067)	(0.045)	(0.075)	(0.056)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.071)	(0.071)	(0.085)
Family size	-0.062	-0.097^{+}	-0.163^{+}	-0.129^{+}	0.003	0.011	-0.092	-0.100^{+}	-0.126^{+}	-0.096	-0.079	-0.109
	(0.067)	(0.045)	(0.076)	(0.057)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.070)	(0.071)	(0.085)
Born in Utrecht	-0.051	-0.086	-0.163^{+}	-0.128^{+}	0.005	0.013	-0.092	-0.100^{+}	-0.129^{+}	-0.099	-0.079	-0.109
	(0.068)	(0.045)	(0.076)	(0.057)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.070)	(0.071)	(0.085)
Religion:												
Catholic	-0.061	-0.095^{+}	-0.163^{+}	-0.129^{+}	0.005	0.013	-0.092	-0.100^{+}	-0.128^{+}	-0.098	-0.079	-0.109
	(0.069)	(0.045)	(0.078)	(0.058)	(0.033)	(0.039)	(0.044)	(0.045)	(0.053)	(0.070)	(0.071)	(0.055)
Dutch Reformed	-0.059	-0.093^{+}	-0.163	-0.128	0.006	0.014	-0.092	-0.100	-0.127^{+}	-0.098	-0.079	-0.109
	(0.068)	(0.045)	(0.077)	(0.057)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.070)	(0.071)	(0.085)
Calvinist	-0.050	-0.088	-0.167^{+}	-0.129^{+}	0.007	0.015	-0.092	-0.100^{+}	-0.126^{+}	-0.096	-0.079	-0.109
	(0.072)	(0.046)	(0.081)	(0.058)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.070)	(0.071)	(0.085)
Other religion	-0.060	-0.093^{+}	-0.162^{+}	-0.128^{+}	0.007	0.015	-0.092	-0.100^{+}	-0.126^{+}	-0.096	-0.079	-0.109
	(0.067)	(0.045)	(0.075)	(0.056)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.070)	(0.071)	(0.085)
Famine exposure:												
postnatal	-0.061	-0.094^{+}	-0.162^{+}	-0.128^{+}	0.003	0.010	-0.092^{+}	-0.099^{+}	-0.127^{+}	-0.097	-0.079	-0.109
	(0.067)	(0.045)	(0.075)	(0.056)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.070)	(0.071)	(0.085)
third trimester	-0.062	-0.095^{+}	-0.162^{+}	-0.128^{+}	0.005	0.013	-0.092^{+}	-0.100^{+}	-0.126^{+}	-0.097	-0.079	-0.108
	(0.067)	(0.045)	(0.075)	(0.056)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.070)	(0.071)	(0.085)
second trimester	-0.059	-0.093^{+}	-0.162^{+}	-0.128^{+}	0.006	0.014	-0.092^{+}	-0.100^{+}	-0.127^{+}	-0.097	-0.079	-0.109
	(0.067)	(0.045)	(0.075)	(0.056)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.070)	(0.071)	(0.085)
first trimester	-0.058	-0.092^{+}	-0.162	-0.128	0.006	0.014	-0.092	-0.100	-0.127^{+}	-0.097	-0.079	-0.109
	(0.067)	(0.045)	(0.075)	(0.056)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.070)	(0.071)	(0.085)
pre-conception	-0.060	-0.094^{+}	-0.162^{+}	-0.128^{+}	0.008	0.016	-0.092^{+}	-0.100^{+}	-0.127^{+}	-0.097	-0.079	-0.109
	(0.067)	(0.045)	(0.075)	(0.056)	(0.033)	(0.039)	(0.044)	(0.048)	(0.053)	(0.070)	(0.071)	(0.085)

Based on adding U to propensity score with probabilities of U from observed probabilities for each covariate. $^{+}p < 0.05$ and $^{**}p < 0.01$

