Pathways from education to mortality, mediated through income

Govert E. Bijwaard*[1] and Kieron J. Barclay[2]

[1]Netherlands Interdisciplinary Demographic Institute
[2]Max Planck Institute for Demographic Research

October 23, 2019

1 Introduction

The fact that individuals with lower socioeconomic status have higher mortality rates has been well established in social science research (Hummer and Lariscy, 2011). Of the various socioeconomic measures that are commonly used to investigate this relationship, the educational gradient has been shown to be particularly robust (Cutler et al., 2011). This educational gradient is apparent not only in less-developed countries, but in the US (Masters et al., 2012) and other Western countries with advanced health care systems (Huisman et al., 2005).

Educational attainment is the most commonly used indicator of socio-economic status in studies of health and mortality. Education is usually completed by early adulthood and remains constant over the life course and precedes other dimensions of socio-economic status. Educational attainment is strongly associated with other measurements of the socio-economic status, like income. It has also been established that lower income increases mortality. The association between education and mortality may, therefore, be running through the influence of education on income.

Despite a substantial association between education and health the causal interpretation of this relation has been challenged. This association may be confounded by factors that influence both education and health (Grossman, 2015). Moreover, surprisingly little research has investigated the underlying causal mechanism of education on health in the presence of one or more intermediate variables, such as income.

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 nonlinear duration outcomes such as the (mixed) proportional hazard model. Recent papers have placed causal mediation analysis within the counterfactual/ potential outcomes framework (Huber, 2014; Imai et al., 2010a,b; VanderWeele, 2015) all assuming sequential unconfoundedness.

Propensity score methods, that assume unconfoundedness, are increasingly used to take account of confounding in observational studies. 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). Huber (2014) derived a method to make causal inference for the direct and indirect effects of a treatment with intermediate variables. based on inverse propensity weighting (IPW). 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 (Austin, 2014; Cole and Hernán, 2004).

*Netherlands Interdisciplinary Demographic Institute (NIDI-KNAW/University of Groningen), PO Box 11650, 2502 AR The Hague, the Netherlands, +31 70 3565224, bijwaard@nidi.nl

Bijwaard and Jones (2019) developed an adjustment of the method of Huber (2014) to analyse the mediating effect in the context of survival models. We extend this further to allow for a sequence of, temporally ordered, mediators.

We seek to find how education affects mortality later in life and how this is mediated through income development. We use Swedish conscription data, linked to information on the parental socioeconomic situation at birth, the parental education, the education of the individual himself, date of death (up till 2012) and annual income for the period 1968 till 2012. We estimate four separate models for the educational gain in mortality using the data of individuals in two adjacent educational levels. For each pair we derive how much of the educational gain in the mortality rate can be attributed to the effect of education on the income development and how much can be viewed as a direct effect of education.

2 Data

The data come from several Swedish population-wide registers which are linked using unique individual identification. The Swedish Military Conscription Data includes demographic information of the conscripts and information obtained at the military examination, including a battery of intelligence tests. The data consist of the population of men born between 1950 and 1960, who were enlisted in the year they turned 18-20. We removed men for which we do not observe the education attainment. and men without a known conscription date. These data are linked to information on the parental socioeconomic situation at birth, the parental education, the education of the individual himself, date of death (up till 2012) and annual income. We aggregated the observed education into five classes: (i) Less than 10 years of education (only primary schooling); (ii) Secondary education (2 years); (iii) Full secondary education (3 years); (iv) Post secondary education (University and PhD) and (v) University and PhD. The data on the time-varying income cover the years 1968-2012.

Selected demographic and parental socioeconomic characteristics at the time of military examinations by education level are presented in Table 1. We see a clear positive relation between the maternal socioeconomic status, the paternal education and the education attained by the military conscript. The higher the social class and education of the parents the higher the education level of the conscript.

The Kaplan-Meier survival curves for the five education categories are shown in Figure 1. Survival increases with the education level and the differences between the education levels increase with age.

