



Deposited via The University of York.

White Rose Research Online URL for this paper:

<https://eprints.whiterose.ac.uk/id/eprint/206929/>

Version: Published Version

---

**Article:**

Bijwaard, Govert and Jones, Andrew Michael (2024) Regression discontinuity design with principal stratification in the mixed proportional hazard model: An application to the long-run impact of education on longevity. *Empirical Economics*. pp. 197-223. ISSN: 0377-7332

<https://doi.org/10.1007/s00181-023-02553-0>

---

**Reuse**

This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:

<https://creativecommons.org/licenses/>

**Takedown**

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing [eprints@whiterose.ac.uk](mailto:eprints@whiterose.ac.uk) including the URL of the record and the reason for the withdrawal request.



# Regression discontinuity design with principal stratification in the mixed proportional hazard model: an application to the long-run impact of education on longevity

Govert E. Bijwaard<sup>1</sup> · Andrew M. Jones<sup>2</sup>

Received: 15 June 2023 / Accepted: 23 December 2023

© The Author(s) 2024

## Abstract

We investigate the long-run impact of education on longevity using data for England and Wales from the Health and Lifestyle Survey. Longevity is modelled by survival analysis using a mixed proportional hazard model. For identification we propose a Regression Discontinuity Design implied by an increase in the minimum school leaving age in 1947 (from 14 to 15) combined with a principal stratification method for estimation of the mortality hazard rate. This method allows us to derive the causal effect of extended education on longevity. In line with earlier studies we do not find credible evidence of a causal impact of the additional years of schooling that were induced by the reform on longevity.

**Keywords** Regression discontinuity design · Education · Mortality · Principal strata · Mixture models

**JEL Classification** C41 · I14 · I24

## 1 Introduction

A popular approach to identify causal effects of education on health and longevity 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. These studies exploit an identification strategy that assumes that these

---

✉ Govert E. Bijwaard  
Bijwaard@nidi.nl; G.E.Bijwaard@rug.nl

Andrew M. Jones  
andrew.jones@york.ac.uk

<sup>1</sup> Netherlands Interdisciplinary Demographic Institute (NIDI-KNAW/University of Groningen), Lange Houtstraat 19, 2502 AR The Hague, The Netherlands

<sup>2</sup> Department of Economics and Related Studies, University of York, York YO10 5DD, UK

changes in the 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 affects health and longevity through its effect on education, one can estimate a causal effect of the additional education on longevity for those who comply with the new law and would not have done so 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) of education on mortality. Here, we use the British Health and Lifestyle Survey (HALS) and exploit an educational reform in 1947 that increased the legal minimum school leaving age in England and Wales from 14 to 15 (Clark and Royer 2013).

A reason why higher education may lead to lower mortality is that the higher educated are more efficient producers of health investment (Grossman 1972, 2006). Grossman (1972) argues that this could be due to (i) *productive efficiency* or (ii) *allocative efficiency*. The former hypothesis posits that the higher educated understand medical advice better and use medical care more efficiently. The allocative efficiency hypothesis on the other hand argues that higher educated individuals choose different, more efficient inputs into health investment, typically thought to be caused by better health knowledge and a more receptive attitude towards new information.

Many studies investigating the impact of education on mortality have used linear models (see e.g., Lleras-Muney 2005; Van Kippersluis et al. 2011; Clark and Royer 2013) to estimate the educational gradient, facilitating ‘standard’ instrumental variable estimation. However, age at death is clearly a duration outcome and, hence, we use a non-linear model for the mortality hazard rate. Duration analysis models the 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 till the end of the observation window, can be modelled directly within this framework. A common characteristic of duration data, including time to death, is that not all individuals experience the event of interest during the observation period. Such right censoring makes inference based on means unreliable. Thus, using survival until survey end would not account for such right-censoring. Another characteristic of duration data is dynamic selection or left truncation: those still alive at the age that the survey starts may not be a random selection of the original population of births. This excludes the comparison of simple survival differences at the end of a survey. We therefore use the (mortality) hazard, or the force of mortality, as this effectively deals with these data characteristics (see e.g., Lancaster 1990; Van den Berg 2001). A common way to accommodate the presence of observed characteristics in a duration model is to specify a proportional hazard (PH) model, in which the hazard is the product of the baseline hazard, that captures the age dependence of the hazard, and a log-linear function of covariates. Neglecting unobserved confounding in inherently non-linear models, such as the proportional hazard model, leads to biased inference. The common approach to address this is to explicitly model the individual-specific effects using unobserved heterogeneity that enters the hazard function multiplicatively, known as the Mixed Proportional hazard (MPH) model.

Studies have attempted to identify the causal effect of education on mortality, 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 there is 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, do account for correlated unobserved factors they assume a particular structure. In contrast, the compulsory schooling change provides a natural instrument to identify the causal effect of education on the mortality rate. However, no unambiguous solution to instrumental variable estimation of the inherently non-linear MPH-model 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, 2018) have shown that both the two-stage predictor substitution (2SPS) and the two-stage residual inclusion (2SRI) methods of Terza et al. (2008) are biased in a Weibull proportional hazard framework, at least under the standard assumptions common in the treatment evaluation literature.