Table B.8: Sensitivity analysis characterizing 'killer' confounders: outcome, selection and mediator effects

	Primary to lower vocational			Lower vocational to lower secondary		
	ω	ξ	ψ	ω	ξ	ψ
$d = 0.1$ & $s = 0.1$	-0.409^{**}	0.232^{**}	-0.085^{+}	0.003	0.097^{**}	-0.032^{**}
	(0.088)	(0.046)	(0.037)	(0.055)	(0.035)	(0.031)
$d = 0.1$ & $s = 0.2$	-0.467^{**}	0.667^{**}	-0.237^{**}	0.004	0.194^{**}	-0.062^{**}
	(0.099)	(0.045)	(0.037)	(0.055)	(0.034)	(0.031)
$d = 0.1$ & $s = 0.3$	-0.579^{**}	1.151^{**}	-0.385^{**}	0.004	0.295^{**}	-0.092^{**}
	(0.116)	(0.052)	(0.035)	(0.056)	(0.035)	(0.031)
$d = 0.1$ & $s = 0.4$	-0.859^{**}	1.733^{**}	-0.537^{**}	0.004	0.398^{**}	-0.122^{**}
	(0.146)	(0.058)	(0.034)	(0.055)	(0.035)	(0.031)
$d = 0.1$ & $s = 0.5$	-2.524^{**}	2.541^{**}	-0.691^{**}	0.004	0.514^{**}	-0.155^{**}
	(0.405)	(0.071)	(0.033)	(0.054)	(0.037)	(0.030)
$d = 0.2$ & $s = 0.1$	-0.782^{**}	0.268^{**}	-0.112^{**}	-0.255^{**}	0.067^{**}	-0.031^{**}
	(0.092)	(0.044)	(0.035)	(0.055)	(0.034)	(0.031)
$d = 0.2$ & $s = 0.2$	-0.889^{**}	0.464^{**}	-0.174^{**}	0.004	0.194^{**}	-0.062^{**}
	(0.107)	(0.046)	(0.036)	(0.055)	(0.034)	(0.031)
$d = 0.2$ & $s = 0.3$	-1.121^{**}	0.913^{**}	-0.320^{**}	0.004	0.295^{**}	-0.092^{**}
	(0.125)	(0.049)	(0.035)	(0.056)	(0.035)	(0.031)
$d = 0.2$ & $s = 0.4$	-1.723^{**}	1.427^{**}	-0.469^{**}	0.004	0.398^{**}	-0.122^{**}
	(0.175)	(0.053)	(0.034)	(0.055)	(0.035)	(0.031)
$d = 0.2$ & $s = 0.5$	-23.794^{**}	2.024^{**}	-0.612^{**}	0.004	0.514^{**}	-0.155^{**}
	(0.507)	(0.060)	(0.033)	(0.054)	(0.037)	(0.030)
$d = 0.3$ & $s = 0.1$	-1.159^{**}	0.318^{**}	-0.139^{**}	-0.512^{**}	0.073^{**}	-0.042^{**}
	(0.093)	(0.045)	(0.034)	(0.056)	(0.035)	(0.031)
$d = 0.3$ & $s = 0.2$	-1.328^{**}	0.485^{**}	-0.192^{**}	-0.241^{**}	0.132^{**}	-0.050^{**}
	(0.109)	(0.044)	(0.034)	(0.056)	(0.033)	(0.031)