Figure 1: Kaplan-Meier survival curves, by education level

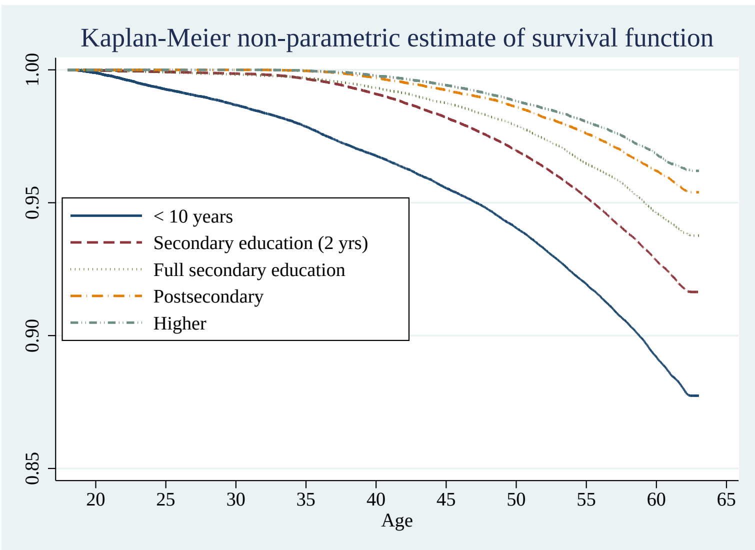

Table 1: Sample characteristics, Swedish Conscripts 1951-1960 ($N = 519,609$)

	Education level				
	primary	secondary (2 years)	secondary (3 years)	Post secondary	university
	Parental EGP				
high grade professional	1%	1%	1%	2%	2%
low grade professional	31%	40%	53%	64%	65%
routine non-manual	8%	7%	6%	4%	4%
small entrepreneur	8%	8%	5%	4%	3%
manual	45%	40%	27%	18%	14%
missing	7%	4%	7%	9%	11%
	Father's education				
less 10 years	54%	48%	32%	22%	18%
Secondary edu (max 12)	26%	27%	25%	19%	17%
Secondary edu (13)	10%	14%	19%	22%	20%
Post secondary	4%	6%	10%	12%	13%
university	4%	4%	13%	24%	32%
missing	2%	1%	1%	1%	0%
	Mother's education				
less 10 years	46%	39%	26%	19%	15%
Secondary edu (max 12)	39%	43%	40%	33%	29%
Secondary edu (13)	5%	5%	8%	9%	9%
Post secondary	5%	7%	12%	15%	16%
university	5%	5%	14%	24%	31%
missing	0%	0%	0%	0%	1%
	Birth order				
Birth order 1	46%	47%	52%	54%	54%
Birth order 2	32%	35%	34%	33%	34%
Birth order 3	14%	12%	11%	9%	9%
Birth order 4	5%	4%	2%	2%	2%
Birth order 5	2%	1%	1%	1%	0%
Birth order > 5	1%	1%	0%	0%	0%
	Family size				
1	20%	16%	16%	13%	13%
2	39%	45%	49%	52%	52%
3	25%	26%	26%	27%	27%
4	10%	8%	7%	6%	6%
5	3%	2%	2%	1%	1%
> 5	3%	2%	1%	1%	1%
	income				
average income	109	133	141	158	182
average log-income	4.45	4.80	4.84	4.96	5.10
$N =$	63,489	218,272	107,578	106,710	23,560

3 Static mediator and outcome

3.1 Counterfactual approach

The intuitive idea behind causation is represented by changes in the outcome due to changes in the treatment. To properly measure causality we should then compare what would have happen to the outcome for different settings of the treatment (education). The problem is that, for each individual, we can only observe the outcome for the education level attained, not for the alternative education level(s). Here we introduce the counterfactual framework for treatment evaluation. In our case the 'treatment' is the education level attained.

Define the potential outcome $Y(d)$ as the outcome an individual would have if the treatment were $D = d$. Thus for a binary treatment, $Y(0)$ is the outcome we could expect if the individual did not get the treatment and $Y(1)$ is the outcome we expect if the individual got the treatment. We only observe one of these outcomes and the observed outcome is $Y = D \cdot Y(1) + (1 - D) \cdot Y(0)$. The individual total effect $Y_i(1) - Y_i(0)$ is never observed, but under some conditions the average *total effect*, $\mathrm{E}\big[Y_i(1) - Y_i(0)\big]$ is identified. A typical condition that allows the identification of the total effect is random assignment of the treatment, because it implies that $\{Y(1), Y(0)\} \coprod D$, with $\coprod$ means independence. With random assignment the potential outcome is independent of the treatment.

