HEDG

HEALTH, ECONOMETRICS AND DATA GROUP

WP 19/05

Education and life-expectancy and how the relationship is mediated through changes in behaviour: a principal stratification approach for hazard rates

Govert E. Bijwaard and Andrew M. Jones

February 2019

http://www.york.ac.uk/economics/postgrad/herc/hedg/wps/

Education and life-expectancy and how the relationship is mediated through changes in behaviour: a principal stratification approach for hazard rates

Govert E. Bijwaard* NIDI

Andrew M. Jones† University of York

February 6, 2019

Abstract

We investigate the causal impact of education on life-expectancy using data for England and Wales from the Health and Lifestyles Survey and how that impact is mediated through changes in health behaviour (smoking, exercise, having breakfast). For identification of the educational gain in mortality we employ a Regression Discontinuity Design implied by an increase in the minimum school leaving age in 1947 (from 14 to 15) together with a principal stratification method for the mortality hazard rate. This method allows us to derive the direct and indirect (through one or more mediators) effect of education on the implied life-expectancy.

Basic maximum likelihood estimation of a standard Gompertz hazard model for the mortality rate suggests that staying in school beyond age 15 years significantly increases life-expectancy by more than 14 years, with large indirect effects running through smoking and exercise. In contrast, estimates from the principal strata method indicate that the educational gain is much smaller (and statistically insignificant) for those who were induced to remain in school beyond age 15. The direct effect of education is even negative for females (but statistically insignificant). Neither, do we find statistically significant indirect effects of education on mortality running through health behaviour.

JEL classification: C41, I14, I24.
Keywords: Regression Discontinuity Design; Education; Mortality; Principal Strata;

*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

A large literature has documented substantial associations between education and longevity and health behaviours. It is commonly assumed that this educational gradient in mortality has a causal interpretation. However, this assumption has been challenged in the literature. The association may partly be explained by confounding factors, observed or unobserved variables, that affect both educational attainment and mortality (Grossman, 2015). For example, education may encourage better health-related decisions such as refraining from smoking or doing exercise. However, education and these health behaviours may be endogenous and depend on uncontrolled confounders, such as cognitive ability or parental background, that affects both the education attained, the health behaviours and mortality (Mazumder, 2008, Clark and Royer, 2013, Fletcher, 2015, McCartney et al., 2013). Ignoring that such common traits exist may render the association spurious and therefore change the policy implications.

The most common approach to identify causal effects of education on health and mortality exploits changes in compulsory schooling policies, usually increases in the minimum age or the legally permitted grade to leave school, as instrumental variables for schooling attainment. All these studies use as identification strategy that the changes in law induced people born in different years (or states) to obtain different levels of schooling for reasons that are plausibly unrelated to factors that may influence their health and mortality. If it is assumed that the change in compulsory schooling law only affected health and mortality through its effect on education, one can estimate a causal effect of education on health for those who comply with the new law and would not do otherwise. Estimates based on these studies point towards a small effect (Mazumder, 2008, 2012, Jones et al., 2011, Van Kippersluis et al., 2011, Fletcher, 2015, Meghir et al., 2018, Basu et al., 2018) or no effect (Albouy and Lequien, 2009, Clark and Royer, 2013, Jürges et al., 2013). Regression discontinuity does not require exclusion restrictions and identification comes from the discontinuity in responses.

Only a few studies have attempted to identify the causal effect of education on mortality rates, using either an inverse propensity weighting method (Bijwaard et al., 2017, Bijwaard and Jones, 2019) or a structural modelling approach (Bijwaard et al., 2015a,b, 2019). However, a critical assumption in propensity score weighting is that of no selection on unobservables. This may be hard to defend. Although the structural models, in which interdependence between education, health, and cognitive ability is explicitly modelled, account for unobserved correlated factors they assume a particular structure. The compulsory schooling change provides a natural instrument to identify the causal effect of education on the mortality rate. Although some studies using compulsory schooling law have looked at the impact of education on mortality (Lleras-Muney, 2005, Clark and Royer, 2013, Van Kippersluis et al., 2011), none have accounted for the dynamic selection and censoring inherent in a duration outcome, such as age at death. We model the mortality hazard rate, the instantaneous probability that an individual dies at a certain age conditional on surviving up to that age. Accounting for right-censoring, when the individual is only known to have survived up until 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 the proportional hazard models, leads to biased inference.

The Mixed Proportional Hazard (MPH) model has been the main model for duration data in economics. However, no unambiguous solution to instrumental variable estimation of the inherently non-linear MPH-model with endogenous covariates has been found. Bijwaard (2009) developed a consistent estimator for the parameters of a semiparametric MPH model with an unspecified distribution of the unobserved heterogeneity and with an endogenous variable for which an instrument exists. In its simplest form, the estimator does not require nonparametric estimation of unknown densities. A

limitation of this method is that the baseline duration dependence is restricted to a piecewise constant function, which may be hard to implement for fast increasing hazard rates like the mortality hazard. Another limitation is that the method is computationally intensive, because it is based on finding the roots of a multidimensional step function which does not have a derivative. The instrumental variable (IV) based methods of Terza et al. (2008) for non-linear models have been used recently for duration models. However, Wan et al. (2015) and Wan et al. (2018) have shown that both their two-stage predictor substitution (2SPS) and the two-stage residual inclusion (2SRI) methods are biased in a Weibull proportional hazard framework, at least under the standard assumptions common in the treatment evaluation literature.

The method we use is based on the principal stratification framework. This framework (Frangakis and Rubin, 2002) is a general potential outcomes framework for causal inference with instruments and/or intermediate variables. Principal stratification has its seeds in the instrumental variable methods as described in Angrist et al. (1996), and Imbens and Rubin (1997), and it has been developed and formalized within the potential outcome approach to causal inference. The commonly applied framework developed by Angrist et al. (1996) to define the *Local Average Treatment Effect* (LATE) in a random experiment with non-compliance is a special case of the principal stratification framework. A principal stratum consists of individuals having the same joint potential outcomes, independent of the treatment assignment (Frangakis and Rubin, 1999, Zhang et al., 2009, Mealli and Mattei, 2012). Therefore comparisons of potential outcomes under different treatment levels within a principal stratum give well-defined causal effects. The principal strata are usually defined in term of four complier types: (*i*) *Always takers*: individuals who take the treatment irrespectively of their assigned treatment (*ii*) *Never takers* individuals who never take the treatment and (*iii*) *Compliers* individuals who only take the treatment if assigned to treatment. The final complier type *Defiers*: individuals who only take the treatment if not assigned to treatment, are ruled out using a monotonicity assumption. When assuming a parametric baseline mortality hazard rate, estimation of the latent complier types and their associated hazard rate is possible using maximum likelihood estimation of the implied mixture model. We assume a Gompertz proportional mortality rate, which assumes an exponential increase in the mortality rate by age. A Gompertz mortality rate is known to provide accurate mortality rates for middle aged individuals (Gavrilov and Gavrilova, 1991). Similar methods for duration outcomes, also based on principal stratification, have been developed by Cuzick et al. (2007), Lin et al. (2014) and Wan et al. (2015).

Surprisingly little research in economics has investigated the underlying causal mechanism of education on mortality in the presence of one or more intermediate variables. One exception is Bijwaard and Jones (2019) who have investigated whether intelligence, as measured by an IQ test, mediates the impact of education on the mortality rate. They use an inverse propensity method and, therefore, assume no unmeasured confounders. The principal stratification framework is very useful when causal estimands are defined in terms of intermediate outcomes, which are on the causal pathway between the treatment and the primary outcome (Mealli and Mattei, 2012). However, mediation analysis with an instrumental variable method is not straightforward. Four approaches exist in the literature to achieve identification in a mediation model. The first approach assumes that the treatment is (as good as) randomly assigned and that there exist no unobserved mediators. The second approach of Imai et al. (2010a,b) assumes that there are no unobserved counfounders and both the treatment and mediator are ignorable. These two approaches remove the need for an instrument and use matching or propensity score methods to obtain causal effects. A third approach by Frölich and Huber (2017) assumes the existence of two instruments, one for the treatment and one for the mediator. Finally, our approach is related to the approach of Dippel et al. (2017) and only requires one instrument. Dippel et al. (2017) postulate that identification of the total and mediated effects is achieved when we assume that confounders which influence the treatment will influence the outcome only through the mediator. This is equivalent to assuming that the instrument is also a valid instrument for the mediators, but conditional on the treatment.

The contribution of this paper is threefold. We study the causal impact of education on the mortality rate using a change in compulsory schooling. Our second contribution is that, we decompose this impact into a direct effect of education and an indirect effect running through health behaviours, such as smoking exercise and breakfast. Finally, the paper provides a methodological innovation in instrumental variable mediation analysis for hazard rate models, using the principal stratification approach to motivate estimation of a mixture model.

Our empirical application using a simple Gompertz mortality rate model with education and health behaviours included suggests that staying in school beyond age 15 years significantly increases life-expectancy by more than 14 years and with a large indirect effect of education running through health behaviours. This overestimates the educational gain of staying in school. Estimates from the principal strata method indicate that the educational gain is much smaller (and statistically insignificant) for those who were induced to remain in school. Both the direct effect of education and indirect effect of education running through health behaviours are statistically insignificant in the principle strata method.

2 Data

We use the British Health and Lifestyle Survey (HALS). This survey was conducted to collect data on health behaviours of the British population, including smoking, alcohol consumption and exercise. We use the first and second waves of the survey combined with the long-term follow-up of deaths. The first wave was conducted in 1984–1985, with a response rate of 73%. In total 9003 individuals (18–99 years old) were interviewed. In 1991–1992 a follow up survey was carried out for which 5352 individuals completed the interviews. Johnston et al. (2015) have used these data to investigate the causal link between education and health knowledge. We use the same measure of schooling, the age at which a respondent left secondary school, which ranges from 14 to 19 years old. Just as for Johnston et al. (2015) our identification strategy utilises educational reforms that increased the legal school leaving age in England and Wales from 14 to 15 (in contrast to Johnston et al. (2015) we only focus on the 1947 reform) and remove all individuals living in Scotland from the sample.

Figure 1 shows how the 1947-reform affects the school leaving age, the probability of leaving school before the age of 15 and the probability to survive until the end of the survey (July 1st, 2009), for all individuals and for males and females separately. The 1947-reform clearly had a large effect on school leaving around the age of 15. Survival seems to increase around the 1947-reform cut-off, but this might be caused by a period effect. We therefore include a linear trend in the days from the cut-off moment (1-1-1933), which may differ before and after the cut-off, in our analysis.[1]

Longitudinal follow-up of the date and cause of death is available up to July 2009 in the Seventh Death Revision of the HALS. We observe the respondents from their survey interview till July 1st, 2009 or till death, which allows us to construct the mortality hazards. Figure 2 depicts the Kaplan-Meier survival curves for individuals born 25 years before or after the cut-off birth of the 1947 reform. According to a log-rank test of survival difference the survival of individuals who left school before age 15 (1947-reform) differs significantly from the survival of individuals who stayed longer in school (also for males and females separately).

Figure 3 depicts the relation between education (leaving school before or after age 15) and the (possible) mediators, smoking, exercise, obesity, good sleep patterns , and having breakfast. It shows clearly that those who stay in school beyond age 15 have better health behaviours.

[1]We also include season of birth dummies, region of birth dummies, a non-white dummy, and when analysing all individuals, a gender dummy as control variables in our analyses.

Figure 1: Age left school and survival to 2009, 1947 reform
Dots represent average schooling and survival from survey entry to survey end by quarter of birth (year of birth for survival)

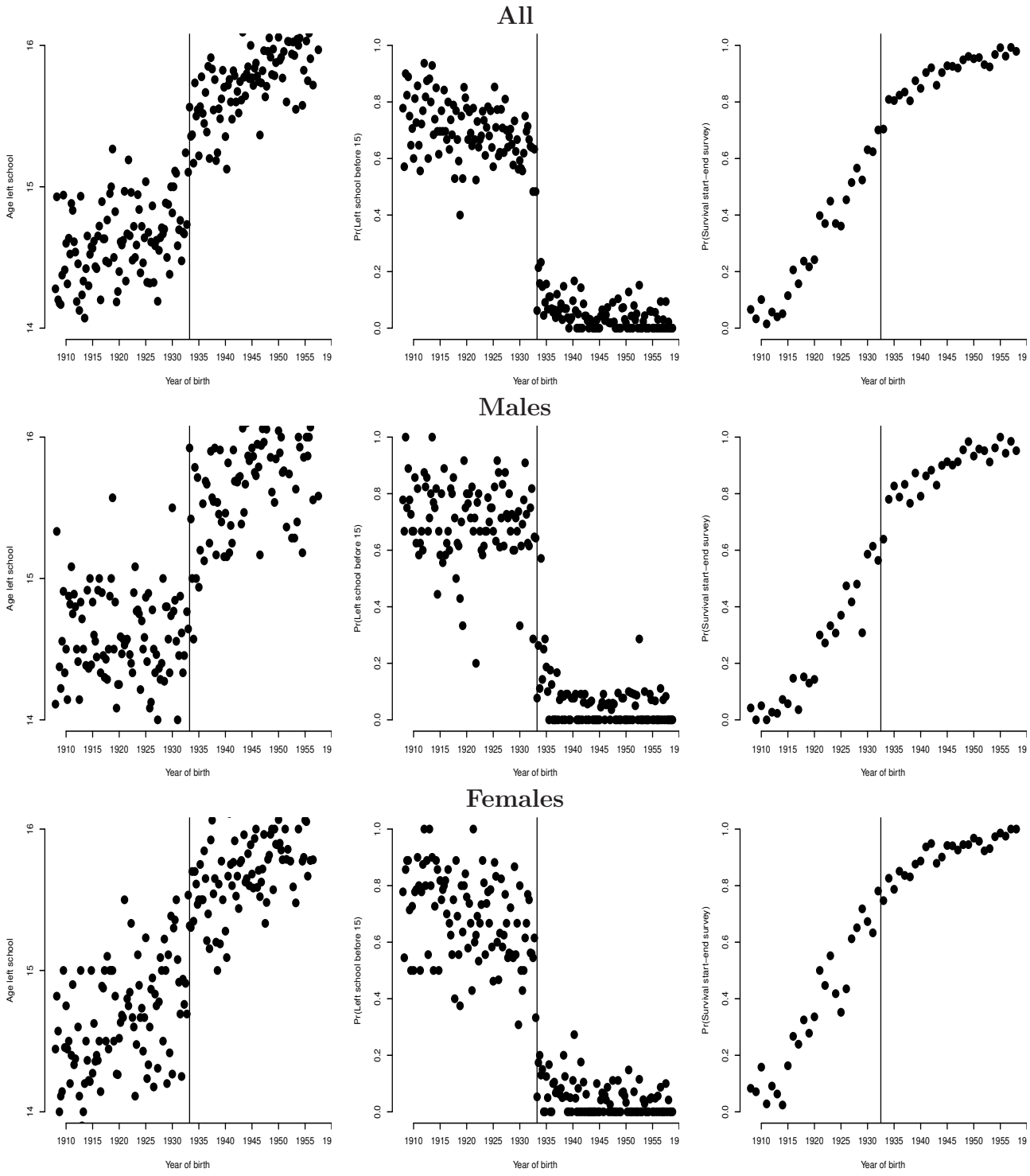


Figure 2: Survival from first survey till end by age left school, only individuals born 25 years before or after the cut-off birth of the 1947 reform