$d = 0.3$ & $s = 0.3$	-1.702^{**}	0.701^{**}	-0.257^{**}	0.004	0.295^{**}	-0.092^{**}
	(0.136)	(0.047)	(0.035)	(0.056)	(0.035)	(0.031)
$d = 0.3$ & $s = 0.4$	-3.054^{**}	1.170^{**}	-0.404^{**}	0.004	0.398^{**}	-0.122^{**}
	(0.292)	(0.049)	(0.035)	(0.055)	(0.035)	(0.031)
$d = 0.3$ & $s = 0.5$	-25.194^{**}	1.634^{**}	-0.531^{**}	0.004	0.514^{**}	-0.155^{**}
	(0.765)	(0.054)	(0.033)	(0.054)	(0.037)	(0.030)
$d = 0.4$ & $s = 0.1$	-1.560^{**}	0.375^{**}	-0.171^{**}	-0.771^{**}	0.079^{**}	-0.053^{**}
	(0.093)	(0.045)	(0.034)	(0.054)	(0.036)	(0.030)
$d = 0.4$ & $s = 0.2$	-1.795^{**}	0.533^{**}	-0.220^{**}	-0.500^{**}	0.139^{**}	-0.060^{**}
	(0.113)	(0.045)	(0.034)	(0.057)	(0.034)	(0.030)
$d = 0.4$ & $s = 0.3$	-2.378^{**}	0.704^{**}	-0.270^{**}	-0.229^{**}	0.200^{**}	-0.068^{**}
	(0.151)	(0.044)	(0.033)	(0.056)	(0.035)	(0.031)
$d = 0.4$ & $s = 0.4$	-30.981^{**}	0.930^{**}	-0.336^{**}	0.004	0.398^{**}	-0.122^{**}
	(4.634)	(0.045)	(0.034)	(0.055)	(0.035)	(0.031)
$d = 0.4$ & $s = 0.5$	-27.204^{**}	1.321^{**}	-0.453^{**}	0.004	0.514^{**}	-0.155^{**}
	(0.862)	(0.049)	(0.033)	(0.054)	(0.037)	(0.030)
$d = 0.5$ & $s = 0.1$	-1.996^{**}	0.445^{**}	-0.211^{**}	-1.052^{**}	0.066^{**}	-0.067^{**}
	(0.097)	(0.044)	(0.034)	(0.052)	(0.036)	(0.030)
$d = 0.5$ & $s = 0.2$	-2.325^{**}	0.594^{**}	-0.255^{**}	-0.765^{**}	0.151^{**}	-0.073^{**}
	(0.123)	(0.044)	(0.035)	(0.054)	(0.035)	(0.030)
$d = 0.5$ & $s = 0.3$	-3.358^{**}	0.750^{**}	-0.302^{**}	-0.491^{**}	0.212^{**}	-0.082^{**}
	(0.227)	(0.044)	(0.034)	(0.057)	(0.036)	(0.030)
$d = 0.5$ & $s = 0.4$	-36.698^{**}	0.552^{**}	-0.335^{**}	-0.218^{**}	0.270^{**}	-0.088^{**}
	(4.947)	(0.044)	(0.033)	(0.055)	(0.035)	(0.031)
$d = 0.5$ & $s = 0.5$	-30.374^{**}	1.054^{**}	-0.378^{**}	0.004	0.514^{**}	-0.155^{**}
	(3.770)	(0.045)	(0.033)	(0.054)	(0.037)	(0.030)