In observational studies it is hardly ever reasonable to assume that the potential outcomes are independent of the (possibly induced by selective choice) treatment. If the treatment selection is based on observed individual characteristics, X it is allowed to assume conditional independence $\{Y(1), Y(0)\} \coprod D|X$. This is called the *unconfoundedness* or *ignorability* assumption (Rosenbaum and Rubin, 1983; Rubin, 1974). If the overlap assumption also holds the total effect is identified by using matching on or weighting by the propensity score. The overlap assumption, or common support assumption, requires that the propensity score, the conditional probability to select treatment given covariates X, is bounded away from zero and one.

If treatment selection depends on unobserved factors that also influence the potential outcome the ignorability assumption fails to hold. One approach to account for such unobserved factors is a structural equation model, that explicitly models the influence of these factors on the treatment and outcome.

3.2 Counterfactual mediation analysis

Recent papers have placed causal mediation analysis within the counterfactual/potential outcomes framework (Huber, 2014; Imai et al., 2010a,b). If a mediator also plays a role in the outcome the potential outcome is also a function of the mediator, $Y(d, m)$, the potential outcome that would have been observed if the treatment was d and the mediator was m. Next the mediator is also affected by the treatment and we have a potential mediator $M(d)$, the mediator that would have been observed if the treatment was equal to d. Then we can define the (average) *total effect* of the treatment as $\mathrm{E}\big[Y_i(1, M_i(1)) - Y_i(0, M_i(0))\big]$ and the *direct effect* of the treatment as $\mathrm{E}\big[Y_i(1, M_i(0)) - Y_i(0, M_i(0))\big]$. The *indirect effect* of the treatment running through the mediator M is defined as $\mathrm{E}\big[Y_i(1, M_i(1)) - Y_i(1, M_i(0))\big]$. Note that the sum of the direct and indirect effect is equal to the total effect of the treatment.

For identification of the direct and indirect effects we need a sequential ignorablility condition (Bijwaard and Jones, 2019; Huber, 2014; Imai et al., 2010a,b),

Sequential Ignorability Assumption.
The following two statements of conditional independence hold:
$\{Y_i(d', m), M_i(d)\} \perp D_i | X$ and $Y_i(d', m) \perp M_i(d) | D_i = d, X$, $\forall d, d'$ and m in the support of M.

which implies that both, conditional on observed covariates X, no unobserved confounder exists that jointly affects the treatment selection, the potential mediator and the potential outcome and that, conditional on observed covariates and the treatment, no unobserved confounder exists that jointly affects the potential mediator and the potential outcome. This can be tested, e.g. see (Bijwaard and Jones, 2019; Imai et al., 2010a,b). Allowing for unobserved factors that influence the treatment, mediator and the outcome, i.e. violating the sequential ignorablility condition, can be dealt with using a structural model.

4 Inverse propensity weighting model using average income

We define the educational effect of moving up one education level in terms of a proportional change in the mortality hazard rate. Because we base our educational effect on (mixed) proportional hazard models of the mortality rate, it is natural also to define the mediator effects proportionally. When we have just one, time-invariant, mediator we can use the Bijwaard and Jones (2019) mediation method for (mixed) proportional hazard models. They assume (next to sequential ignorability) a proportional mediator effect. Then the direct effect, not running through income, of education can be identified using an inverse Propensity Weighting (IPW) method with weights:

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

with weight $W(d)$ for $\theta(d)$, for $d = 0, 1$.

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 $\theta(d)$.

Although we observe income on an annual basis we first assume (average) income is a time-invariant and use the average (log)income as a mediator of the education effect on mortality.

5 Inverse propensity weighting model

Here we explain how we can account for the evolution of income of the life course in estimating the direct and indirect effects of education on mortality.