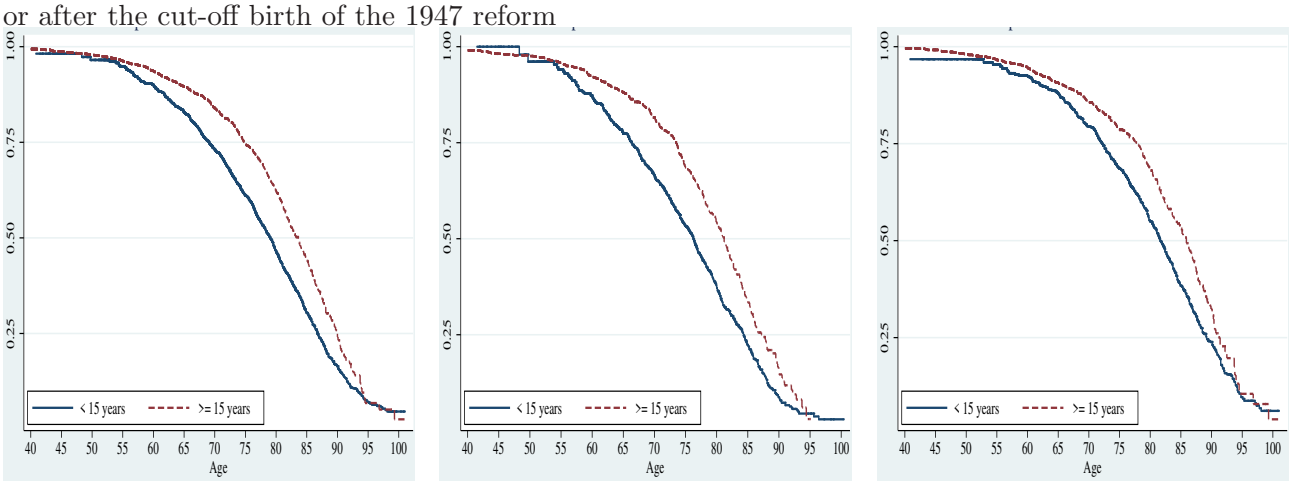
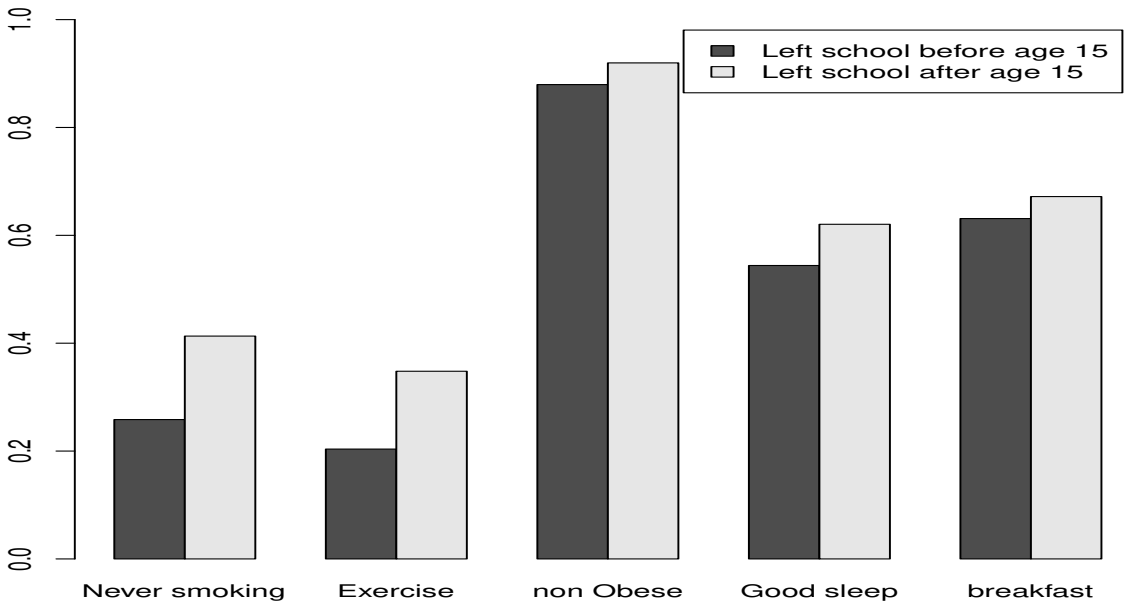


Figure 3: Mediator probability by education level (leaving school before or after age 15)



3 Methodology

We seek to find the impact of improving education, measured by staying longer in school (beyond age 15), on mortality. We are especially interested in disentangling this effect of education into a direct effect and an indirect effect (operating through health behaviours) on mortality. We assume a Gompertz proportional hazard mortality rate, which postulates that the (baseline) hazard increases exponentially with age, $\lambda(t|D,X)) = e^{\beta_0+\alpha t+\beta' X+\gamma D}$. We will use the (implied) life expectancy as the outcome of interest.[2] Assuming that the estimated Gompertz hazard holds, the life expectancy can be very well approximated by (Lenart, 2014)

$$\mathrm{LE}(D|\alpha,\beta,\gamma) = -\exp\left(\tfrac{1}{\alpha}\exp\left[\beta_0+\beta' X+\gamma_D D\right]\right) \times \left(\beta_0+\beta' X+\gamma_D D-\ln(\alpha)+0.5772\right)/\alpha \quad (1)$$

where 0.5772 is the Euler constant.

When we assume a Mixed Proportional Hazard (MPH) model with Gamma distributed unobserved heterogeneity (with unit mean and variance σ^2) the life expectancy can be approximated by (Missov, 2013)

$$\mathrm{LE}(D|\alpha,\beta,\gamma) = \Bigg[\left(1-\tfrac{\sigma^2}{\alpha}e^{\beta_0+\beta' X+\gamma_D D}\right)^{-\frac{1}{\sigma^2}} \times \left(\ln(\alpha)-\beta_0-\beta' X-\gamma_D D-\ln(\sigma^2)\right) - \sum_{j=1}^{\frac{1}{\sigma^2}-1} \tfrac{1}{j}\left(1-\tfrac{\sigma^2}{\alpha}e^{\beta_0+\beta' X+\gamma_D D}\right)^{j-\frac{1}{\sigma^2}}\Bigg]/\alpha \quad (2)$$

3.1 Standard mediation analysis

Traditionally, causal mediation analysis has been formulated within the framework of linear structural equation models (LSEM) (Baron and Kenny, 1986). The LSEM assumes the following two equations: Let Y be the outcome, mortality, D the 'treatment', higher education , M the mediator, e.g. smoking behaviour, and X other included covariates.

$$M = \beta_{M0}+\beta_M' X+\gamma_M D+\epsilon_M \quad (3)$$
$$Y = \beta_{Y0}+\beta_Y' X+\gamma_D D+\theta_M M+\epsilon_Y \quad (4)$$

where the error terms ϵ_M and ϵ_Y are assumed to be mutually independent, with variance σ_M^2 and σ_Y^2. Then, the direct effect of education, D, is γ_D and the indirect effect, running through the mediator, is $\theta_M \cdot \gamma_M$. Note that this simple multiplication does not hold for a non-linear model.

3.2 Principle strata hazard rate model

The identification of the causal effect of health on mortality is complicated by the potential endogeneity of education. The association between health and education may partly be explained by confounding factors such as cognitive ability and parental background, which affect both education choices and health (McCartney et al., 2013). For this reason the coefficient of education (and the mediators) in a standard (proportional) hazard models for the mortality rate, such as a Gompertz model, is likely to be upward biased. To address this endogeneity we use a fuzzy regression discontinuity design, as implied by the change in school leaving age of the 1947-reform in England and Wales.

A method using people born just around a threshold to identify causal effects is known as *Regression discontinuity design* (RDD), (Imbens and Lemieux, 2008, Lee and Lemieux, 2010). The basic idea

[2] In principle the direct and indirect effect can also be defined in terms of hazard (ratios). But these effects depend on age, t, and are therefore difficult to interpret.

behind RDD is that assignment to treatment (in our case continue schooling after age 15) is determined, either completely or partly, by the value of the instrument (the change in law) being on either side of a fixed threshold. Because people born before the reform could still stay in school beyond age 15 we have a fuzzy RDD. We use a local, separate for each side of the threshold, linear regression to account for distance from the threshold (Gelman and Imbens, 2019).

The fuzzy RDD we have can be viewed as an instrumental variable method, with being affected by the change in the law (i.e. born after 1-4-1933) used as instrument for staying longer in school (beyond age 15). Define for the policy change (treatment assignment) the (potential) binary variable $D(z)$, with $Z = 1$ if an individual was affected by the policy change and zero otherwise and $D = 1$ if the individual stayed in school to age 15 or beyond. We seek to investigate the role of mediation factors (health behaviours) on the impact of education on mortality. These mediators are possibly affected by education. We use the *principal strata* formulation of the problem (Frangakis and Rubin, 2002). This implies we have four (latent) complier types for education: *always takers* are individuals who always stay longer in school irrespectively of whether they were affected by the policy change; *never takers* are individuals who never stay in school beyond age 15 and, *Compliers* are individuals who only stay longer in school because they were induced to do so through the policy change. We assume that *Defiers*, individuals who only stay in school beyond age 15 because they were not induced through the policy change, do not exist. Our work is novel in that we consider inherently non-linear hazard models, instead of linear models. We assume that the (potential) hazard depends on the complier-type. We also assume that the value, i.e. the probability that individual exhibits 'good' health behaviour $\Pr(M = 1|c)$, and the impact of the mediator depends on the complier-type.

We assume the complier type influences only the scale of the mortality rate,γ_{d1} and γ_{d0} and through the mediators M_j, θ_{jd1} and θ_{jd0}. Thus, the potential hazard for an individual of complier type d is

$$\lambda^{(d)}(t|\cdot) \;=\; v\lambda_0(t;\alpha)\exp\Bigg(\gamma_{d1}Z \;+\; \gamma_{d0}(1 \;-\; Z) \;+\; \sum_{j=1}^{m} M_j\big[\theta_{jd1}Z \;+\; \theta_{jd0}(1 \;-\; Z)\big] \;+\; \beta' X\Bigg) \quad (5)$$

where $d = \{a(lways), n(ever), c(omplier)\}$ are the complier types. We either assume that $v \equiv 1$ (PH-model) or that v follows a unit mean Gamma distribution with variance σ^2 (MPH-model).

We impose some (standard, Imbens and Rubin (1997)) additional assumptions[3]:
Weak exclusion restriction: requires that the treatment assignment (and mediators) are unrelated to potential outcomes for never-takers and always-takers. This implies that the effect of always-takers and of never-takers is independent of treatment assignment (or education level for the mediator effects). Thus, ($\gamma_{d1} = \gamma_{d0} = \gamma_d$ and $\theta_{jd1} = \theta_{jd0} = \theta_{jd}$ for $d = \{a, n\}; j = 1, \ldots, m$. Note that this restriction is inherent to the RDD, the policy change only effects the outcome through the induced change in treatment (prolonging the time in school). Classical instrumental variable analysis requires an exclusion restriction (Angrist et al., 1996). Still, for compliers both the education effect and the mediator effect may differ by the instrument value Z. Implying that those who are induced to stay longer in school behave differently than those who were not induced.

For identification of the LATE (sometimes called the compliers average causal effects (CACE), (Mealli and Mattei, 2012)), we already made the monotonicity assumptions, ruling out defiers: *Monotonicity* Angrist et al. (1996): $D(1) \geq D(0)$.
Mediator assumption

$$Z \not\perp M|D \quad \text{and} \quad Z \perp \lambda^{(d)}(t|M)|D \quad (6)$$

Figure 4 provides a graphical illustration of the relationship between the instrument, education, the mediator(s) and the mortality rate using a directed acyclic graph, where each arrow represents a

[3]We also impose the stable unit treatment value assumption (SUTVA), which requires that the mortality of an individual is not affected by the instrument assigned to other individuals, the education attained or health behaviour of other individuals.

causal path (Pearl, 2000, 2012). It states that the instrument Z, the change in compulsory schooling law, influences the education choice D, which influences the mediator(s) M. Both education and the mediator(s) influence the mortality hazard λ. The model allows for three sources of endogenous effects: an unmeasured confounder U_D that causes D and M, an unmeasured confounder U_M that is caused by D and causes M and λ and, an unmeasured confounder U_λ that causes M and λ.