The change in the compulsory schooling law offers a (fuzzy) Regression Discontinuity Design (RDD), as it generated a discontinuity in the “treatment” (number of schooling years) for those affected when the reform was implemented. We use the local randomization framework of the RDD (Lee 2008; Lee and Lemieux 2010; Cattaneo et al. 2015), where the treatment assignment (staying on longer at school) is assumed to be as-if randomly assigned in a small interval around the reform implementation date. In the local randomization framework of the RDD a principal stratification into complier types follows naturally (Imbens 2016).

The principal stratification framework (Frangakis and Rubin 2002) is a general potential outcomes framework for causal inference with instruments and/or intermediate variables. Principal stratification has its roots in instrumental variable methods, as described in Angrist et al. (1996); 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 who have 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 irrespective of their assigned treatment (*ii*) *Never takers* individuals who never take the treatment (*iii*) *Compliers* individuals who only take the treatment if assigned to treatment (*iv*) *Defiers*: individuals who only take the treatment if not assigned to treatment. Defiers 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, with 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); Wan et al. (2015).

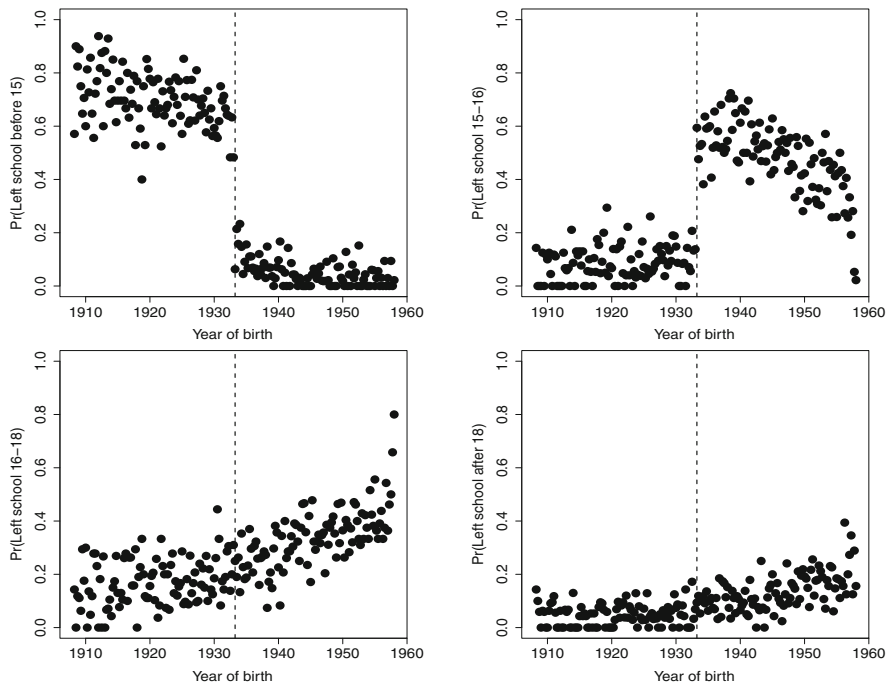
The contribution of this paper is to provide a methodological innovation in instrumental variable analysis for hazard rate models, using the principal stratification approach to motivate estimation of a mixture model.

## 2 Data and descriptive statistics

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 wave 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 only 5352 individuals completed the interviews. We therefore focus on the first wave. 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 we remove all individuals living in Scotland from the sample). On 1 April 1947, the legal school leaving age was raised to age 15 in Britain, while until 31 March 1947 children in Britain could leave school when they reached 14 years of age. This reform affected children who turned 14 after 31 March 1947 (born after 31 March 1933) as they had to stay at school longer.

Figure 1 shows how the 1947 reform affects the school leaving age, the probability of leaving school before the age of 15, the probability of leaving school between age 15 and 16, the probability of leaving school between age 16 and 18 and, the probability of leaving school after age 16. The 1947 reform clearly had a large effect on school leaving around the age of 15, but not on leaving school after age 16.

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 probability to survive until the end of the survey



**Fig. 1** Probability to leave school before age 15, age 15–16, age 16–18 and after age 18 (around the cut-off birth date of 1–4–1933) Dots represent average schooling from survey entry to survey end by quarter of birth

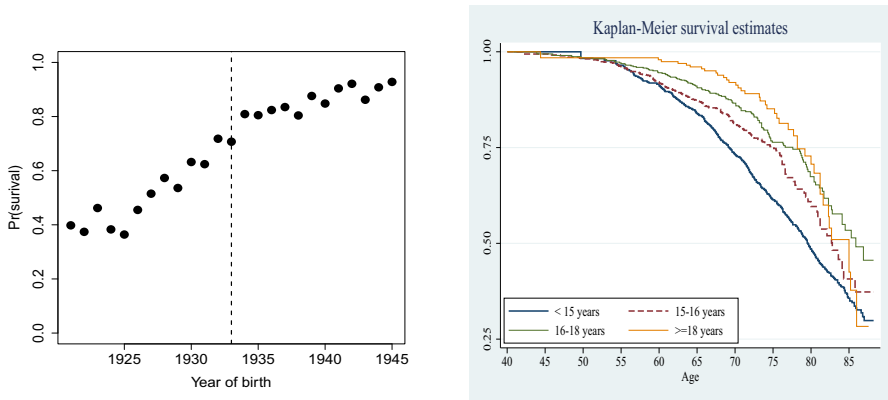
(July 1st, 2009) and the Kaplan–Meier survival curves for individuals born within 12 years before or after the cut-off birth of the 1947 reform.