Income is observed to change at fixed moments in time. We also assume that sequential ignorability holds for every income observation (the sequential mediators), conditional on observed individual characteristics and the income history up till the income we consider. We extend the Bijwaard and Jones (2019) method to (sequentially ordered) mediators, the income at fixed time points. An important difference with dynamic mediation analysis in the literature (e.g. Daniel et al. (2015)) is that our 'treatment', the education level, is time-invariant (fixed after completion). We therefore do not need to consider how the treatment changes over time with the mediator.

Before we turn to the analysis of (M)PH model we show the intuition using a linear model with 2 (temporally) ordered mediators. Define the potential values $M_1(d), M_2(d, m_1), Y(d, m_1, m_2), M_2(d, M_1(d')))$ and $Y\big(d, M_1(d'), M_2(d'', M_1(d'))\big)$. Thus the potential outcome depends on (suppressing the dependence on observed characteristics) the treatment $D = d$, the first potential mediator, $M_1(d')$ (which depend on the education level) and the second potential mediator (which depends on the education and the first mediator). The expected values for the potential output is:

$$\begin{aligned} \mathrm{E}\Big[Y\big(d, M_1(d'), M_2(d'', M_1(d'))\big)\Big] &= \int \mathrm{E}\Big[Y|D = d, M_1 = m_1, M_2 = m_2, X = x\Big] \\ &f_{M2|D,M_1,X}(m_2|D = d'', M_1 = m_1, X = x) \cdot f_{M1|D,X}(m_1|D = d', X = x) f_X(x)\, dm_2 dm_1' dm_1 dx \end{aligned} \tag{2}$$

where $f(\cdot|\cdot)$ is a conditional density function. All the densities can, using similar reasoning as Huber (2014) (and (Bijwaard and Jones, 2019)) and Bayes' rule be written in terms of propensity scores, $\Pr(D = d|\cdot, X = x)$:

$$\begin{aligned}\mathrm{E}\Big[Y\big(d, M_1(d'), M_2(d'', M_1(d'))\big)\Big] = \int \mathrm{E}\Big[Y|D = d, M_1 = m_1, M_2 = m_2, X = x\Big]\\ \frac{\Pr(D = d''|M_2, M_1 = m_1', X = x)}{\Pr(D = d''|M_1 = m_1', X = x)} \cdot \frac{\Pr(D = d'|M_1 = m_1, X = x)}{\Pr(D = d'|X = x)} f_X(x)\, dm_2 dm_1' dm_1 dx \quad (3)\end{aligned}$$

From (3) we can derive the weights needed to calculate the direct, and different indirect effects in a (M)PH hazard model. The weights for the total effect are

$$W_{tot} = \frac{D}{\Pr(D = 1|X)} + \frac{(1 - D)}{\Pr(D = 0|X)} \quad (4)$$

The weights for the direct effect for income of those with $D = 0$, $W(0)$, are derived from $Y\big(1, M_1(0), M_2(0, M_1(0))\big)$ and $Y\big(0, M_1(0), M_2(0, M_1(0))\big)$ and the weights for the direct effect for income of those with $D = 1$, $W(1)$, are derived from $Y\big(1, M_1(1), M_2(1, M_1(1))\big)$ and $Y\big(0, M_1(1), M_2(1, M_1(1))\big)$[1]

$$\begin{aligned}W_{Dir}(0) &= \frac{\Pr(D = 0|\overline{M}_2, X)}{\Pr(D = 0|X)} \left[\frac{Y \cdot D}{\Pr(D = 1|\overline{M}_2, X)} - \frac{Y \cdot (1 - D)}{\Pr(D = 0|\overline{M}_2, X)}\right] \quad (5)\\ W_{Dir}(1) &= \frac{1}{\Pr(D = 1|X)} \left[Y \cdot D - Y \cdot (1 - D)\frac{\Pr(D = 1|\overline{M}_2, X)}{\Pr(D = 0|\overline{M}_2, X)}\right] \quad (6)\end{aligned}$$

with $\overline{M}_2 = \{M_1, M_2\}$. The weights for the indirect effect through M_1 are derived from $Y\big(1, M_1(1), M_2(1, M_1(1))\big)$ and $Y\big(1, M_1(0), M_2(1, M_1(0))\big)$ or $Y\big(1, M_1(1), M_2(0, M_1(1))\big)$ and $Y\big(1, M_1(0), M_2(0, M_1(0))\big)$