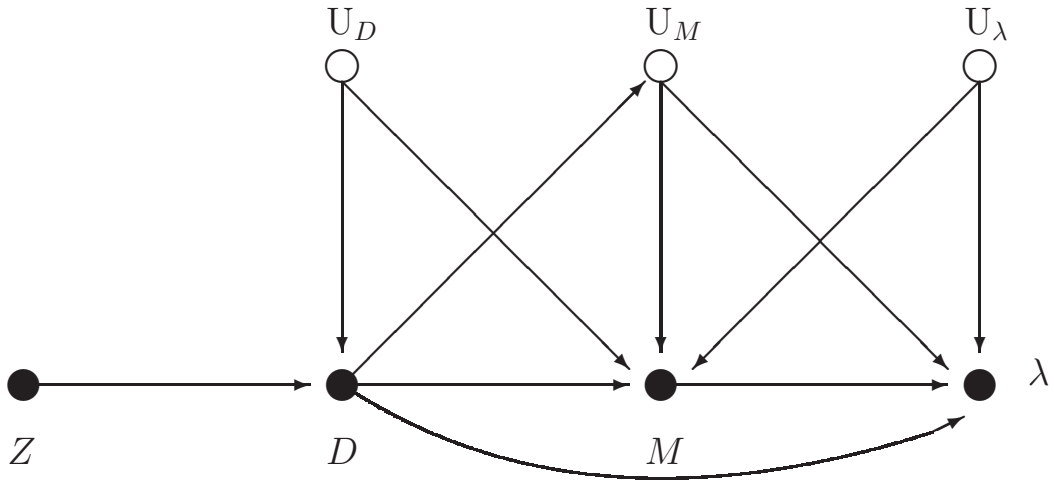


Figure 4: Mediation model with an instrument

3.3 Estimation of principle strata hazard rate model

Based on the assumption of a known functional form of the baseline hazard, e.g. Weibull ($\lambda_0(t) = \alpha t^{\alpha-1}$) or Gompertz ($\lambda_0(t) = e^{\alpha t}$), it is very easy to derive the likelihood function contribution of individual i with 3 mediators, see Appendix A for the full likelihood

$$L_i = \prod_{Z_i}\prod_{D_i}\prod_{M_{1i}}\prod_{M_{2i}}\prod_{M_{3i}}\Big[\lambda_i\big(t_i|Z_i, D_i, M_{1i}, M_{2i}, M_{3i}\big)^{\delta_i} S\big(t|Z_i, D_i, M_{1i}, M_{2i}, M_{3i}\big)\Big]^{I\left(Z_i, D_i, M_{1i}, M_{2i}, M_{3i}\right)} \quad (7)$$

with $S\big(t|Z_i, D_i, M_{1i}, M_{2i}, M_{3i}\big)$ is the survival rate at age t for an individual with $Z_i, D_i, M_{1i}, M_{2i}, M_{3i}$, e.g.,

$$S(t|Z=0, D=0, M_1=1, M_2=0, M_3=1) = p^n p^n_{m_1}(1-p^n_{m_2})p^n_{m_3}\exp\Big(-\int_0^t \lambda_0(s)e^{\beta' X+\theta_{1n}+\theta_{3n}}ds\Big) + p^c p^0_{m_1}(1-p^0_{m_2})p^0_{m_3}\exp\Big(-\int_0^t \lambda_0(s)e^{\beta' X+\gamma_0+\theta_{10}+\theta_{30}}ds\Big)$$

where p^n, p^c are the complier-type probabilities for never-takers and compliers and $p^n_{m_j} = \Pr(M_j = 1|\text{never taker})$ and $p^0_{m_j} = \Pr(M_j = 1|\text{complier}, Z = 0)$ for $j = 1, 2, 3$. All these probabilities are estimated jointly together with the other parameters of the model. This implies that our model is a latent class model with the complier types as latent classes.

Regression discontinuity designs identify a treatment effect locally around the threshold. Standard in the literature is a local continuity assumption, implying that persons close to threshold are comparable except for assignment variable. The standard approach to account for divergence is to include a local polynomial of the running variable, in our case the date of birth, estimated separately at each side of the threshold. Gelman and Imbens (2019) argue that a linear polynomial, and not higher order polynomials, is the best approach. We, therefore, include linear functions of the birthdate from April 1933, for each side of the threshold, in the Gompertz model.

In regression discontinuity design there is always a trade-off between bias (a small window) and variance (a large window). We choose a window of 10 years. In Section 5 we discuss the results for a window of 15 years and for smaller windows.

3.4 Causal quantities

We use the (implied) life expectancy as the outcome of interest. In causal mediation analysis, researchers are typically interested in decomposing the treatment effects on the outcome into two effects: the *indirect effect* that operates through the mediator and the *direct effect* that does not operate through the mediator. When using instrumental variables usually only the Local Average Treatment Effect (LATE) for the compliers is identified Angrist et al. (1996). First we define the local direct, $\zeta(d)$, and indirect, $\xi(d)$, effects with three mediators for $d = 0, 1$ as (using (1)

$$\zeta(d) = \Big[\mathrm{LE}\big(1, M_1(d), M_2(d), M_3(d)\big) - \mathrm{LE}\big(0, M_1(d), M_2(d), M_3(d)\big)|\mathit{complier}\Big] \tag{8}$$

$$\xi_1(d) = \Big[\mathrm{LE}\big(d, M_1(1), M_2(d), M_3(d)\big) - \mathrm{LE}\big(d, M_1(0), M_2(d), M_3(d)\big)|\mathit{complier}\Big] \tag{9}$$

$$\xi_2(d) = \Big[\mathrm{LE}\big(d, M_1(d), M_2(1), M_3(d)\big) - \mathrm{LE}\big(d, M_1(d), M_2(0), M_3(d)\big)|\mathit{complier}\Big] \tag{10}$$

$$\xi_3(d) = \Big[\mathrm{LE}\big(d, M_1(d), M_2(d), M_3(1)\big) - \mathrm{LE}\big(d, M_1(d), M_2(d), M_3(1)\big)|\mathit{complier}\Big] \tag{11}$$

The derivation of these causal quantities are given in Appendix B. Additionally we have the quantity of *local mediated interaction* for $d = 0, 1$

$$\begin{aligned} IM(d) = \Big[&\mathrm{LE}\big(d, M_1(1), M_2(1), M_3(1)\big) + 2\mathrm{LE}\big(d, M_1(0), M_2(0), M_3(0)\big) - \mathrm{LE}\big(d, M_1(1), M_2(0), M_3(0)\big) \\ &- \mathrm{LE}\big(d, M_1(0), M_2(1), M_3(0)\big) - \mathrm{LE}\big(d, M_1(0), M_2(0), M_3(1)\big)|\mathit{complier}\Big] \end{aligned} \tag{12}$$

The local average mediated interaction measures the interactive effect of both mediators, see Taguri et al. (2018).

The local average indirect effect of M_j, $\xi_j(d)$, represents the average change in the outcome among the compliers when mediator M_j was changed from the value that would have been realised under the treatment condition to the value that would have been realised under the control condition while the treatment remains at value d. The local average direct effect, $\zeta(d)$, represents the average change in the outcome among the compliers when the treatment was changed from the control condition to the treatment condition while the mediator remains at the value that would have been realised under the treatment condition d. Note that the the sum of local direct, the local indirect effects and the local mediated interaction equals the 'standard' LATE:

$$\begin{aligned} LATE &= \Big[\mathrm{LE}\big(1, M_1(1), M_2(1), M_3(1)\big) - \mathrm{LE}\big(0, M_1(0), M_2(0), M_3(0)\big)|\mathit{complier}\Big] \\ &= \zeta(0) + IM(1) + \sum_{j=1}^{3} \xi_j(1) = \zeta(1) + IM(0) + \sum_{j=1}^{3} \xi_j(0) \end{aligned} \tag{13}$$

4 Results

4.1 Basic Gompertz model

Before we report the results of the principle strata Gompertz model we discuss the results of using a standard Gompertz model (with or without gamma distributed unobserved heterogeneity) for the mortality rate when education (a dummy for staying in school beyond age 15) is one of the included variables. We also include three health behaviours that may be associated with education: smoking (never smoking), doing exercise and having a good breakfast and their interaction with education. All the estimates use the mortality data of individuals born in a bandwidth of 10 years around April 1933. Table C.1 in Appendix C provides the full parameter estimates. Including these three health behaviours in the hazard rate makes the effects of staying in school beyond age 15 statistically insignificant. The health behaviours all decrease the mortality hazard (all are statistically significant for the sample of males and females together and for the sample of females, never smoking and breakfast are significant for the sample of males). None of the interactions between staying in school beyond age 15 and the health behaviours are statistically significant.

Based on these estimated parameters we calculate the implied life-expectancy for each education and health behaviour combination and the implied total, direct and indirect effects (similar to equations (8) to (12)). Table 1 reports these estimated educational gains on the implied life-expectancy. We find a rather large total effect of education (staying in school beyond age 15) of fourteen (females) to nineteen (males) additional years. However, the direct effect of education is much smaller and not statistically significant.

The main impact of education seems to run through changes in health behaviour. Smoking explains six to eleven years of increasing life expectancy, exercise explain four and a half to eight years, and having breakfast explains (not statistically significant for males) three to five years. For all three groups (males and females together, and males or females separately) the MPH model with gamma distributed unobserved heterogeneity gives lower educational gains than the PH model.

Table 1: Estimated educational gain in life-expectancy at age 18 basic Gompertz model

	all		males		females	
	PH	MPH[b]	PH	MPH[b]	PH	MPH[b]
Total effect	17.787**	17.025**	21.276**	19.252**	14.704**	14.501**
	(2.641)	(2.294)	(4.528)	(3.503)	(3.104)	(2.894)
	direct effect					
Direct effect $D = 1$	−0.922	−2.579	2.294	−1.460	−3.444	−3.905
	(3.533)	(3.423)	(5.850)	(5.450)	(4.386)	(4.346)
Direct effect $D = 0$	2.609	3.183	3.593	5.649	1.894	1.705
	(1.884)	(1.971)	(3.107)	(3.149)	(2.336)	(2.530)
	indirect effect, $D = 1$					
never smoking	8.270**	7.940**	9.719**	8.364**	7.119**	7.707**
	(1.846)	(1.574)	(3.296)	(2.443)	(2.124)	(2.010)
exercise	6.636**	5.892**	8.623**	6.419**	4.755+	4.511+
	(1.897)	(1.599)	(3.174)	(1.682)	(2.261)	(2.044)
breakfast	4.508**	4.057**	3.197	1.964	5.492**	5.129**
	(1.639)	(1.478)	(2.701)	(2.167)	(2.020)	(1.951)
	indirect effect, $D = 0$					
never smoking	7.433**	7.637**	8.905**	10.645**	6.249**	5.659+
	(1.852)	(1.850)	(3.266)	(3.091)	(2.150)	(2.314)
exercise	7.032**	7.916**	6.209+	6.917+	7.339**	8.191**
	(1.985)	(1.984)	(3.027)	(3.012)	(2.619)	(2.691)
breakfast	4.236**	4.047+	3.856**	3.144	4.554+	4.0551+
	(1.409)	(1.595)	(2.162)	(2.546)	(1.843)	(2.080)
	mediation interaction effect					
IM(0)	0.008**	0.005+	0.011	0.006	0.006+	0.004
	(0.003)	(0.003)	(0.006)	(0.005)	(0.003)	(0.003)
IM(1)	−4.237**	−4.047+	−3.857	−3.144	−4.555+	−4.551+
	(1.411)	(1.596)	(2.169)	(2.549)	(1.845)	(2.081)

[a] never smoking, exercise and breakfast as mediation factors, linear local trend in month-year of birth (from 1-4-1933). Separate trend before and after the reform.

[b] Gamma unobserved heterogeneity. Regional, season of birth dummies, a non-white dummy, and a gender dummy (only in all) are also included. The bandwidth around 1-4-1933 included in the estimation is 10 years. The window of included births is twice the bandwidth. $^{+}p < 0.05$ $^{**}p < 0.01$

4.2 Principle strata Gompertz model

The standard Gompertz model presented above does not account for endogeneity of staying in school. The principle strata Gompertz model described in Section 3.2 that exploits the policy reform of 1947 as an instrument for staying in school seeks to solve this endogeneity issue. The full estimation results are given in Table D.1 in Appendix D here we only discuss the relevant coefficients for compliers: γ and θ for compliers with $Z = 0$ or $Z = 1$. None of the complier coefficients of education (γ_1 vs γ_0) or the mediator effects of the health behaviours differ significantly. The estimated mediator (health behaviour) probabilities by complier type are also important for the education gain and are reported in Table D.2 in Appendix D. The never-takers probability is very small (around 8%). In principle the data should not contain never-takers, as starting from 1947 everybody had to remain in school till the age 15. In Section 5 we test the robustness of our results when we either discard the people who leave school before age 15 after 1947 or assume that those people remain in school. About 60% of the people are compliers, which is the subpopulation of our (local) treatment effects. Although the mediator probabilities are always higher for the compliers with $Z = 1$ (induced to stay longer in school) than for the compliers with $Z = 0$ (no mandatory stay in school till the age of 15) they only differ significantly between these two compliers for having breakfast.

Based on the estimated coefficients of the model we calculate the life expectancy for the compliers and the implied education gain (using equations (8) to (12)). Table 2 presents these educational gains in life-expectancy. It is only for the sample of males with the PH model, that we find a statistically significant educational gain (LATE) on life expectancy, this is of about fifteen years. Using a MPH model with gamma distributed unobserved heterogeneity we do not find any significant educational gain. All the direct and indirect effects of education are also statistically insignificant when using the principle strata Gompertz model.