Note that the survival gaps, depicted in the right-hand plot of Fig. 2, are based on the raw survival data and these could exist for a multitude of reasons, including selection, reverse causality and, potentially, a causal impact of education on mortality. 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).

### 3 Regression discontinuity design and principal stratification

Understanding the causal effect of a treatment  $D$  (education) on an outcome  $Y$  (longevity) is fundamental goal of social science. The identification of the causal effect is complicated by the potential endogeneity of education. The association between longevity and education may partly be explained by confounding factors such as cognitive ability and parental background, which affect both education choices and longevity (McCartney et al. 2013).

To address this endogeneity we use a fuzzy regression discontinuity design, as implied by the change in minimum school leaving age of the 1947-reform in England



**Fig. 2** Survival from first survey till end of follow up by age left school, only individuals born within 12 years before or after the cut-off birth date (31–3–1933) for the 1947 reform Dots represent average survival from survey entry to survey end by year of birth for survival

and Wales, in a principal stratification framework. Note that a standard (proportional) hazard model for the mortality rate, such as a Gompertz model, using only observation within the RDD bandwidth is likely to be biased, as it still does not account for the endogeneity. Our instrumental variable method, based on principal stratification, a non-linear extension of the commonly applied linear Local Average Treatment effect (LATE) approach (Angrist et al. 1996), can provide an unbiased estimate of the effect of education on longevity.

### 3.1 Regression discontinuity as a local randomized experiment

Before we elaborate on the non-linear analysis, we define the instrument used in this study and the regression discontinuity design. A method using observations close to a threshold to identify causal effects is known as a *regression discontinuity design* (RDD), (Imbens and Lemieux 2008; Lee and Lemieux 2010). The basic idea behind RDD is that assignment to treatment (in our case, continuing 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 (i.e., the “running variable” is the birth date and the threshold is the date the reform was implemented, 1–4–1933). Because people born before the reform could still stay in school beyond age 15 we have a fuzzy RDD.

In the local randomization-based approach to the RD design (Lee 2008; Lee and Lemieux 2010; Cattaneo et al. 2015), it is hypothesized that, within some finite window of an administrative threshold (e.g., a test score or age cutoff) that determines treatment assignment, subjects are “as-if” randomly assigned to treatment and control.

Formally, let  $W_0 = [r_c - h, r_c + h]$  with  $r_c$  the threshold and  $h$  the window width, the local randomization assumption can be stated as the following two assumptions:

- (A) The distribution of the running variable (birth date) in the window  $W_0$  is known and does not depend on the potential outcomes.

(B) Inside  $W_0$ , the potential outcomes (potential mortality) depend on the running variable solely through the treatment indicator (stay at school beyond age 15).

Assuming that the birth date will have no effect on the (potential) mortality is unrealistic. However, as Cattaneo et al. (2017) show, if the effect of the birthday on the potential mortalities can be captured by a polynomial of order  $p$  on the distance of the individual birthday from the threshold, it is possible to allow that the potential outcomes depend on the running value (birthday). We add a local polynomial in the distance from the threshold with the order of the polynomial chosen to minimize the AIC given the bandwidth (just as in Lee and Lemieux 2010). The fuzzy RDD can be viewed as an instrumental variable method, with the change in the law used as instrument for staying longer in school.

### 3.2 Choice of bandwidth

In general, choosing a bandwidth involves finding an optimal balance between precision and bias. On the one hand, using a larger bandwidth yields more precise estimates as more observations are available to estimate the regression. On the other hand, the specification is less likely to be accurate when a larger bandwidth is used, which can bias the estimate of the treatment effect.

In the local randomization approach the choice of the optimal bandwidth is based on a sequential randomization test. We follow the practical steps suggested by Cattaneo et al. (2015) to establish whether local randomization is plausible in small windows around the cut-off and determine the size of such a window. The procedure involves a simple difference-in-means test for the predetermined covariates comparing their values on each side of the cut-off. This test is carried out for each candidate window. If the  $p$ -value regarding the null that a covariate has the same value for both sides of the cut-off is below 0.15 (Cattaneo et al. 2015; Cattaneo and Titiunik 2022), then that window is rejected and we attempt the procedure with a smaller window. A window is selected if one cannot reject the null for any of the predetermined covariates using a threshold  $p$ -value of 0.15.

### 3.3 Assumptions for identification of causal effects

Following the literature, we define causal effects using the potential outcomes (or counterfactual) framework. Define for the policy change (treatment assignment) the (potential) discrete  $D(z)$ , with  $Z = 1$  if an individual was affected by the policy change and zero otherwise and  $D = 0$  if the individual left school before age 15,  $D = 1$  if the individual left school at age 15 to age 16,  $D = 2$  if the individual left school at age 16 to age 18 and,  $D = 3$  if the individual left school at age 18 or beyond. We assume that the policy change does not affect the choice to stay at school after age 16.