Based on adding U to propensity score under the assumption that $\Pr(U = 1) = 0.4$ and $p_{11} - p_{10} = 0$, the differences $d = p_{01} - p_{00}$ and $s = p_{1.} - p_{0.}$. No effect would give $\omega = 0, \xi = 0$ and $\psi = 0$. $^{+}p < 0.05$ and $^{**}p < 0.01$

Table B.8: (continued)

	Lower secondary to higher		
	ω	ξ	ψ
$d = 0.1$ & $s = 0.1$	0.004^{**}	0.062^{**}	-0.020^{**}
	(0.065)	(0.046)	(0.037)
$d = 0.1$ & $s = 0.2$	0.005^{**}	0.121^{**}	-0.037^{**}
	(0.067)	(0.046)	(0.037)
$d = 0.1$ & $s = 0.3$	0.005^{**}	0.182^{**}	-0.054^{**}
	(0.066)	(0.046)	(0.038)
$d = 0.1$ & $s = 0.4$	0.004^{**}	0.258^{**}	-0.076^{**}
	(0.066)	(0.047)	(0.038)
$d = 0.1$ & $s = 0.5$	0.004^{**}	0.355^{**}	-0.103^{**}
	(0.069)	(0.052)	(0.040)
$d = 0.2$ & $s = 0.1$	-0.225^{**}	-0.131^{**}	0.031^{**}
	(0.065)	(0.045)	(0.038)
$d = 0.2$ & $s = 0.2$	0.005^{**}	0.121^{**}	-0.037^{**}
	(0.067)	(0.046)	(0.037)
$d = 0.2$ & $s = 0.3$	0.005^{**}	0.182^{**}	-0.054^{**}
	(0.066)	(0.046)	(0.038)
$d = 0.2$ & $s = 0.4$	0.004^{**}	0.258^{**}	-0.076^{**}
	(0.066)	(0.047)	(0.038)
$d = 0.2$ & $s = 0.5$	0.004^{**}	0.355^{**}	-0.103^{**}
	(0.069)	(0.052)	(0.040)
$d = 0.3$ & $s = 0.1$	-0.495^{**}	-0.370^{**}	0.095^{**}
	(0.068)	(0.046)	(0.038)
$d = 0.3$ & $s = 0.2$	-0.173^{**}	-0.029^{**}	0.003^{**}
	(0.067)	(0.046)	(0.037)
$d = 0.3$ & $s = 0.3$	0.005^{**}	$0.18,^{**}$	-0.054^{**}
	(0.066)	(0.046)	(0.038)
$d = 0.3$ & $s = 0.4$	0.004^{**}	0.258^{**}	-0.076^{**}
	(0.066)	(0.047)	(0.038)
$d = 0.3$ & $s = 0.5$	0.004^{**}	0.355^{**}	-0.103^{**}
	(0.069)	(0.052)	(0.040)
$d = 0.4$ & $s = 0.1$	-0.756^{**}	-0.613^{**}	0.162^{**}
	(0.069)	(0.047)	(0.039)
$d = 0.4$ & $s = 0.2$	-0.450^{**}	-0.270^{**}	0.069^{**}
	(0.069)	(0.047)	(0.038)
$d = 0.4$ & $s = 0.3$	-0.123^{**}	0.075^{**}	-0.026^{**}
	(0.066)	(0.047)	(0.037)
$d = 0.4$ & $s = 0.4$	0.004^{**}	0.258^{**}	-0.076^{**}
	(0.066)	(0.047)	(0.038)
$d = 0.4$ & $s = 0.5$	0.004^{**}	0.355^{**}	-0.103^{**}
	(0.069)	(0.052)	(0.040)
$d = 0.5$ & $s = 0.1$	-1.037^{**}	-0.874^{**}	0.236^{**}
	(0.064)	(0.047)	(0.039)
$d = 0.5$ & $s = 0.2$	-0.723^{**}	-0.525^{**}	0.141^{**}
	(0.071)	(0.047)	(0.039)
$d = 0.5$ & $s = 0.3$	-0.412^{**}	-0.173^{**}	0.041^{**}
	(0.069)	(0.048)	(0.039)
$d = 0.5$ & $s = 0.4$	-0.071^{**}	0.196^{**}	-0.059^{**}
	(0.065)	(0.047)	(0.038)
$d = 0.5$ & $s = 0.5$	0.004^{**}	0.355^{**}	-0.103^{**}
	(0.069)	(0.052)	(0.040)

Based on adding U to propensity score under the assumption that $\Pr(U = 1) = 0.4$ and $p_{11} - p_{10} = 0$, the differences $d = p_{01} - p_{00}$ and $s = p_{1.} - p_{0.}$. No effect would give $\omega = 0$, $\xi = 0$ and $\psi = 0$. $^{+}p < 0.05$ and $^{**}p < 0.01$