Table 2: Estimated educational gain in life-expectancy at age 18 principle strata Gompertz model

	all		males		females	
	PH	MPH[b]	PH	MPH[b]	PH	MPH[b]
LATE	7.591	4.660	15.013^{+}	11.317	−0.086	−1.208
	(4.730)	(4.101)	(7.168)	(6.814)	(5.699)	(5.815)
	direct effect					
dir(1)	−0.017	1.156	7.244	9.658	−5.972	−5.728
	(3.514)	(3.402)	(5.586)	(5.013)	(4.117)	(4.205)
dir(0)	−0.017	1.156	7.239	9.660	−5.971	−5.730
	(3.512)	(3.403)	(5.584)	(5.013)	(4.116)	(4.206)
	indirect effect, $D = 1$					
never smoking	4.690	2.561	2.437	−0.118	5.144	4.435
	(2.850)	(2.108)	(3.657)	(2.352)	(3.647)	(3.462)
exercise	1.982	1.022	3.186	0.963	0.983	0.833
	(1.606)	(1.374)	(3.012)	(2.365)	(1.819)	(1.769)
breakfast	0.936	−0.079	2.144	0.814	−0.241	−0.748
	(2.391)	(2.249)	(4.509)	(4.193)	(2.633)	(2.836)
	indirect effect, $D = 0$					
never smoking	4.690	2.561	2.440	−0.118	5.143	4.437
	(2.850)	(2.108)	(3.660)	(2.351)	(3.647)	(3.464)
exercise	1.982	1.021	3.189	0.962	0.983	0.833
	(1.606)	(1.373)	(3.014)	(2.364)	(1.819)	(1.770)
breakfast	0.936	−0.079	2.146	0.813	−0.241	−0.748
	(2.392)	(2.248)	(4.513)	(4.190)	(2.633)	(2.836)
	mediation interaction effect					
IM(0)	0.001	−0.000	0.017	−0.000	0.000	0.000
	(0.001)	(0.001)	(0.027)	(0.001)	(0.001)	(0.001)
IM(1)	−0.001	0.000	−0.013	0.000	−0.000	−0.000
	(0.001)	(0.001)	(0.021)	(0.001)	(0.001)	(0.001)

[a] never smoking, exercise and breakfast as mediation factors, linear local trend in month-year of birth (from 1-4-1933). Separate trend before and after the reform.

[b] Gamma unobserved heterogeneity. Regional, season of birth dummies, a non-white dummy, and a gender dummy (only in all) are also included. The bandwidth around 1-4-1933 included in the estimation is 10 years. The window of included births is twice the bandwidth. $^{+}p < 0.05$ $^{**}p < 0.01$

5 Robustness

In this section we check how robust our results are to a different choice of the 'third' mediator', to the choice of the bandwidth and to excluding never takers.

The reason we chose smoking, exercise and breakfast behaviour as potential mediators of education on mortality is because only these three behaviours showed a statistically significant relation with staying in school beyond age 15, see Table C.2 in Appendix C. We tested whether the estimated educational effects change when using another health behaviour as a third mediator (still using smoking and exercise behaviour as mediators). Table E.1 in Appendix E presents the estimated total, direct and indirect effects of education on the life-expectancy using sleeping-well, prudent alcohol use or not being obese as the third mediator. For males we find a statistically significant total effect of education when using a PH model for all three alternative mediators. However, the effect is much smaller and statistically insignificant when using a MPH model with gamma distributed unobserved heterogeneity.

A common issue in using fuzzy regression discontinuity designs is the choice of the bandwidth of whom to select around the threshold. There is always a trade-off between bias (a small window) and precision (a large window). We had chosen a window of 10 years around the threshold of born in April 1933. Table E.2 in Appendix E presents the educational gains in life-expectancy implied by re-estimating the complier model using different bandwidths, 15, 9, 8 or 7 years. Note that the MPH model with gamma unobserved heterogeneity did not converge in all cases. For all bandwidths smaller than 10 years all estimated education effects remain statistically insignificant, but close to the reported effects in Table 2. Choosing a larger bandwidth of 15 years would increase the estimated educational effects. The total educational gain (LATE) would be statistically significant for males and females together (eleven years for the MPH model) and for males only(twelve years). The effect of education running through smoking would be significant using a 15 year bandwidth for males and females together (seven years) and for females only (six years). The effect of education running through exercise would be significant for males and females together (two years) and for males only (four years). We consider, however, a bandwidth of 15 years around the policy change too large.

Another issue is that we identified a few, 8%, never-takers although in principle they should not exist. We, therefore, check the implication of either (1) removing those people born after April 1933 who leave school before age 15 from the analyses or (2) assuming that those people have stayed beyond age 15 in school. Again we re-estimate the principle strata Gompertz model and calculate the implied educational gains. Table E.3 in Appendix E presents these estimated educational gains with adjustment for never takers. Removing the never-takers from the analyses has little effect on the estimated educational gains. Only for males the total effect drops a little (and is not statistically significant for the PH model anymore). Assuming that the never-takers remain in fact beyond age 15 in school has a larger (but still rather small) impact on the estimated educational gains. All of the estimated educational gains (total, direct and indirect) for the MPH model with gamma distributed unobserved heterogeneity remain statistically insignificant.

6 Conclusion

We investigate the causal educational gain in life-expectancy and how it is mediated through difference in health behaviour (smoking, exercise, having breakfast), using data for England and Wales from the Health and Lifestyle Survey. For causal identification of the educational gain we employ a Regression Discontinuity Design implied by the increase in the minimum school leaving age in 1947 (from 14 to 15) together with a principal stratification method for the mortality hazard rate. The principal stratification framework is a general potential outcomes framework for causal inference with instruments and/or intermediate variables. It defines complier types (always takers, compliers and never takers) for both educational attainment, depending on the schooling reform, and the level of the mediators, depending on the education attainment. This method allows us to derive the direct and

indirect (through one or more mediators) effect of education on the implied life-expectancy.

Our empirical analysis using a simple Gompertz mortality rate model suggests that staying in school beyond age 15 years significantly increases life-expectancy by more than 14 years and with a large indirect effects of education running through health behaviours. This overestimates the educational gain of staying in school. Estimates from the principal strata method indicate that the educational gain is much smaller (and statistically insignificant) for those who were induced to remain in school beyond age 15. The direct effect of education is statistically insignificant and even negative for males and females together and for the model using only females. Neither do we find any statistically significant indirect effects of education running through changes in health behaviour.

We also conducted a few robustness test, allowing for a different choice of the 'third' mediator', smaller bandwidths around the threshold and ruling out never-takers.

The new principle strata hazard rate model we introduced here could be used for other applications with a duration outcome and a suitable instrument for the endogenous treatment and mediators. The method can also be adjusted for parametric baseline hazards other than the Gompertz.

References

Albouy, V. and Lequien, L. (2009). Does compulsory education lower mortality? *Journal of Health Economics*, 28(1):155–168.

Angrist, J. D., Imbens, G. W., and Rubin, D. B. (1996). Identification of causal effects using instrumental variables (with discussion). *Journal of the American Statistical Association*, 91:444–472.

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.

Basu, A., Jones, A. M., and Rosa Dias, P. (2018). Heterogeneity in the impact of type of schooling on adult health and lifestyle. *Journal of Health Economics*, 57:1–14.

Bijwaard, G. E. (2009). Instrumental variable estimation for duration data. In Engelhardt, H., Kohler, H.-P., and Fürnkranz-Prskawetz, A., editors, *Causal Analysis in Population Studies: Concepts, Methods, Applications*, pages 111–148. Springer–Verlag.

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*, 56:forthcoming.

Bijwaard, G. E., Myrskylä, M., and Tynelius, P. (2019). Education, cognitive ability and cause-specific mortality: A structural approach. *Population Studies*, 72:forthcoming.

Bijwaard, G. E., Myrskylä, M., Tynelius, P., and Rasmussen, F. (2017). Educational gains in cause-specific mortality: Accounting for cognitive ability and family-level confounders using propensity score weighting. *Social Science & Medicine*, 184:49–56.

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

Bijwaard, G. E., van Poppel, F., Ekamper, P., and Lumey, L. H. (2015b). Gains in life expectancy associated with higher education in men. *PLoS ONE*, 10:e0141200.

Clark, D. and Royer, H. (2013). The effect of education on adult mortality and health: Evidence from Britain. *American Economic Review*, 103(6):2087–2120.

Cuzick, J., Sasieni, P., Myles, J., and Tyrer, J. (2007). Estimating the effect of treatement in a proportional hazards model in the presence of non–compliance and contamination. *Journal of the Royal Statistical Society: Series B*, 69:565–588.

Dippel, C., Gold, R., Heblich, S., and Pinto, R. (2017). Instrumental variables and causal mechanisms: Unpacking the effect of trade on workers and voters. Technical Report 23209, National Bureau of Economic Research.

Fletcher, J. M. (2015). New evidence of the effects of education on health in the US: Compulsory schooling laws revisited. *Social Science & Medicine*, 127:101–107.

Frangakis, C. E. and Rubin, D. B. (1999). Addressing complications of intention-to-treat analysis in the combined presence of all-or-none treatment-noncompliance and subsequent missing outcomes. *Biometrika*, 86(2):365–379.

Frangakis, C. E. and Rubin, D. B. (2002). Principal stratification in causal inference. *Biometrics*, 58:21–29.

Frölich, M. and Huber, M. (2017). Direct and indirect treatment effects–causal chains and mediation analysis with instrumental variables. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(5):1645–1666.

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

Gelman, A. and Imbens, G. (2019). Why high-order polynomials should not be used in regression discontinuity designs. *Journal of Business & Economic Statistics*, 37:"forthcoming".

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

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.

Imbens, G. W. and Lemieux, T. (2008). Regression discontinuity designs: A guide to practice. *Journal of Econometrics*, 142:615–635.

Imbens, G. W. and Rubin, D. B. (1997). Bayesian inference for causal effects in randomized experiments with noncompliance. *The Annals of Statistics*, 25:305–327.

Johnston, D. W., Lordan, G., Shields, M. A., and Suziedelyte, A. (2015). Education and health knowledge: Evidence from UK compulsory schooling reform. *Social Science & Medicine*, 127:92–100.

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

Jürges, H., Kruk, E., and Reinhold, S. (2013). The effect of compulsory scholling on health: Evidence from biomarkers. *Journal of Population Economics*, 26:645–672.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48:281–355.

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.

Lin, H., Li, Y., Jiang, L., and Li, G. (2014). A semiparametric linear transformation model to estimate causal effects for survival data. *Canadian Journal of Statistics*, 42:18—-35.

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

Mazumder, B. (2008). Does education improve health: A reexamination of the evidence from compulsory schooling laws. *Federal Reserve Bank of Chicago Economic Perspectives*, 33(2):2–16.

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

McCartney, G., Collins, C., and Mackenzie, M. (2013). What (or who) causes health inequalities: Theories, evidence and implications? *Health Policy*, 113:221—-227.

Mealli, F. and Mattei, A. (2012). A refreshing account of principal stratification. *The International Journal of Biostatistics*, 8:1–19.

Meghir, C., Palme, M., and Simeonova, E. (2018). Education, cognition and health: Evidence from a social experiment. *American Economic Journal: Applied Economics*, 10(2):234–256.

Missov, T. I. (2013). Gamma-Gompertz life expectancy at birth. *Demographic Research*, 28:259–270.

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

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

Taguri, M., Featherstone, J., and Cheng, J. (2018). Causal mediation analysis with multiple causally non–ordered mediators. *Statistical Methods in Medical Research*, 27(1):3–19.

Terza, J., Basu, A., and Rathouz, P. (2008). Two-stage instrumental variable methods: Addressing endogeneity in health econometric modeling. *Journal of Health Economics*, 27:531–543.

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

Van Kippersluis, H., O'Donnell, O., and van Doorslaer, E. (2011). Long run returns to education: Does schooling lead to an extended old age? *Journal of Human Resources*, 46(4):695–721.

Wan, F., Small, D., Bekelman, J. E., and Mitra, N. (2015). Bias in estimating the causal hazard ratio when using two-stage instrumental variable methods. *Statistics in Medicine*, 34:2235–2265.

Wan, F., Small, D., and Mitra, N. (2018). A general approach to evaluating the bias of 2-stage instrumental variable estimators. *Statistics in Medicine*, 37:1997–2015.

Zhang, J. L., Rubin, D. B., and Mealli, F. (2009). Likelihood-based analysis of causal effects of job-training programs using principal stratification. *Journal of the American Statistical Association*, 104(485):166–176.