We use the *principal strata* formulation of the problem (Frangakis and Rubin 2002). This implies we have six (latent) complier types ( $P$ ) for education: *always takers 1* are individuals who always leave school at age 15 to age 16 irrespectively of whether they were affected by the policy change (i.e.,  $D(1) = D(0) = 1$ ;  $P = a_1$ ); *always takers 2* are individuals who always leave school at age 16 to age 18 irrespectively of

whether they were affected by the policy change (i.e.,  $D(1) = D(0) = 2$ ;  $P = a_2$ ); *always takers* are individuals who always stay beyond age 18 at school irrespectively of whether they were affected by the policy change (i.e.,  $D(1) = D(0) = 3$ ;  $P = a_3$ ); *never takers* are individuals who never stay in school beyond age 15 (i.e.,  $D(1) = D(0) = 0$ ;  $P = n$ ). Under our identification strategy *always takers* and *never takers* do not contribute to identification of the local treatment effect. *Compliers* are individuals who only stay in school to age 15 to age 16 in school because they were induced to do so through the policy change (i.e.,  $D(1) = 1$  and  $D(0)=0$ ;  $P=c$ ). It is the *compliers* that identify the local treatment effect of an extended education.

Following the literature on potential outcomes we impose the following assumptions:

**Assumption 1: Stable unit value assumption (SUTVA)**

SUTVA implies that potential outcomes, for each person  $i$  are unrelated to the treatment status (education) of other individuals.

**Assumption 2: Ignorable instrument**

$$\{Y(d), D(z)\} \perp Z | X$$

This assumption typically holds in a randomized experiment. The assumption is also plausible in observational studies where  $Z$  represents an instrumental variable that is regarded as exogenous after (possibly) conditioning on observed covariates.

**Assumption 3: Exclusion restrictions**  $\forall z = 0, 1; d = 0, 1, 2, 3$ :

$$Y(z, d) = Y(d)$$

This assumption states that the instrument  $Z$  can only affect the outcome through its effect on education. This implies that the potential outcome can be written as  $Y(d)$ . This also implies that the effect of always-takers and of never-takers is independent of treatment assignment. Note that this restriction is inherent in the RDD, the policy change only effects the outcome through the induced change in treatment (prolonging the time in school).

**Assumption 4: Monotonicity**  $D(1) \geq D(0)$

Assumption 4 rules out the existence of *Defiers*, individuals who only stay in school to age 15 to age 16 because they were not induced to do so through the policy change. This implies that the educational effect on the outcome is only identified for compliers while the educational effect for never takers and always takers is not identified.

### 3.4 Principal strata hazard rate model

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 use the principal strata framework to show under the identification of the causal effects. Denote the complier type probabilities by  $P^{a_1}, P^{a_2}, P^{a_3}, P^{a_c}, P^{a_n}$ , the probability of being an always taker, never taker or complier (possibly conditional on  $X$ ), which can be derived from cross tabulation of education and the instrument

$\Pr(D = d|Z = z)$ :

$$\begin{aligned} \Pr(D = 0|Z = 1) &= P^n \\ \Pr(D = 0|Z = 0) &= p^n + p^c \\ \Pr(D = 1|Z = 1) &= p^{a1} + p^c \\ \Pr(D = 1|Z = 0) &= p^{a1} \\ \Pr(D = 2|Z = 1) &= \Pr(D = 2|Z = 0) = p^{a2} \\ \Pr(D = 3|Z = 1) &= \Pr(D = 3|Z = 0) = p^{a3} \end{aligned}$$

Thus  $p^c = \Pr(D = 1|Z = 1) - \Pr(D = 1|Z = 0)$ . All these probabilities are estimated jointly with the other parameters of the model. This implies that our specification is a latent class model with the complier types modelled as latent classes, with the LATE identified for the sub-set of compliers:

$$LATE = E\left[Y(1) - Y(0) \mid P = c\right] \tag{1}$$

Estimating a principal strata model gives the required functions (see the next sub-Section). Note that in our application, the compliers are the sub-population who have additional years of schooling induced by the change in the minimum school leaving age and the impact of this change in education can therefore be regarded as being due to plausibly exogenous variation.

We assume a Gompertz proportional hazard mortality rate, which postulates that the (baseline) hazard increases exponentially with age (e.g.,  $\lambda(t|X) = e^{\beta_0 + \alpha t + \beta' X}$ ).<sup>1</sup> We use the (implied) life expectancy as the outcome of interest.<sup>2</sup> Assuming that the estimated Gompertz hazard holds, the life expectancy can be very well approximated by Lenart (2014):

$$\mu_X = -\exp\left(\frac{1}{\alpha} \exp[\beta_0 + \beta' X]\right) \times (\beta_0 + \beta' X - \ln(\alpha) - e^{\beta' X} + 0.5772) / \alpha \tag{2}$$

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  $\sigma^2$ ) the life expectancy can be approximated by Missov (2013):

$$\begin{aligned} \mu_X &= \left[ \left(1 - \frac{\sigma^2}{\alpha} e^{\beta_0 + \beta' X}\right)^{-\frac{1}{\sigma^2}} \left(\ln(\alpha) - \beta_0 - \beta' X - \ln(\sigma^2)\right) \right. \\ &\quad \left. - \sum_{j=1}^{\frac{1}{\sigma^2}-1} \frac{1}{j} \left(1 - \frac{\sigma^2}{\alpha} e^{\beta_0 + \beta' X}\right)^{j-\frac{1}{\sigma^2}} \right] / \alpha \end{aligned} \tag{3}$$

<sup>1</sup> Another parametric mortality rate, e.g., a Weibull model ( $\lambda_0(t) = \alpha t^{\alpha-1}$ ) is also possible.