$$\begin{aligned}W_{I,M_1}(0) &= \frac{Y \cdot (1 - D)}{\Pr(D = 0|\overline{M}_2, X)} \left[\frac{\Pr(D = 1|\overline{M}_2, X)}{\Pr(D = 1|X)} - \frac{\Pr(D = 1|\overline{M}_2, X)}{\Pr(D = 1|M_1, X)} \frac{\Pr(D = 0|M_1, X)}{\Pr(D = 0|X)}\right] \quad (7)\\ W_{I,M_1}(1) &= \frac{Y \cdot D)}{\Pr(D = 1|\overline{M}_2, X)} \left[\frac{\Pr(D = 0|\overline{M}_2, X)}{\Pr(D = 0|M_1, X)} \frac{\Pr(D = 0|M_1, X)}{\Pr(D = 0|X)} - \frac{\Pr(D = 0|\overline{M}_2, X)}{\Pr(D = 0|X)}\right] \quad (8)\end{aligned}$$

The weight for the indirect effect through M_2 is derived from $Y\big(1, M_1(1), M_2(1, M_1(0))\big)$ and $Y\big(1, M_1(1), M_2(0, M_1(0))\big)$

$$\begin{aligned}W_{I,M_2}(0) &= \frac{Y \cdot (1 - D)}{\Pr(D = 0|\overline{M}_2, X)} \left[\frac{\Pr(D = 1|\overline{M}_2, X)}{\Pr(D = 1|X)} \frac{\Pr(D = 0|M_1, X)}{\Pr(D = 0|X)} - \frac{\Pr(D = 0|\overline{M}_2, X)}{\Pr(D = 0|X)}\right] \quad (9)\\ W_{I,M_2}(1) &= \frac{Y \cdot D)}{\Pr(D = 1|\overline{M}_2, X)} \left[\frac{\Pr(D = 1|\overline{M}_2, X)}{\Pr(D = 1|X)} - \frac{\Pr(D = 1|\overline{M}_2, X)}{\Pr(D = 1|M_1, X)} \frac{\Pr(D = 0|M_1, X)}{\Pr(D = 0|X)}\right] \quad (10)\end{aligned}$$

However, we are only interested in the direct and sum of all indirect effects, which is the total effect minus the direct effect.

Extending this to K sequential mediators, assuming that (i) The potential income depends only on the educational attainment, the previous income and observed characteristics and (ii) The weights for

[1] In principle we could also define direct effects based on (i) $Y\big(1, M_1(0), M_2(1, M_1(0))\big)$ and $Y\big(0, M_1(0), M_2(1, M_1(0))\big)$ or (ii) $Y\big(1, M_1(1), M_2(0, M_1(1))\big)$ and $Y\big(0, M_1(1), M_2(0, M_1(1))\big)$. We think the others are more natural.

the direct effect (not running through any of the K income observations) are

$$W_{Dir}(0) = \frac{\Pr(D=0|\overline{M}_K, X)}{\Pr(D=0|X)} \left[\frac{Y \cdot D}{\Pr(D=1|\overline{M}_K, X)} - \frac{Y \cdot (1-D)}{\Pr(D=0|\overline{M}_K, X)} \right] \tag{11}$$

$$W_{Dir}(1) = \frac{1}{\Pr(D=1|X)} \left[Y \cdot D - Y \cdot (1-D) \frac{\Pr(D=1|\overline{M}_K, X)}{\Pr(D=0|\overline{M}_K, X)} \right] \tag{12}$$

with $\overline{M}_K = \{M_1, \ldots, M_K\}$. Thus for estimating the direct effects of education on mortality we only need to estimate the propensity scores conditional on the covariates and at each time t_k conditional on the covariates and the income history up till t_k.