Appendix A Likelihood

Based on the assumption of a Gompertz[4] functional form of the baseline hazard and three mediators it is very easy to derive the likelihood function contribution of individual i.

$$L_i = \prod_{Z_i}\prod_{D_i}\prod_{M_{1i}}\prod_{M_{2i}}\prod_{M_{3i}}\Big[\lambda_i\big(t_i|Z_i,D_i,M_{1i},M_{2i},M_{3i}\big)^{\delta_i}\times S\big(t|Z_i,D_i,M_{1i},M_{2i},M_{3i}\big)\Big]^{I\left(Z_i,D_i,M_{1i},M_{2i},M_{3i}\right)}$$

with $S\big(t|Z_i,D_i,M_{1i},M_{2i},M_{3i}\big)$ is the survival rate at age t for an individual with $Z_i,D_i,M_{1i},M_{2i},M_{3i}$, with

$$\begin{aligned}S(t|Z=0,D=0,M_1,M_2,M_3) = {} & p^n(1-p^n_{m_1})^{1-M_1}(p^n_{m_1})^{M_1}(1-p^n_{m_2})^{1-M_2}(p^n_{m_2})^{M_2}(1-p^n_{m_3})^{1-M_3}(p^n_{m_3})^{M_3}\\ & \times\exp\Big(-\int_0^t e^{\alpha_n s+\gamma_n+\beta' X+\sum_j\theta_{jn}M_j}ds\Big)+\\ & p^c(1-p^0_{m_1})^{1-M_1}(p^0_{m_1})^{M_1}(1-p^0_{m_2})^{1-M_2}(p^0_{m_2})^{M_2}(1-p^0_{m_3})^{1-M_3}(p^0_{m_3})^{M_3}\\ & \times\exp\Big(-\int_0^t e^{\alpha_0 s+\gamma_0+\beta' X+\sum_j\theta_{j0}M_j}ds\Big)\end{aligned}\quad(A.1)$$

$$\begin{aligned}S(t|Z=1,D=0,M_1,M_2,M_3) = {} & p^n(1-p^n_{m_1})^{1-M_1}(p^n_{m_1})^{M_1}(1-p^n_{m_2})^{1-M_2}(p^n_{m_2})^{M_2}(1-p^n_{m_3})^{1-M_3}(p^n_{m_3})^{M_3}\\ & \times\exp\Big(-\int_0^t e^{\alpha_n s+\gamma_n+\beta' X+\sum_j\theta_{jn}M_j}ds\Big)\end{aligned}\quad(A.2)$$

$$\begin{aligned}S(t|Z=0,D=1,M_1,M_2,M_3) = {} & p^a(1-p^a_{m_1})^{1-M_1}(p^a_{m_1})^{M_1}(1-p^a_{m_2})^{1-M_2}(p^a_{m_2})^{M_2}(1-p^a_{m_3})^{1-M_3}(p^a_{m_3})^{M_3}\\ & \times\exp\Big(-\int_0^t e^{\alpha_a s+\gamma_a+\beta' X+\sum_j\theta_{ja}M_j}ds\Big)\end{aligned}\quad(A.3)$$

$$\begin{aligned}S(t|Z=1,D=1,M_1,M_2,M_3) = {} & p^a(1-p^a_{m_1})^{1-M_1}(p^a_{m_1})^{M_1}(1-p^a_{m_2})^{1-M_2}(p^a_{m_2})^{M_2}(1-p^a_{m_3})^{1-M_3}(p^a_{m_3})^{M_3}\\ & \times\exp\Big(-\int_0^t e^{\alpha_n s+\gamma_n+\beta' X+\sum_j\theta_{ja}M_j}ds\Big)+\\ & p^c(1-p^1_{m_1})^{1-M_1}(p^1_{m_1})^{M_1}(1-p^1_{m_2})^{1-M_2}(p^1_{m_2})^{M_2}(1-p^1_{m_3})^{1-M_3}(p^1_{m_3})^{M_3}\\ & \times\exp\Big(-\int_0^t e^{\alpha_1 s+\gamma_1+\beta' X+\sum_j\theta_{j1}M_j}ds\Big)\end{aligned}\quad(A.4)$$

and $p^f = \Pr(\textit{complier} = f)$, $f = \{a,c,n\}$ and $p^n_{m_j} = \Pr(M_j = 1|\textit{complier type} = n), p^a_{m_j} = \Pr(M_j = 1|\textit{complier type} = a), p^1_{m_j} = \Pr(M_j = 1|\textit{complier type} = c, Z = 1)$ and $p^0_{m_j} = \Pr(M_j = 1|\textit{complier type} = c, Z = 0)$.[5] The hazard rate $\lambda(t|Z = z, D = d, M_1, M_2, M_3) = -\partial\log S(t|Z = z, D = d, M_1, M_2, M_3)/\partial t$.

[4] It is straightforward to derive the likelihood for other known baseline hazards.

[5] Note that it is possible to allow the probabilities to depend on observed exogenous variables, X_c, which may be different from X.

Appendix B Derivation of causal quantities

Given:

$$f\Big(D(1), M_1\big(D(1)\big), M_2\big(D(1)\big), M_3\big(D(1)\big)\Big) - f\Big(D(0), M_1\big(D(0)\big), M_2\big(D(0), M_3\big(D(1)\big)\big)\Big) =$$
$$\Big[f\big(1, M_1(1), M_2(1), M_3(1)\big) \cdot D(1) + f\big(0, M_1(0), M_2(0), M_3(0)\big) \cdot \big(1 - D(1)\big)\Big]$$
$$- \Big[f\big(1, M_1(1), M_2(1), M_3(1)\big) \cdot D(0) + f\big(0, M_1(0), M_2(0), M_3(0)\big) \cdot \big(1 - D(0)\big)\Big]$$
$$= \Big[f\big(1, M_1(1), M_2(1), M_3(1)\big) - f\big(0, M_1(0), M_2(0), M_3(0)\big)\Big] \cdot \Big[D(1) - D(0)\Big]$$

This implies that:

$$\mathrm{E}\Big[f\Big(D(1), M_1\big(D(1)\big), M_2\big(D(1), M_3\big(D(1)\big)\Big) - f\Big(D(0), M_1\big(D(0)\big), M_2\big(D(0), M_2\big(D(0)\big)\Big)\Big] =$$
$$= \mathrm{E}\Big[f\big(1, M_1(1), M_2(1), M_3(1)\big) - f\big(0, M_1(0), M_2(0), M_3(0)\big)\Big| D(1) - D(0) = 1\Big] \cdot \Pr\Big[D(1) - D(0) = 1\Big]$$

and the Local Average Treatment Effect is given by:

$$\mathrm{LATE} = \mathrm{E}\Big[f\big(1, M_1(1), M_2(1), M_3(1)\big) - f\big(0, M_1(0), M_2(0), M_3(0)\big)\Big| \mathit{complier}\Big] \tag{B.1}$$

Given:

$$f\Big(D(1), M_1\big(D(z)\big), M_2\big(D(z)\big), M_3\big(D(z)\big)\Big) - f\Big(D(0), M_1\big(D(z)\big), M_2\big(D(z)\big), M_3\big(D(z)\big)\Big) =$$
$$\Big[f\big(1, M_1\big(D(z)\big), M_2\big(D(z)\big) M_3\big(D(z)\big)\big) \cdot D(1) + f\big(0, M_1\big(D(z)\big), M_2\big(D(z)\big) M_3\big(D(z)\big)\big) \cdot \big(1 - D(1)\big)\Big]$$
$$- \Big[f\big(1, M_1\big(D(z)\big), M_2\big(D(z)\big) M_3\big(D(z)\big)\big) \cdot D(0) + f\big(0, M_1\big(D(z)\big), M_2\big(D(z)\big) M_3\big(D(z)\big)\big) \cdot \big(1 - D(0)\big)\Big]$$
$$= \Big[f\big(1, M_1\big(D(z)\big), M_2\big(D(z)\big) M_3\big(D(z)\big)\big) - f\big(0, M_1\big(D(z)\big), M_2\big(D(z)\big) M_3\big(D(z)\big)\big)\Big] \cdot \Big[D(1) - D(0)\Big]$$

we derive the local direct effect:

$$\zeta(d) = \mathrm{E}\Big[f\big(1, M_1(d), M_2(d), M_3(d)\big) - f\big(0, M_1(d), M_2(d), M_3(d)\big)\Big| \mathit{complier}\Big] \tag{B.2}$$

Given:

$$f\Big(D(z), M_1\big(D(1)\big), M_2\big(D(z)\big), M_3\big(D(z)\big)\Big) - f\Big(D(z), M_1\big(D(0)\big), M_2\big(D(z)\big), M_3\big(D(z)\big)\Big) =$$
$$\Big[f\big(D(z), M_1(1), M_2(D(z)), M_3(D(z))\big) \cdot D(1) + f\big(D(z), M_1(0), M_2(D(z)), M_3(D(z))\big) \cdot \big(1 - D(1)\big)\Big]$$
$$- \Big[f\big(D(z), M_1(1), M_2(D(z)), M_3(D(z))\big) \cdot D(0) + f\big(D(z), M_1(0), M_2(D(z)), M_3(D(z))\big) \cdot \big(1 - D(0)\big)\Big]$$
$$= \Big[f\big(D(z), M_1(1), M_2(D(z)), M_3(D(z))\big) - f\big(D(z), M_1(0), M_2(D(z)), M_3(D(z))\big)\Big] \cdot \Big[D(1) - D(0)\Big]$$

the local indirect effect of M_1 is:

$$\xi_1(d) = \mathrm{E}\Big[f(d, M_1(1), M_2(d), M_3(d)) - f(d, M_1(0), M_2(d), M_3(d)\Big| \mathit{complier}\Big] \tag{B.3}$$

and from

$$f\Big(D(z), M_1(D(z)), M_2\big(D(1), M_3(D(z))\big)\Big) - f\Big(D(z), M_1(D(z)), M_2\big(D(0)\big), M_3(D(z))\Big) =$$
$$\Big[f\big(D(z), M_1(D(z)), M_2(1), M_3(D(z))\big) \cdot D(1) + f\big(D(z), M_1(D(z)), M_2(0), M_3(D(z))\big) \cdot \big(1 - D(1)\big)\Big]$$
$$- \Big[f\big(D(z), M_1(D(z)), M_2(1), M_3(D(z))\big) \cdot D(0) + f\big(D(z), M_1(D(z)), M_2(0), M_3(D(z))\big) \cdot \big(1 - D(0)\big)\Big]$$
$$= \Big[f\big(D(z), M_1(D(z)), M_2(1), M_3(D(z))\big) - f\big(D(z), M_1(D(z)), M_2(0), M_3(D(z))\big)\Big] \cdot \Big[D(1) - D(0)\Big]$$

the local indirect effect of M_2 is:

$$\xi_2(d) = \mathrm{E}\Big[f(d, M_1(d), M_2(1), M_3(d)) - f(d, M_1(d), M_2(0), M_3(d)\Big| \mathit{complier}\Big] \tag{B.4}$$

Finally, given:

$$f\Big(D(z), M_1(D(z)), M_2\big(D(z), M_3(D(1))\big)\Big) - f\Big(D(z), M_1(D(z)), M_2\big(D(z)\big), M_3(D(0))\Big) =$$
$$\Big[f\big(D(z), M_1(D(z)), M_2(D(z)), M_3(D(1))\big)\cdot D(1) + f\big(D(z), M_1(D(z)), M_2(D(z)), M_3(D(0))\big)\cdot\big(1-D(1)\big)\Big]$$
$$-\Big[f\big(D(z), M_1(D(z)), M_2(D(z)), M_3(D(1))\big)\cdot D(0) + f\big(D(z), M_1(D(z)), M_2(D(z)), M_3(D(0))\big)\cdot\big(1-D(0)\big)\Big]$$
$$= \Big[f\big(D(z), M_1(D(z)), M_2(D(z)), M_3(D(1))\big) - f\big(D(z), M_1(D(z)), M_2(D(z)), M_3(D(0))\big)\Big]\cdot\Big[D(1)-D(0)\Big]$$

the local indirect effect of M_3 is:

$$\xi_3(d) = \mathrm{E}\Big[f(d, M_1(d), M_2(d), M_3(1)) - f(d, M_1(d), M_2(d), M_3(1)\Big| \mathit{complier}\Big] \tag{B.5}$$

This results in the following estimation equations :

$$\begin{aligned}
\zeta(1) \;=\;& p_1^1 p_2^1 p_3^1\Big[\mathrm{LE}^{(c)}(1,1,1,1|1) - \mathrm{LE}^{(c)}(0,1,1,1|1)\Big] \\
&+ (1-p_1^1) p_2^1 p_3^1\Big[\mathrm{LE}^{(c)}(1,0,1,1|1) - \mathrm{LE}^{(c)}(0,0,1,1|1)\Big] \\
&+ p_1^1 (1-p_2^1) p_3^1\Big[\mathrm{LE}^{(c)}(1,1,0,1|1) - \mathrm{LE}^{(c)}(0,1,0,1|1)\Big] \\
&+ p_1^1 p_2^1 (1-p_3^1)\Big[\mathrm{LE}^{(c)}(1,1,1,0|1) - \mathrm{LE}^{(c)}(0,1,1,0|1)\Big] \\
&+ (1-p_1^1)(1-p_2^1) p_3^1\Big[\mathrm{LE}^{(c)}(1,0,0,1|1) - \mathrm{LE}^{(c)}(0,0,0,1|1)\Big] \\
&+ (1-p_1^1) p_2^1 (1-p_3^1)\Big[\mathrm{LE}^{(c)}(1,0,1,0|1) - \mathrm{LE}^{(c)}(0,0,1,0|1)\Big] \\
&+ p_1^1 (1-p_2^1)(1-p_3^1)\Big[\mathrm{LE}^{(c)}(1,1,0,0|1) - \mathrm{LE}^{(c)}(0,1,0,0|1)\Big] \\
&+ (1-p_1^1)(1-p_2^1)(1-p_3^1)\Big[\mathrm{LE}^{(c)}(1,0,0,0|1) - \mathrm{LE}^{(c)}(0,0,0,0|1)\Big]
\end{aligned} \tag{B.6}$$

$$\begin{aligned}
\zeta(0) \;=\;& p_1^0 p_2^0 p_3^0\Big[\mathrm{LE}^{(c)}(1,1,1,1|0) - \mathrm{LE}^{(c)}(0,1,1,1|0)\Big] \\
&+ (1-p_1^0) p_2^0 p_3^0\Big[\mathrm{LE}^{(c)}(1,0,1,1|0) - \mathrm{LE}^{(c)}(0,0,1,1|0)\Big] \\
&+ p_1^0 (1-p_2^0) p_3^0\Big[\mathrm{LE}^{(c)}(1,1,0,1|0) - \mathrm{LE}^{(c)}(0,1,0,1|0)\Big] \\
&+ p_1^0 p_2^0 (1-p_3^0)\Big[\mathrm{LE}^{(c)}(1,1,1,0|0) - \mathrm{LE}^{(c)}(0,1,1,0|0)\Big] \\
&+ (1-p_1^0)(1-p_2^0) p_3^0\Big[\mathrm{LE}^{(c)}(1,0,0,1|0) - \mathrm{LE}^{(c)}(0,0,0,1|0)\Big] \\
&+ (1-p_1^0) p_2^0 (1-p_3^0)\Big[\mathrm{LE}^{(c)}(1,0,1,0|0) - \mathrm{LE}^{(c)}(0,0,1,0|0)\Big] \\
&+ p_1^0 (1-p_2^0)(1-p_3^0)\Big[\mathrm{LE}^{(c)}(1,1,0,0|0) - \mathrm{LE}^{(c)}(0,1,0,0|0)\Big] \\
&+ (1-p_1^0)(1-p_2^0)(1-p_3^0)\Big[\mathrm{LE}^{(c)}(1,0,0,0|0) - \mathrm{LE}^{(c)}(0,0,0,0|0)\Big]
\end{aligned} \tag{B.7}$$