<sup>2</sup> In principle the causal effect can also be defined in terms of hazard (ratios). But these effects depend on age,  $t$ , and are therefore difficult to interpret.

Due to right-censoring (which is affected differently by education ) we cannot use the average duration directly to estimate the model parameters. Using a hazard rate model effectively accounts for censoring (and possible time-varying covariates) and allows us to estimate all the parameters. Life-expectancy can then be derived from the estimated parameters.

We assume the complier type influences only the scale of the mortality rate,  $\gamma_1$  (for a complier who is induced to continue schooling due to the instrument,  $Z = 1$ ) and  $\gamma_0$  (for a complier who is induced not to continue schooling due to the instrument  $Z = 0$ ). Thus, the potential hazard for an individual of complier type  $P = \{a(lways), n(ever), c(omplier)\}$  is:

$$\lambda^{(P)}(t|\cdot) = v\lambda_0(t; \alpha) \exp\left(\gamma_{a_1}I(P = a_1) + \gamma_{a_2}I(P = a_2) + \gamma_{a_3}I(P = a_3) + \gamma_nI(P = n) + (\gamma_1D + \gamma_0(1 - D))I(P = c) + \beta'X\right) \quad (4)$$

Note that, due to Assumption 3, for always takers (1,2, or 3)  $\gamma_{a_1}, \gamma_{a_2}, \gamma_{a_3}$  do not depend on  $D$ , similarly for never takers we only have  $\gamma_n$ . For compliers the education level  $D$  is either zero or one. We either assume that  $v \equiv 1$  (PH-model) or that  $v$  follows a unit mean Gamma distribution with variance  $\sigma^2$  (MPH-model).

### 3.5 Estimation of principal strata hazard rate model

Based on the assumption of a known functional form of the baseline hazard, such as a Gompertz ( $\lambda_0(t) = e^{\alpha t}$ ), we can derive the likelihood function contribution of individual  $i$ , see Appendix A for the full likelihood<sup>3</sup>:

$$L_i = \prod_{Z_i} \prod_{D_i} \left[ \lambda_i(t_i|Z_i, D_i)^{\delta_i} \times S(t|Z_i, D_i) \right]^{I(Z_i, D_i)} \quad (5)$$

where  $S(t|Z_i, D_i)$  is the survival rate at age  $t$  for an individual with  $Z_i, D_i$ , e.g.:

$$S(t|Z = 0, D = 0) = p^n \exp\left(-\int_0^t \lambda_0(s)e^{\beta'X} ds\right) + p^c \exp\left(-\int_0^t \lambda_0(s)e^{\beta'X + \gamma_0} ds\right)$$

or:

$$S(t|Z = 1, D = 3) = p^{a_3} \exp\left(-\int_0^t \lambda_0(s)e^{\beta'X + \gamma_{a_3}} ds\right)$$

RDDs identify a treatment effect locally around the threshold. A local continuity assumption is standard in the literature, implying that persons close to the threshold are comparable except for their values of the assignment variable. The standard approach

<sup>3</sup> An alternative would be a Bayesian approach, as suggested by Li et al (2015).

to account for divergence is to include a local polynomial of the running variable, in our case the date of birth, estimated separately on each side of the threshold. We let the AIC determine the order of the polynomial functions of the time of birthdate from April 1933, separately for each side of the threshold. In Sect. 5 we discuss robustness checks based on smaller windows.

## 4 Empirical results

Our identification strategy relies on the mixed proportional hazard with principal stratification to identify a LATE. This LATE focuses on the *compliers*, who were influenced by the increase in minimum school leaving age, and the relevant treatment relates to the binary comparison of those who left school before age 15 and those who left school at ages 15 or 16. Before we report the results of this principal strata model we discuss the results for a standard Gompertz model (with or without gamma distributed unobserved heterogeneity) for the mortality rate when a dummy for staying in school beyond age 15 is one of the included variables. This provides a benchmark for our causal inference.

We base the bandwidth choice around April 1933 on a randomization test, see Sect. 3.2 and use a bandwidth of 12 years. This seems a large bandwidth around the cutoff date but is comparable to other bandwidths used in the literature, e.g., Clark and Royer (2013) use a bandwidth of 15 years and both Van Kippersluis et al. (2011) and Johnston et al. (2015) use a bandwidth of 10 years. The first panel of Table 1 give the estimated effect on the mortality hazard. The first two columns of Table 1 provide the estimated coefficients for the basic (M)PH Gompertz model.<sup>4</sup> In this benchmark model staying in school beyond age 15 is associated with a reduction of the mortality hazard by 33% ( $=1 - e^{-0.403}$ ). Including a gamma distributed unobserved heterogeneity (MPH) increases the association with staying in school. We calculate the implied life-expectancy of leaving school after age 15, based on the estimated parameters and using Eqs. (2) or (3), which are reported in the second panel of Table 1. Again the first two columns report the estimated educational gains in the implied life-expectancy for the standard Gompertz model. In the basic Gompertz model we find a total association between staying in school beyond age 15 and life-expectancy of 5.7–6.0 years.