6 Results

We assume a Gompertz proportional hazard model for the mortality. As controls we include variables that influence both the educational attainment and mortality: birth order, family size, birth year, birth month (to account for possible seasonal effects), father's education, mother's education and parental EGP. Table 2 reports the estimated effect on the mortality hazard of moving up one educational level. The first column provides the results from 'standard' (Gompertz) proportional hazard regression. Not adjusting for selective education choice, comparing the unadjusted estimates with the IPW estimates, overestimates the impact of education (especially for the lowest education group). We conclude from these analyses that men with only two years of secondary education would reduce their mortality rate with 36% (=$1 - e^{-0.464}$) if they had attained one more year of secondary education. For the other education levels the gain is lower but still substantial (22%-24%).

Table 2: Impact of education levels on the mortality rate, total effect

	Unadjusted	*IPW estimate*
Primary to	-0.360^{**}	-0.243^{**}
secondary (2 years)	(0.013)	(0.013)
Secondary (2 years) to	-0.464^{**}	-0.445^{**}
secondary (3 years)	(0.021)	(0.021)
Secondary (3 years) to	-0.293^{**}	-0.271^{**}
Post secondary	(0.028)	(0.028)
Post secondary to	-0.253^{**}	-0.242^{**}
higher	(0.026)	(0.030)

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

In Table 3 we present the decomposition of the effects of education on the mortality rate, running through the effect it has on income (η) and the direct effect (θ). We see a clear difference among the low educated, only primary education, the medium educated, two years of secondary education and the higher educated. For the low educated the effect of improving education on mortality is mainly a direct effect of education. For the medium educated half of the educational gain is running through an increase in income, while for the higher educated the direct effect of education is negative (increasing mortality) and the main effect of education improvement is running through and increase in income.

Table 3: Decomposition of the educational gradient on the mortality rate, into an effect running though income and running through other pathways

	other pathways		*income*	
	$\theta(1)$	$\theta(0)$	$\eta(0)$	$\eta(1)$
Primary to	-0.219^{**}	-0.217^{**}	-0.025	-0.026
secondary (2 years)	(0.013)	(0.013)	(0.018)	(0.018)
Secondary (2 years) to	-0.202^{**}	-0.288^{**}	-0.244^{**}	-0.157^{**}
secondary (3 years)	(0.020)	(0.020)	(0.029)	(0.029)
Secondary (3 years) to	0.157^{**}	0.059^{+}	-0.429^{**}	-0.330^{**}
Post secondary	(0.029)	(0.026)	(0.040)	(0.038)
Post secondary to	0.179^{**}	0.080^{**}	-0.421^{**}	-0.322^{**}
higher	(0.031)	(0.028)	(0.043)	(0.041)

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

References

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.

Baron, R. M. and Kenny, D. A. (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 Jones, A. M. (2019). An IPW estimator for mediation effects in hazard models: with an application to schooling, cognitive ability and mortality. *Empirical Economics*, 57(1):129–175.

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

Cutler, D. M., Lleras-Muney, A., and Vogl, T. (2011). Socioeconomic status and health: Dimensions and mechanisms. In Glied, S. and Smith, P. C., editors, *The Oxford Handbook of Health Economics*, pages 124–163. Oxford UP, Oxford.

Daniel, R. M., De Stavola, B. L., Cousens, S. N., and Vansteelandt, S. (2015). Causal mediation analysis with multiple mediators. *Biometrics*, 71:1–14.

Grossman, M. (2015). The relationship between health and schooling: What's new? *Nordic Journal of Health Economics*, 3(1):7–17.

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

Huisman, M., Kunst, A. E., Bopp, M., J.-K.Borgan, Borrell, C., Costa, G., Deboosere, P., Gadeyne, S., Glickman, M., Marinacci, C., Minder, C., Valkonen, T., and Mackenbach, J. P. (2005). Educational inequalities in cause-specific mortality in middle-aged and older men and women in eight western European populations. *The Lancet*, 365(9458):493–500.

Hummer, R. A. and Lariscy, J. T. (2011). Educational attainment and adult mortality. In Rogers, R. G. and Grimmins, E. M., editors, *International Handbook of Adult Mortality*, volume 2, pages 241–260. Springer, Dordrecht, the Netherlands.

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

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

Masters, R. K., Hummer, R. A., and Powers, D. A. (2012). Educational differences in US adults mortality: A cohort perspective. *American Sociological Review*, 77(4):548–572.

Rosenbaum, P. and Rubin, D. B. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika*, 70:41–50.

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

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