$$
\begin{aligned}
\xi_1(1) &= p_2^1 p_3^1\Big[p_1^1\mathrm{LE}^{(c)}(1,1,1,1|1) - p_1^0\mathrm{LE}^{(c)}(1,1,1,1|0) + (p_1^0 - p_1^1)\mathrm{LE}^{(c)}(1,0,1,1|1)\Big] \qquad \text{(B.8)}\\
&\quad + (1-p_2^1)p_3^1\Big[p_1^1\mathrm{LE}^{(c)}(1,1,0,1|1) - p_1^0\mathrm{LE}^{(c)}(1,1,0,1|0) + (p_1^0 - p_1^1)\mathrm{LE}^{(c)}(1,0,0,1|1)\Big]\\
&\quad + p_2^1(1-p_3^1)\Big[p_1^1\mathrm{LE}^{(c)}(1,1,1,0|1) - p_1^0\mathrm{LE}^{(c)}(1,1,1,0|0) + (p_1^0 - p_1^1)\mathrm{LE}^{(c)}(1,0,1,0|1)\Big]\\
&\quad + (1-p_2^1)(1-p_3^1)\Big[p_1^1\mathrm{LE}^{(c)}(1,1,0,0|1) - p_1^0\mathrm{LE}^{(c)}(1,1,0,0|0) + (p_1^0 - p_1^1)\mathrm{LE}^{(c)}(1,0,0,0|1)\Big]\\
\xi_1(0) &= p_2^0 p_3^0\Big[p_1^1\mathrm{LE}^{(c)}(1,1,1,1|1) - p_1^0\mathrm{LE}^{(c)}(1,1,1,1|0) + (p_1^0 - p_1^1)\mathrm{LE}^{(c)}(0,0,1,1|0)\Big] \qquad \text{(B.9)}\\
&\quad + (1-p_2^0)p_3^0\Big[p_1^1\mathrm{LE}^{(c)}(1,1,0,1|1) - p_1^0\mathrm{LE}^{(c)}(1,1,0,1|0) + (p_1^0 - p_1^1)\mathrm{LE}^{(c)}(0,0,0,1|0)\Big]\\
&\quad + p_2^0(1-p_3^0)\Big[p_1^1\mathrm{LE}^{(c)}(1,1,1,0|1) - p_1^0\mathrm{LE}^{(c)}(1,1,1,0|0) + (p_1^0 - p_1^1)\mathrm{LE}^{(c)}(0,0,1,0|0)\Big]\\
&\quad + (1-p_2^0)(1-p_3^0)\Big[p_1^1\mathrm{LE}^{(c)}(1,1,0,0|1) - p_1^0\mathrm{LE}^{(c)}(1,1,0,0|0) + (p_1^0 - p_1^1)\mathrm{LE}^{(c)}(0,0,0,0|0)\Big]\\
\xi_2(1) &= p_1^1 p_3^1\Big[p_2^1\mathrm{LE}^{(c)}(1,1,1,1|1) - p_2^0\mathrm{LE}^{(c)}(1,1,1,1|0) + (p_2^0 - p_2^1)\mathrm{LE}^{(c)}(1,1,0,1|1)\Big] \qquad \text{(B.10)}\\
&\quad + (1-p_1^1)p_3^1\Big[p_2^1\mathrm{LE}^{(c)}(1,0,1,1|1) - p_2^0\mathrm{LE}^{(c)}(1,0,1,1|0) + (p_2^0 - p_2^1)\mathrm{LE}^{(c)}(1,0,0,1|1)\Big]\\
&\quad + p_1^1(1-p_3^1)\Big[p_2^1\mathrm{LE}^{(c)}(1,1,1,0|1) - p_2^0\mathrm{LE}^{(c)}(1,1,1,0|0) + (p_2^0 - p_2^1)\mathrm{LE}^{(c)}(1,1,0,0|1)\Big]\\
&\quad + (1-p_1^1)(1-p_3^1)\Big[p_2^1\mathrm{LE}^{(c)}(1,0,1,0|1) - p_2^0\mathrm{LE}^{(c)}(1,0,1,0|0) + (p_2^0 - p_2^1)\mathrm{LE}^{(c)}(1,0,0,0|1)\Big]\\
\xi_2(0) &= p_1^0 p_3^0\Big[p_2^1\mathrm{LE}^{(c)}(1,1,1,1|1) - p_2^0\mathrm{LE}^{(c)}(1,1,1,1|0) + (p_2^0 - p_2^1)\mathrm{LE}^{(c)}(0,1,0,1|0)\Big] \qquad \text{(B.11)}\\
&\quad + (1-p_1^0)p_3^0\Big[p_2^1\mathrm{LE}^{(c)}(1,0,1,1|1) - p_2^0\mathrm{LE}^{(c)}(1,0,1,1|0) + (p_2^0 - p_2^1)\mathrm{LE}^{(c)}(0,0,0,1|0)\Big]\\
&\quad + p_1^0(1-p_3^0)\Big[p_2^1\mathrm{LE}^{(c)}(1,1,1,0|1) - p_2^0\mathrm{LE}^{(c)}(1,1,1,0|0) + (p_2^0 - p_2^1)\mathrm{LE}^{(c)}(0,1,0,0|0)\Big]\\
&\quad + (1-p_1^0)(1-p_3^0)\Big[p_2^1\mathrm{LE}^{(c)}(1,0,1,0|1) - p_2^0\mathrm{LE}^{(c)}(1,0,1,0|0) + (p_2^0 - p_2^1)\mathrm{LE}^{(c)}(0,0,0,0|0)\Big]\\
\xi_3(1) &= p_1^1 p_2^1\Big[p_3^1\mathrm{LE}^{(c)}(1,1,1,1|1) - p_3^0\mathrm{LE}^{(c)}(1,1,1,1|0) + (p_3^0 - p_3^1)\mathrm{LE}^{(c)}(1,1,1,0|1)\Big] \qquad \text{(B.12)}\\
&\quad + (1-p_1^1)p_2^1\Big[p_3^1\mathrm{LE}^{(c)}(1,0,1,1|1) - p_3^0\mathrm{LE}^{(c)}(1,0,1,1|0) + (p_3^0 - p_3^1)\mathrm{LE}^{(c)}(1,0,1,0|1)\Big]\\
&\quad + p_1^1(1-p_2^1)\Big[p_3^1\mathrm{LE}^{(c)}(1,1,0,1|1) - p_3^0\mathrm{LE}^{(c)}(1,1,0,1|0) + (p_3^0 - p_3^1)\mathrm{LE}^{(c)}(1,1,0,0|1)\Big]\\
&\quad + (1-p_1^1)(1-p_2^1)\Big[p_3^1\mathrm{LE}^{(c)}(1,0,0,1|1) - p_3^0\mathrm{LE}^{(c)}(1,0,0,1|0) + (p_3^0 - p_3^1)\mathrm{LE}^{(c)}(1,0,0,0|1)\Big]\\
\xi_3(0) &= p_1^0 p_2^0\Big[p_3^1\mathrm{LE}^{(c)}(1,1,1,1|1) - p_3^0\mathrm{LE}^{(c)}(1,1,1,1|0) + (p_3^0 - p_3^1)\mathrm{LE}^{(c)}(0,1,1,0|0)\Big] \qquad \text{(B.13)}\\
&\quad + (1-p_1^0)p_2^0\Big[p_3^1\mathrm{LE}^{(c)}(1,0,1,1|1) - p_3^0\mathrm{LE}^{(c)}(1,0,1,1|0) + (p_3^0 - p_3^1)\mathrm{LE}^{(c)}(0,0,1,0|0)\Big]\\
&\quad + p_1^0(1-p_2^0)\Big[p_3^1\mathrm{LE}^{(c)}(1,1,0,1|1) - p_3^0\mathrm{LE}^{(c)}(1,1,0,1|0) + (p_3^0 - p_3^1)\mathrm{LE}^{(c)}(0,1,0,0|0)\Big]\\
&\quad + (1-p_1^0)(1-p_2^0)\Big[p_3^1\mathrm{LE}^{(c)}(1,0,0,1|1) - p_3^0\mathrm{LE}^{(c)}(1,0,0,1|0) + (p_3^0 - p_3^1)\mathrm{LE}^{(c)}(0,0,0,0|0)\Big]
\end{aligned}
$$

with $\mathrm{LE}^{(c)}(d, m_1, m_2, m_3|z)$ is the expected life-expectancy given $D(z) = d$ and $M_1 = m_1, M_2 = m_2$ and $M_3 = m_3$ for a complier and $p_m^z = \Pr(M_m = 1|\text{complier}, Z = z),\ z = 0, 1\ m = 1, 2, 3$. The local

average treatment effect is estimated by:

$$\begin{aligned}
\text{LATE} \;=\; & p_1^1 p_2^1 p_3^1 \text{LE}^{(c)}(1,1,1,1|1) - p_1^0 p_2^0 p_3^0 \text{LE}^{(c)}(1,1,1,1|0) \\
& + (1-p_1^1) p_2^1 p_3^1 \text{LE}^{(c)}(1,0,1,1|1) - (1-p_1^0) p_2^0 p_3^0 \text{LE}^{(c)}(1,0,1,1|0) \\
& + p_1^1 (1-p_2^1) p_3^1 \text{LE}^{(c)}(1,1,0,1|1) - p_1^0 (1-p_2^0) p_3^0 \text{LE}^{(c)}(1,1,0,1|0) \\
& + p_1^1 p_2^1 (1-p_3^1) \text{LE}^{(c)}(1,1,1,0|1) - p_1^0 p_2^0 (1-p_3^0) \text{LE}^{(c)}(1,1,1,0|0) \\
& + (1-p_1^1)(1-p_2^1) p_3^1 \text{LE}^{(c)}(1,0,0,1|1) - (1-p_1^0)(1-p_2^0) p_3^0 \text{LE}^{(c)}(1,0,0,1|0) \\
& + (1-p_1^1) p_2^1 (1-p_3^1) \text{LE}^{(c)}(1,0,1,0|1) - (1-p_1^0) p_2^0 (1-p_3^0) \text{LE}^{(c)}(1,0,1,0|0) \\
& + p_1^1 (1-p_2^1)(1-p_3^1) \text{LE}^{(c)}(1,1,0,0|1) - p_1^0 (1-p_2^0)(1-p_3^0) \text{LE}^{(c)}(1,1,0,0|0) \\
& + (1-p_1^1)(1-p_2^1)(1-p_3^1) \text{LE}^{(c)}(1,0,0,0|1) - (1-p_1^0)(1-p_2^0)(1-p_3^0 \text{LE}^{(c)}(1,0,0,0|0)
\end{aligned} \tag{B.14}$$

and the local mediated interactions can be derived from (13) and the equations above.

Appendix C Additional tables: Simple Gompertz model

Table C.1: Estimated (mixed) proportional Gompertz mortality rate

	all		males		females	
	PH	MPH[b]	PH	MPH[b]	PH	MPH[b]
schoolage 15	−0.196	−0.331	−0.238	−0.650	−0.163	−0.184
	(0.141)	(0.209)	(0.205)	(0.381)	(0.201)	(0.275)
Never smoking	−0.558**	−0.795**	−0.590**	−1.225**	−0.538**	−0.612+
	(0.134)	(0.200)	(0.207)	(0.396)	(0.179)	(0.243)
exercise	−0.528**	−0.824**	−0.411+	−0.796+	−0.632**	−0.886**
	(0.145)	(0.222)	(0.196)	(0.359)	(0.218)	(0.336)
breakfast	−0.318**	−0.421+	−0.256	−0.362	−0.392+	−0.492+
	(0.104)	(0.167)	(0.141)	(0.293)	(0.156)	(0.231)
schoolage 15 ×	−0.063	−0.031	−0.054	0.263	−0.075	−0.221
Never smoking	(0.186)	(0.243)	(0.292)	(0.446)	(0.250)	(0.338)
schoolage 15 ×	0.030	0.211	−0.160	0.058	0.222	0.398
exercise	(0.199)	(0.263)	(0.279)	(0.430)	(0.289)	(0.378)
schoolage 15 ×	−0.020	−0.001	0.044	0.136	−0.081	−0.062
breakfast	(0.159)	(0.222)	(0.227)	(0.382)	(0.227)	(0.300)
τ_1[c]	−0.018	−0.040	−0.017	−0.065	−0.015	−0.032
	(0.014)	(0.022)	(0.020)	(0.040)	(0.021)	(0.030)
τ_2[c]	−0.054**	−0.038	−0.063+	−0.023	−0.043	−0.035
	(0.021)	(0.025)	(0.029)	(0.041)	(0.030)	(0.035)
constant	−8.714**	−10.353**	−7.738**	−10.450**	−9.500**	−10.667**
	(0.403)	(0.724)	(0.543)	(1.143)	(0.597)	(1.073)
α	0.075**	0.104**	0.066**	0.115**	0.086**	0.108**
	(0.006)	(0.012)	(0.008)	(0.019)	(0.008)	(0.018)
σ^2		1.044**		1.536**		0.970
		(0.364)		(0.531)		(0.719)

[b] Gamma unobserved heterogeneity.
[c] Linear local trend in month-year of birth (from 1-4-1933). Separate trend before and after the reform. The bandwidth around 1-4-1933 included in the estimation is 10 years. The window of included births is twice the bandwidth. $^{+}p < 0.05$ $^{**}p < 0.01$