The standard Gompertz model presented above does not account for the potential endogeneity of staying in school. Similar to the ‘standard’ RDD analysis that involves an instrumental variable method, like 2SLS, the principal strata Gompertz model described in Sect. 3.4, that exploits the policy reform of 1947 as an instrument for staying in school, seeks to solve this endogeneity issue.

The third and fourth columns of Table 1 report the difference in the hazard parameters for the compliers:  $\gamma_1 - \gamma_0$ . The full estimation results are given in Table 10 in Appendix D. We find a large and statistically significant effect of staying in school till age 15–16 instead of leaving school before age 15 on the hazard rate for the compliers, when the distance from the threshold is ignored (i.e., the order of the polynomial in the running variable is zero), see Table 10 in Appendix D. However, this estimate lacks

<sup>4</sup> The full parameter estimates are in Table 7 in Appendix D.

**Table 1** Estimated impact of schooling on the mortality hazard and estimated educational gain in life-expectancy,  $N = 2750$ 

Gompertz		Principal strata	
PH	MPH <sup>1</sup>	PH	MPH <sup>1</sup>
<i>Mortality hazard</i>			
-0.403**	-0.497**	0.176	0.074
(0.082)	(0.117)	(0.205)	(0.229)
<i>Educational gain</i>			
5.65**	6.04**	-2.45	-0.85
(1.20)	(1.22)	(2.83)	(2.67)

Based on AIC the order of the local polynomial trend in month-year of birth (from 1-4-1933) is determined, see Table 6 in 1. The bandwidth around 1-4-1933 included in the estimation is based on the Randomization test and is 12 years. The window of included births is twice the bandwidth. <sup>+</sup> $p < 0.05$ ; <sup>\*\*</sup> $p < 0.01$

<sup>1</sup>Gamma unobserved heterogeneity

credibility as it excludes a direct influence of birth date (the period effect or a secular trend) on the mortality hazard. The model that performs best on statistical criteria, based on the AIC, contains a second order polynomial in the distance to the threshold and leads to a positive and statistically insignificant effect.

Based on these estimated parameters we calculate the life-expectancy for each level of education, using Eqs. (2) or (3). The third and fourth columns of Table 1 report the estimated educational gains for the principal strata models. The total educational gain for the preferred model, with a second order polynomial in the running variable, estimates a statistically insignificant decrease in life-expectancy.

## 5 Robustness checks

In this section we check how robust our results are for males and females separately, to including covariates, to the choice of the bandwidth and to adjusting for never takers.

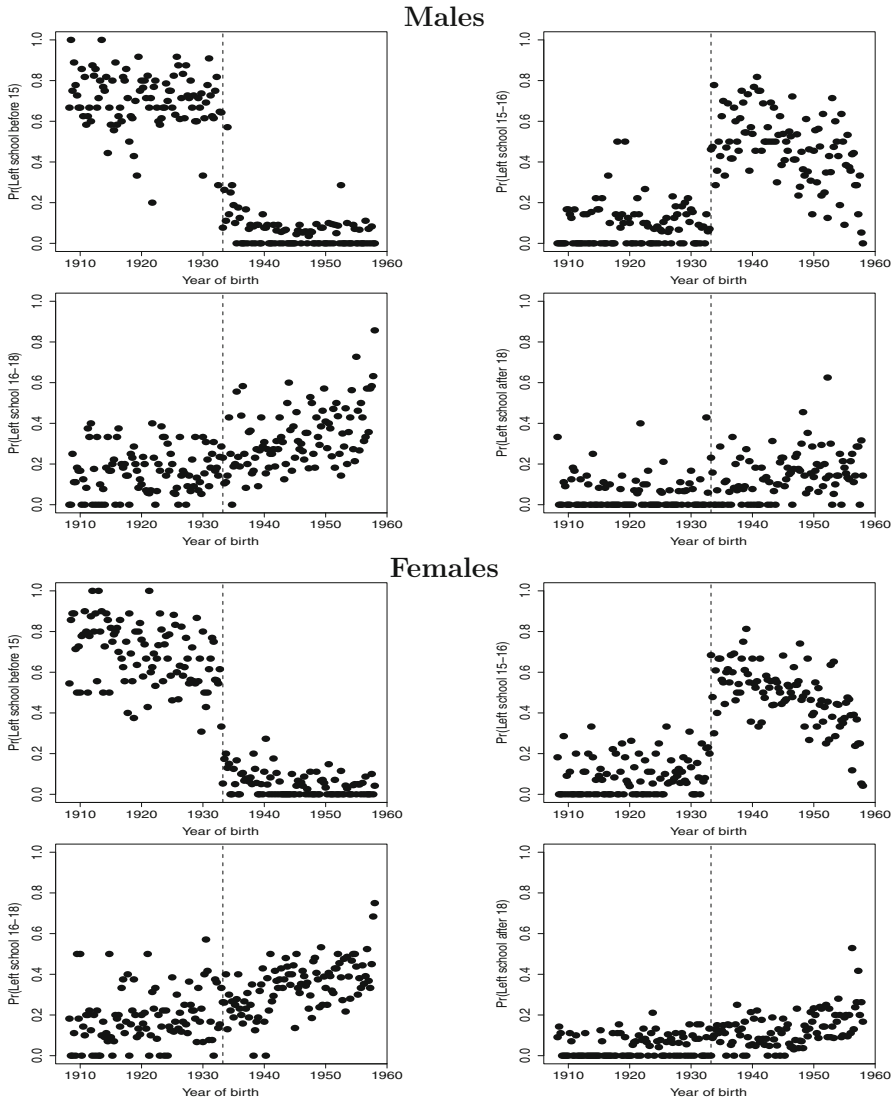
Figure 3 shows how the 1947 reform affects the school leaving age for males and females separately. Again, the 1947 reform clearly had a large effect on school leaving around the age of 15, but not on leaving school after age 16 for both males and females.

Figure 4 depicts the probability to survive until the end of the survey (July 1st, 2009) and the Kaplan–Meier survival curves for individuals born within 12 years before or after the cut-off birth of the 1947 reform for males and females separately.

Note that, as stated earlier the survival gaps, depicted in the right-hand plots of Fig. 4, are based on the raw survival data and could exist for many reasons including both selection bias or a causal impact.

We re-estimate the standard Gompertz and the principal strata model separately for males and females. The first two columns of Table 2 provide the estimated coefficients for the basic (M)PH Gompertz model.<sup>5</sup> Again the standard Gompertz model predicts a large reduction of the mortality hazard from staying in school beyond age 15. Including

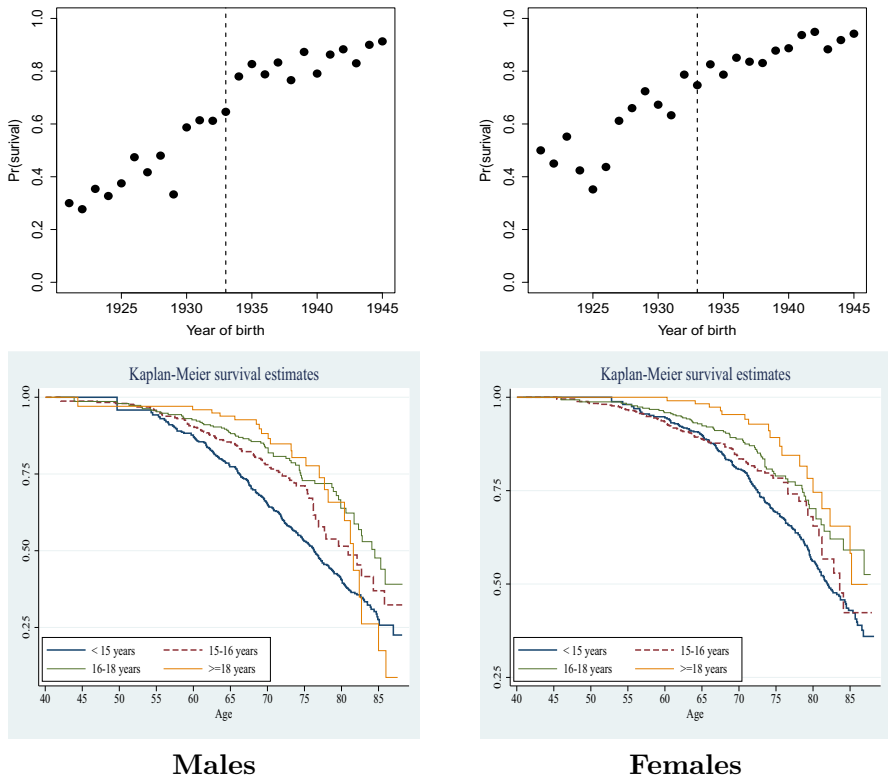
<sup>5</sup> The full parameter estimates are in Table 8 and 9 in Appendix D.



**Fig. 3** Probability to leave school before age 15, age 15–16, age 16–18 and after age 18 (around the cut-off birth date of 1–4–1933) for males and females separately Dots represent average schooling from survey entry to survey end by quarter of birth

a gamma distributed unobserved heterogeneity (MPH) hardly affects the size of this effect. For men, the preferred model leads to a negative but statistically insignificant effect of staying longer in school on the hazard rate among the compliers for women the estimate is positive (and also statistically insignificant).

We base the bandwidth choice around April 1933 again on a randomization test and use a bandwidth of 14 years (males only) or 13 years (females only). Note that the different bandwidths imply that the sum of the sample size of the males and of



**Fig. 4** Survival from first survey till end of follow up by age left school, only individuals born within 12 years before or after the cut-off birth date (31–3–1933) for males and females separately Dots represent average survival from survey entry to survey end by year of birth for survival

the females is not equal to the total sample size ( $N = 2750$ ). The second panel of Table 2 reports the estimated educational gains in the implied life-expectancy, the first two columns for the standard Gompertz model and the third and fourth columns for the principal strata model. In the standard Gompertz model we find a total association between staying in school beyond age 15 and life-expectancy of 5.4 years (males) and 5.1 years (females). The estimated educational gains using the principal strata model are not statistically significant and for women the estimate is negative.