Table C.1: Estimated (mixed) proportional Gompertz mortality rate (continued)

	all		males		females	
	PH	MPH[b]	PH	MPH[b]	PH	MPH[b]
wales	0.275	0.292	0.288	0.488	0.244	0.214
	(0.169)	(0.234)	(0.240)	(0.402)	(0.241)	(0.313)
north	0.448**	0.732**	0.599**	1.260**	0.310	0.453
	(0.152)	(0.233)	(0.212)	(0.429)	(0.221)	(0.303)
north west	0.264+	0.351	0.274	0.572	0.274	0.260
	(0.130)	(0.182)	(0.183)	(0.325)	(0.188)	(0.241)
yorks/humber	0.005	−0.035	0.067	0.093	−0.048	−0.109
	(0.164)	(0.216)	(0.235)	(0.365)	(0.232)	(0.288)
west midlands	0.340+	0.463+	0.331	0.688	0.338	0.358
	(0.150)	(0.212)	(0.213)	(0.356)	(0.213)	(0.282)
east midlands	0.230	0.264	0.255	0.454	0.215	0.146
	(0.156)	(0.213)	(0.212)	(0.345)	(0.235)	(0.299)
east anglia	−0.143	−0.237	−0.094	−0.159	−0.244	−0.363
	(0.217)	(0.284)	(0.290)	(0.452)	(0.331)	(0.409)
south west	0.155	0.166	0.295	0.530	−0.037	−0.128
	(0.163)	(0.219)	(0.217)	(0.365)	(0.253)	(0.314)
greater london	0.131	0.278	−0.005	0.072	0.222	0.477
	(0.146)	(0.208)	(0.205)	(0.315)	(0.211)	(0.348)
nonwhite	−0.506	−0.693	−0.354	−0.671	−0.839	−0.940
	(0.296)	(0.361)	(0.344)	(0.468)	(0.589)	(0.668)
winter	−0.023	−0.034	−0.122	−0.100	0.087	0.041
	(0.105)	(0.147)	(0.150)	(0.255)	(0.150)	(0.199)
spring	−0.147	−0.168	−0.279	−0.419	−0.028	−0.038
	(0.106)	(0.147)	(0.151)	(0.255)	(0.151)	(0.196)
autumn	−0.208	−0.165	−0.335+	−0.232	−0.065	−0.112
	(0.111)	(0.154)	(0.153)	(0.259)	(0.166)	(0.213)
male	0.321**	0.460**				
	(0.077)	(0.112)				

[b] Gamma unobserved heterogeneity. The bandwidth around 1-4-1933 included in the estimation is 10 years. The window of included births is twice the bandwidth. $^{+}p < 0.05$ $^{**}p < 0.01$

Table C.2: Linear regression of staying in school beyond age 15 on mediator

	whole sample	Males only	Females only
Never smoking	0.150^{**}	0.129^{**}	0.168^{**}
	(0.025)	(0.035)	(0.035)
Exercise	0.101^{**}	0.110^{**}	0.099^{**}
	(0.024)	(0.036)	(0.033)
Breakfast	0.116^{**}	0.141^{**}	0.096^{**}
	(0.025)	(0.038)	(0.034)
Sleeping well	0.050	0.080^{+}	0.029
	(0.026)	(0.040)	(0.035)
prudent alcohol	0.023	0.036	0.016
	(0.016)	(0.034)	(0.011)
Not Obese	0.052^{**}	0.021	0.072^{**}
	(0.016)	(0.021)	(0.023)

[a] Linear local trend in month-year of birth (from 1-4-1933). Separate trend before and after the reform. The bandwidth around 1-4-1933 included in the estimation is 10 years. The window of included births is twice the bandwidth. Regional, season of birth, non-white dummies, and a gender dummy (only in whole sample) are also included. $^{+}p < 0.05$ $^{**}p < 0.01$.

Appendix D Parameters principle strata Gompertz model

Table D.1: Estimated coefficients education and mediators in principle strata Gompertz mortality rate

	all		males		females	
	PH	MPH[a]	PH	MPH[a]	PH	MPH[a]
	education					
γ_a	-9.025^{**}	-10.650^{**}	-7.984^{**}	-11.285^{**}	-10.181^{**}	-10.743^{**}
	(0.443)	(0.827)	(0.599)	(1.375)	(0.687)	(1.470)
γ_n	-8.548^{**}	-10.105^{**}	-8.085^{**}	-11.417^{**}	-9.267^{**}	-9.774^{**}
	(0.481)	(0.833)	(0.743)	(1.460)	(0.665)	(1.358)
γ_0	-8.768^{**}	-10.225^{**}	-7.697^{**}	-10.409^{**}	-9.933^{**}	-10.491^{**}
	(0.420)	(0.755)	(0.555)	(1.158)	(0.661)	(1.455)
γ_1	-8.767^{**}	-10.342^{**}	-8.189^{**}	-11.568^{**}	-9.405^{**}	-9.927^{**}
	(0.444)	(0.800)	(0.637)	(1.382)	(0.641)	(1.364)
	never smoking					
θ_a	-0.404^{+}	-0.597^{+}	-0.617^{+}	-1.044^{+}	-0.221	-0.313
	(0.183)	(0.246)	(0.294)	(0.474)	(0.246)	(0.348)
θ_n	-0.238	-0.318	-0.348	-0.709	0.074	0.087
	(0.540)	(0.648)	(1.156)	(1.244)	(0.578)	(0.664)
θ_0	-0.625^{**}	-0.876^{**}	-0.605^{+}	-1.356^{**}	-0.663^{**}	-0.712^{+}
	(0.178)	(0.252)	(0.287)	(0.518)	(0.247)	(0.298)
θ_1	-1.311^{**}	-1.263^{**}	-0.889	-0.853	-1.601^{+}	-1.598^{+}
	(0.484)	(0.443)	(0.680)	(0.702)	(0.663)	(0.626)
	no exercise					
θ_a	-0.378	-0.483	-0.467	-0.771	-0.233	-0.243
	(0.201)	(0.258)	(0.313)	(0.562)	(0.267)	(0.294)
θ_n	-0.847	-1.003	-0.824	-1.123	-0.810	-0.934
	(0.617)	(0.678)	(1.117)	(1.320)	(0.696)	(0.776)
θ_0	-0.459^{+}	-0.741^{**}	-0.348	-0.754	-0.607^{+}	-0.712
	(0.190)	(0.276)	(0.251)	(0.494)	(0.309)	(0.412)
θ_1	-0.774^{+}	-0.818^{+}	-0.819	-0.727	-0.785	-0.858
	(0.339)	(0.368)	(0.568)	(0.738)	(0.435)	(0.485)
	breakfast					
θ_a	-0.306	-0.401	-0.224	-0.214	-0.397	-0.470
	(0.194)	(0.272)	(0.288)	(0.514)	(0.270)	(0.351)
θ_n	-0.572	-0.547	-0.828	-0.746	-0.317	-0.333
	(0.443)	(0.527)	(0.796)	(0.988)	(0.482)	(0.550)
θ_0	-0.262	-0.379	-0.145	-0.264	-0.331	-0.379
	(0.138)	(0.207)	(0.190)	(0.404)	(0.214)	(0.267)
θ_1	-0.366	-0.350	-0.297	-0.311	-0.400	-0.360
	(0.262)	(0.309)	(0.399)	(0.601)	(0.362)	(0.408)
α	0.075^{**}	0.101^{**}	0.068^{**}	0.120^{**}	0.088^{**}	0.098^{**}
	(0.006)	(0.012)	(0.008)	(0.020)	(0.009)	(0.025)
σ^2		0.963^{**}		1.268^{**}		0.654
		(0.197)		(0.227)		(0.748)

[a] Gamma unobserved heterogeneity. The estimated coefficients of the control variables are very similar to those reported in Table C.1 and are available upon request. $^{+}p < 0.05$ $^{**}p < 0.01$

Table D.2: Estimated complier and mediator probability in principle strata Gompertz model with never smoking, exercise and breakfast as mediation factors

	all		males		females	
	PH	MPH[a]	PH	MPH[a]	PH	MPH[a]
	Complier probability education					
always-takers	0.344**	0.345**	0.311**	0.310**	0.375**	0.375**
	(0.015)	(0.015)	(0.021)	(0.021)	(0.021)	(0.021)
never-takers	0.078**	0.078**	0.086**	0.086**	0.074**	0.074**
	(0.008)	(0.008)	(0.013)	(0.013)	(0.010)	(0.010)
compliers	0.577**	0.577**	0.603**	0.604**	0.551**	0.551**
	(0.017)	(0.017)	(0.025)	(0.025)	(0.023)	(0.023)
	probability never smoking					
always-takers	0.485**	0.487**	0.380**	0.381**	0.559**	0.559**
	(0.026)	(0.025)	(0.039)	(0.039)	(0.033)	(0.033)
never-takers	0.259**	0.259**	0.239**	0.241**	0.273**	0.273**
	(0.046)	(0.046)	(0.067)	(0.068)	(0.064)	(0.064)
compliers, $Z = 0$	0.295**	0.295**	0.222**	0.223**	0.363**	0.363**
	(0.041)	(0.041)	(0.057)	(0.057)	(0.058)	(0.058)
compliers, $Z = 1$	0.410**	0.409**	0.337**	0.338**	0.434**	0.434**
	(0.053)	(0.054)	(0.081)	(0.081)	(0.072)	(0.072)
	probability exercise					
always-takers	0.358**	0.358**	0.360**	0.359**	0.359**	0.359**
	(0.024)	(0.024)	(0.038)	(0.039)	(0.031)	(0.031)
never-takers	0.236**	0.235**	0.230**	0.231**	0.232**	0.232**
	(0.034)	(0.034)	(0.065)	(0.064)	(0.059)	(0.059)
compliers, $Z = 0$	0.205**	0.205**	0.169**	0.171**	0.237**	0.237**
	(0.036)	(0.036)	(0.049)	(0.049)	(0.052)	(0.052)
compliers, $Z = 1$	0.314**	0.312**	0.336**	0.336**	0.294**	0.294**
	(0.048)	(0.048)	(0.074)	(0.074)	(0.065)	(0.065)
	probability breakfast					
always-takers	0.797**	0.798**	0.807**	0.807**	0.791**	0.791**
	(0.021)	(0.021)	(0.031)	(0.032)	(0.027)	(0.027)
never-takers	0.579**	0.579**	0.655**	0.653**	0.514**	0.514**
	(0.053)	(0.053)	(0.074)	(0.074)	(0.073)	(0.073)
compliers, $Z = 0$	0.581**	0.580**	0.487**	0.486**	0.667**	0.667**
	(0.044)	(0.044)	(0.065)	(0.065)	(0.058)	(0.058)
compliers, $Z = 1$	0.607**	0.606**	0.728**	0.727**	0.498**	0.498**
	(0.048)	(0.048)	(0.063)	(0.063)	(0.070)	(0.070)

[a] Gamma unobserved heterogeneity. $^{+}p < 0.05$ $^{**}p < 0.01$

Appendix E Additional tables: Principle strata Gompertz robustness

Table E.1: Estimated educational gain in life-expectancy at age 18 principle strata Gompertz model, using different third mediator

third mediator[a]	all		males		females	
	PH	MPH[b]	PH	MPH[b]	PH	MPH[b]
	LATE					
sleeping well	11.252	5.741	15.252^{+}	11.124	3.632	−4.497
	(6.403)	(4.613)	(6.925)	(5.704)	(8.078)	(4.854)
prudent alcohol	7.398	4.138	14.177^{+}	11.161	0.907	−1.847
	(4.681)	(3.912)	(6.539)	(6.099)	(6.335)	(5.890)
not obese	9.276	4.904	16.259^{+}	11.808	1.274	−2.775
	(4.974)	(4.028)	(7.021)	(6.044)	(6.115)	(5.430)
	direct effect, dir(1)					
sleeping well	4.827	3.194	12.480	11.951	−6.471	-13.715^{+}
	(5.631)	(5.157)	(7.388)	(6.345)	(7.706)	(6.413)
prudent alcohol	3.268	3.898	10.364	10.952	−20.785	−16.772
	(4.763)	(4.620)	(6.260)	(5.725)	(18.977)	(20.095)
not obese	3.320	2.364	11.298	5.498	−6.789	−5.281
	(7.717)	(7.705)	(14.030)	(9.586)	(8.171)	(7.570)
	direct effect,dir(0)					
sleeping well	4.826	3.195	12.478	11.950	−6.469	-13.717^{+}
	(5.630)	(5.157)	(7.390)	(6.344)	(7.703)	(6.414)
prudent alcohol	3.267	3.897	10.361	10.953	−20.763	−16.787
	(4.763)	(4.618)	(6.260)	(5.724)	(18.934)	(20.148)
not obese	3.319	2.365	11.293	5.499	−6.787	−5.281
	(7.716)	(7.705)	(14.0.6)	(9.587)	(8.167)	(7.571)

[a] Never smoking and exercise included as first two mediation factors.
[b] Gamma unobserved heterogeneity. $^{+}p < 0.05$ $^{**}p < 0.01$

Table E.1: continued

bandwidth[a]	all		males		females	
	PH	MPH[b]	PH	MPH[b]	PH	MPH[b]
	indirect effect Never smoking, $D = 1$					
sleeping well	8.145	4.236	3.539	0.587	9.001	6.208
	(4.282)	(2.498)	(3.900)	(2.441)	(5.385)	(2.871)
prudent alcohol	5.310	2.927	1.494	−0.554	6.779	5.114
	(2.973)	(2.133)	(3.347)	(2.282)	(4.174)	(3.322)
not obese	6.284	3.426	3.094	0.361	7.116	4.907
	(3.119)	(2.112)	(3.666)	(2.254)	(4.051)	(2.163)
	indirect effect Never smoking, $D = 0$					
sleeping well	8.146	4.237	3.536	0.588	9.003	6.207
	(4.283)	(2.499)	(3.897)	(2.442)	(5.385)	(2.954)
prudent alcohol	5.311	2.928	1.492	−0.554	6.780	5.112
	(2.973)	(2.133)	(3.344)	(2.283)	(4.174)	(3.320)
not obese	6.285	3.426	3.090	0.362	7.114	4.906
	(3.120)	(2.112)	(3.663)	(2.255)	(4.051)	(2.953)
	indirect effect exercise, $D = 1$					
sleeping well	1.927	0.975	1.765	0.115	2.216	1.684
	(1.666)	(1.359)	(2.502)	(1.727)	(2.218)	(1.992)
prudent alcohol	1.627	0.697	1.837	−0.006	1.840	1.502
	(1.500)	(1.264)	(2.277)	(1.628)	(2.042)	(1.986)
not obese	1.877	0.812	1.712	−0.409	1.823	1.116
	(1.666)	(1.339)	(2.822)	(1.768)	(2.116)	(1.905)
	indirect effect exercise, $D = 0$					
sleeping well	1.926	0.975	1.763	0.115	2.216	1.684
	(1.667)	(1.359)	(2.500)	(1.728)	(2.218)	(1.992)
prudent alcohol	1.626	0.697	1.836	−0.006	1.840	1.502
	(1.499)	(1.265)	(2.275)	(1.629)	(2.043)	(1.985)
not obese	1.876	0.812	1.710	−0.409	1.823	1.116
	(1.665)	(1.339)	(2.819)	(1.769)	(2.116)	(1.905)
	indirect effect third mediator, $D = 1$					
sleeping well	−3.647	−2.664	2.146	−1.527	−1.117	1.327
	(2.964)	(2.511)	(4.509)	(2.970)	(4.006)	(3.387)
prudent alcohol	0.936	−3.382	0.485	0.768	13.059	8.319
	(2.392)	(3.792)	(5.107)	(4.471)	(17.935)	(19.448)
not obese	−2.205	−1.698	0.161	6.356	−0.875	−3.516
	(6.571)	(6.258)	(13.136)	(8.155)	(6.709)	(6.261)
	indirect effect third mediator, $D = 0$					
sleeping well	−3.644	−2.666	2.144	−1.529	−1.117	1.327
	(2.961)	(2.514)	(4.513)	(2.973)	(4.007)	(3.386)
prudent alcohol	0.936	−3.385	0.484	0.769	13.069	8.312
	(2.391)	(3.796)	(5.100)	(4.474)	(17.963)	(19.415)
not obese	−2.203	−1.699	0.161	6.358	−0.876	−3.516
	(6.565)	(6.262)	(13.118)	(8.155)	(6.709)	(6.261)

[a] Never smoking and exercise included as first two mediation factors.

[b] Gamma unobserved heterogeneity. $^{+}p < 0.05$ $^{**}p < 0.01$. We do not report the mediated interaction because they are all close to zero, just as in Table 2.

Table E.2: Estimated educational gain in life-expectancy at age 18 principle strata Gompertz model, by bandwidth

bandwidth[a]	all		males		females	
	PH	MPH[b]	PH	MPH[b]	PH	MPH[b]
	LATE					
15 year	11.834**	10.559**	12.891^{+}	11.813**	8.028	4.509
	(4.172)	(3.880)	(5.107)	(4.530)	(5.593)	(4.263)
9 year	7.674	6.412	16.522	11.111	−2.015	–
	(5.265)	(5.709)	(8.633)	(6.427)	(5.475)	(–)
8 year	8.877	6.523	–	13.790	−2.368	−7.688
	(5.928)	(6.163)	(–)	(9.428)	(5.932)	(5.136)
7 year	8.181	5.896	13.367	11.096	1.688	−7.508
	(6.108)	(5.451)	(10.219)	(7.130)	(7.423)	(4.987)
	direct effect, dir(1)					
15 year	2.549	2.780	5.750	7.000	−3.718	−3.432
	(3.983)	(3.930)	(5.242)	(5.136)	(5.250)	(3.240)
9 year	−0.899	2.792	7.014	11.605	−8.212	–
	(3.774)	(5.835)	(6.012)	(7.063)	(4.365)	()
8 year	0.763	5.419	7.820	14.467	−8.150	−16.268^{+}
	(4.129)	(6.963)	(5.986)	(7.098)	(4.858)	(7.444)
7 year	2.421	4.708	7.030	11.349	−4.260	−16.936^{+}
	(4.399)	(6.251)	(6.962)	(7.418)	(5.189)	(6.859)
	direct effect,dir(0)					
15 year	2.548	2.781	5.748	7.004	−3.717	−3.433
	(3.982)	(3.932)	(5.240)	(5.137)	(5.247)	(3.242)
9 year	−0.899	2.793	7.009	11.604	−8.211	–
	(3.772)	(5.835)	(6.008)	(7.060)	(4.363)	()
8 year	0.763	5.419	7.810	14.465	−8.148	−16.268^{+}
	(4.127)	(6.692)	(5.980)	(7.097)	(4.857)	(7.444)
7 year	2.420	4.708	7.027	11.348	−4.260	−16.936^{+}
	(4.398)	(6.249)	(6.960)	(7.415)	(5.188)	(6.859)
	indirect effect Never smoking, $D = 1$					
15 year	7.976**	6.814**	3.552	2.411	10.639**	6.248^{+}
	(2.804)	(2.548)	(2.536)	(2.125)	(4.046)	(2.857)
9 year	4.839	5.264	3.554	1.792	4.596	–
	(2.935)	(2.994)	(4.204)	(2.750)	(3.126)	()
8 year	3.504	3.889	3.361	2.180	4.111	5.281
	(2.791)	(3.080)	(4.492)	(4.352)	(3.286)	(2.901)
7 year	2.523	2.684	0.440	−0.450	4.931	5.697
	(2.720)	(2.691)	(2.322)	(2.159)	(4.409)	(3.054)

[a] The bandwidth around 1-4-1933. The window of included births is twice the bandwidth. Note that for some bandwidths the model did not converge (especially for females)

[b] Gamma unobserved heterogeneity. $^{+}p < 0.05$ $^{**}p < 0.01$

Table E.2: continued

bandwidth[a]	all		males		females	
	PH	MPH[b]	PH	MPH[b]	PH	MPH[b]
	indirect effect Never smoking, $D = 0$					
15 year	7.977^{**}	6.816^{**}	3.550	2.414	10.640^{**}	6.248^{+}
	(2.804)	(2.548)	(2.537)	(2.127)	(4.046)	(2.857)
9 year	4.839	5.265	3.551	1.793	4.597	–
	(2.935)	(2.995)	(4.201)	(2.751)	(3.126)	()
8 year	3.503	3.889	3.554	2.180	4.110	5.281
	(2.791)	(3.080)	(4.204)	(4.352)	(3.285)	(2.901)
7 year	2.522	2.685	0.439	−0.451	4.932	5.697
	(2.720)	(2.691)	(2.321)	(2.161)	(4.409)	(3.054)
	indirect effect exercise, $D = 1$					
15 year	2.791^{+}	2.395^{+}	4.140^{+}	3.058	1.725	0.493
	(1.219)	(1.165)	(2.055)	(1.837)	(1.489)	(1.123)
9 year	2.653	1.725	3.343	0.306	1.600	–
	(2.018)	(1.729)	(3.584)	(1.960)	(2.222)	()
8 year	3.636	1.715	–	2.025	0.317	1.325
	(2.649)	(2.339)	()	(3.056)	(2.415)	(2.187)
7 year	2.988	1.573	4.486	1.331	1.592	1.043
	(2.867)	(2.137)	(5.596)	(2.565)	(2.996)	(2.285)
	indirect effect exercise, $D = 0$					
15 year	2.270^{+}	2.396^{+}	4.138^{+}	3.062	1.725	0.493
	(1.219)	(1.165)	(2.054)	(1.839)	(1.489)	(1.223)
9 year	2.652	1.725	3.339	0.307	1.600	–
	(2.017)	(1.729)	(3.581)	(1.962)	(2.223)	()
8 year	3.635	1.715	–	2.026	0.317	1.325
	(2.648)	(2.339)	()	(3.057)	(2.415)	(2.187)
7 year	2.987	1.574	4.483	1.333	1.592	1.043
	(2.866)	(2.137)	(5.593)	(2.568)	(2.997)	(2.285)
	indirect effect breakfast, $D = 1$					
15 year	−1.483	−1.432	−0.548	−0.662	−0.619	1.200
	(2.141)	(2.056)	(2.859)	(2.652)	(2.903)	(2.230)
9 year	1.081	−3.369	2.618	−2.591	1.014	–
	(2.530)	(2.854)	(4.921)	(3.398)	(2.299)	()
8 year	0.974	−4.498	5.133	−4.880	4.596	1.974
	(2.867)	(3.330)	(5.394)	(3.721)	(3.126)	(4.274)
7 year	0.251	−3.069	1.415	−1.132	4.596	2.688
	(3.008)	(3.082)	(5.442)	(3.820)	(3.126)	(4.160)
	indirect effect breakfast, $D = 0$					
15 year	−1.482	−1.434	−0.548	−0.663	−0.619	1.200
	(2.140)	(2.059)	(2.861)	(2.657)	(2.903)	(2.230)
9 year	1.081	−3.372	2.615	−2.595	−0.001	–
	(2.529)	(2.860)	(4.916)	(3.406)	(2.716)	()
8 year	0.974	−4.502	5.126	−4.884	4.596	1.974
	(2.866)	(3.337)	(5.388)	(3.724)	(3.126)	(4.274)
7 year	0.251	−3.073	1.414	−1.135	4.596	2.688
	(3.007)	(3.088)	(5.438)	(3.830)	(3.126)	(4.160)

[a] The bandwidth around 1-4-1933. The window of included births is twice the bandwidth.

[b] Gamma unobserved heterogeneity. $^{+}p < 0.05$ $^{**}p < 0.01$. We do not report the mediated interaction because they are all close to zero, just as in Table 2.

Table E.3: Estimated educational gain in life-expectancy at age 18 principle strata Gompertz model adjustment for Never takers

	Removing Never takers					
	all		males		females	
	PH	MPH[b]	PH	MPH[b]	PH	MPH[b]
LATE	7.410	4.502	12.276	8.363	2.434	1.916
	(4.233)	(3.641)	(6.495)	(5.111)	(5.333)	(5.788)
	direct effect					
dir(1)	0.246	1.205	6.146	8.005	−3.942	−3.923
	(3.257)	(3.171)	(5.585)	(4.804)	(3.780)	(3.816)
dir(0)	0.246	1.205	6.143	8.005	−3.941	−3.923
	(2.256)	(3.172)	(5.583)	(4.803)	(3.779)	(3.816)
	indirect effect, $D = 1$					
never smoking	4.780	2.665	2.540	0.103	5.687	5.365
	(2.572)	(1.879)	(3.348)	(2.094)	(3.482)	(3.654)
exercise	1.719	0.806	2.602	0.435	1.025	0.967
	(1.484)	(1.235)	(2.637)	(1.753)	(1.726)	(1.721)
breakfast	0.665	−0.173	0.992	−0.181	−0.336	−0.493
	(2.209)	(2.022)	(4.159)	(3.259)	(2.438)	(2.553)
	indirect effect, $D = 0$					
never smoking	4.780	2.665	2.538	0.103	5.687	5.634
	(2.572)	(1.880)	(3.346)	(2.095)	(3.482)	(3.653)
exercise	1.718	0.806	2.600	0.435	1.025	0.967
	(1.483)	(1.235)	(2.635)	(1.754)	(1.726)	(1.721)
breakfast	0.665	−0.174	0.991	−0.181	−0.336	−0.493
	(2.210)	(2.023)	(4.155)	(3.261)	(2.438)	(2.553)

[a] never smoking, exercise and breakfast as mediation factors

[b] Gamma unobserved heterogeneity. $^{+}p < 0.05$ $^{**}p < 0.01$. We do not report the mediated interaction because they are all close to zero, just as in Table 2.

Table E.3: Estimated educational gain in life-expectancy at age 18 principle strata Gompertz model adjustment for Never takers (continued)

	Assume Never takers have higher education					
	all		males		females	
	PH	MPH[b]	PH	MPH[b]	PH	MPH[b]
LATE	5.331	2.975	12.883^{+}	9.436	−1.361	−2.294
	(3.793)	(3.406)	(6.451)	(6.163)	(4.516)	(4.963)
	direct effect					
dir(1)	−0.202	0.799	6.442	8.771	−5.364	−5.378
	(2.946)	(2.941)	(4.950)	(4.502)	(3.526)	(3.612)
dir(0)	−0.202	0.799	6.439	8.772	−5.364	−5.379
	(2.945)	(2.941)	(4.948)	(4.502)	(3.525)	(3.613)
	indirect effect, $D = 1$					
never smoking	3.468	1.825	1.768	−0.451	3.914	3.468
	(2.160)	(1.647)	(3.018)	(1.936)	(2.681)	(2.659)
exercise	1.331	0.562	2.638	0.555	0.606	0.508
	(1.263)	(1.086)	(2.613)	(2.081)	(1.439)	(1.424)
breakfast	0.736	−0.211	2.040	0.559	−0.517	−0.892
	(1.999)	(1.895)	(3.976)	(3.706)	(2.207)	(2.421)
	indirect effect, $D = 0$					
never smoking	3.467	1.825	1.767	−0.451	3.915	3.467
	(2.160)	(1.647)	(3.016)	(1.937)	(2.682)	(2.658)
exercise	1.330	0.562	2.635	0.555	0.606	0.508
	(1.263)	(1.087)	(2.611)	(2.082)	(1.440)	(1.424)
breakfast	0.735	−0.211	2.038	0.560	−0.517	−0.891
	(1.998)	(1.895)	(3.973)	(3.708)	(2.208)	(2.420)

[a] never smoking, exercise and breakfast as mediation factors

[b] Gamma unobserved heterogeneity. $^{+}p < 0.05$ $^{**}p < 0.01$. We do not report the mediated interaction because they are all close to zero, just as in Table 2.