In a local randomization view of RDD additional covariates, data beyond the outcome and the running variable, are used to find the bandwidth around the threshold (see Sect. 3.2). Researchers often use additional covariates to reduce the variance of their empirical estimates. A common strategy is to include the covariates additively separably and linearly-in-parameters in a local linear RD regression (Calonico et al. 2019). Table 3 provides the estimated effect on the hazard and the estimated educational gain using a RDD (principal strata model) with additional covariates (the exogenous variables, region and sex). These results do not differ substantially from those without covariates and are still statistically insignificant.

**Table 2** Estimated coefficient of schooling on the mortality hazard, males and females separately

Gompertz		Principal strata	
PH	MPH <sup>1</sup>	PH	MPH <sup>1</sup>
Coefficient of schooling			
<i>Males only</i>			
-0.391**	-0.391**	-0.331	-0.331
(0.100)	(0.100)	(0.219)	(0.253)
<i>Females only</i>			
-0.452**	-0.479**	0.408	0.407
(0.100)	(0.116)	(0.264)	(0.267)
Educational gain in life-expectancy			
<i>Males only</i>			
5.36**	5.36**	4.49	4.49
(1.43)	(1.43)	(2.97)	(2.87)
<i>Females only</i>			
5.10**	5.07**	-4.73	-4.64
(1.27)	(1.26)	(3.03)	(3.10)

Including covariates (region and sex, only for all individuals). Based on AIC the order of the local polynomial trend in month-year of birth (from 1-4-1933) is determined, see Table 6 in 1. The bandwidth around 1-4-1933 included in the estimation is based on the Randomization test and is 14 (males only) or 13 (females only) years. The window of included births is twice the bandwidth. <sup>+</sup> $p < 0.05$ , <sup>\*\*</sup> $p < 0.01$ ,  $N = 1478$  for males,  $N = 1657$  for females

<sup>1</sup> Gamma unobserved heterogeneity

**Table 3** Estimated impact of schooling on the mortality hazard and estimated educational gain in life-expectancy, principal strata model including covariates

PH	MPH <sup>1</sup>
<i>Mortality hazard</i>	
0.185	0.089
(0.204)	(0.228)
<i>Educational gain</i>	
-2.53	-0.97
(2.77)	(2.51)

Including covariates (region and sex, only for all individuals). Based on AIC the order of the local polynomial trend in month-year of birth (from 1-4-1933) is determined, see Table 6 in 1. The bandwidth around 1-4-1933 included in the estimation is based on the Randomization test and is 12 (all), 14 (males only) or 13 (females only) years. The window of included births is twice the bandwidth. <sup>+</sup> $p < 0.05$ ; <sup>\*\*</sup> $p < 0.01$

<sup>1</sup> Gamma unobserved heterogeneity

A common issue in using fuzzy regression discontinuity designs is the choice of the bandwidth for whom to select around the threshold. There is always a trade-off between bias and precision. The chosen window of 12 years around the threshold of being born in April 1933 is based on the randomization test. Table 4 reports the total

**Table 4** Estimated educational gain in life-expectancy principal strata model, by bandwidth

Bandwidth	PH	Gamma <sup>†</sup>
	10	2.10 (2.78)
9	1.37 (2.97)	1.70 (2.95)
8	0.88 (3.20)	1.59 (2.97)
7	3.34 (3.36)	3.77 (3.27)
6	4.25 (4.04)	4.60 (3.81)
5	3.69 (4.27)	4.05 (4.11)

The window of included births is twice the bandwidth. Based on AIC the order of the local polynomial trend in month–year of birth (from 1–4–1933) is determined. <sup>+</sup> $p < 0.05$ ; <sup>\*\*</sup> $p < 0.01$

<sup>†</sup>PH: no unobserved heterogeneity; Gamma: Gamma unobserved heterogeneity

educational gains in life-expectancy estimated for the principal strata model using smaller bandwidths from 10 down to 5 years. For all bandwidths the estimated total educational gains are statistically insignificant.

Another issue is that we identified a few (5%) never-takers, individuals that report leaving school before the age of 15 in the post-policy period when the legal school leaving age had been increased. We, therefore, check what happens if either (1) we remove never takers from the sample or (2) assume that these people stayed in school beyond age 15. Again we re-estimate the complier model and calculate the educational gains in life-expectancy. Table 5 presents the estimated total educational gains with these two adjustments for never takers. None of the estimated educational gains are statistically significant.

Finally, using a local treatment framework implies that all the assumptions (in Sect. 3.3) hold. We visually test the exclusion restriction (assumption 3) and the monotonicity assumption (Assumption 4). A joint graphical ‘test’ of the exclusion restriction and Monotonicity (Kitagawa 2015; Mourifié and Wan 2017) is:

$$S(t|D = 1, Z = 1) \Pr(D = 1|Z = 1) \geq S(t|D = 1, Z = 0) \Pr(D = 1|Z = 0)$$

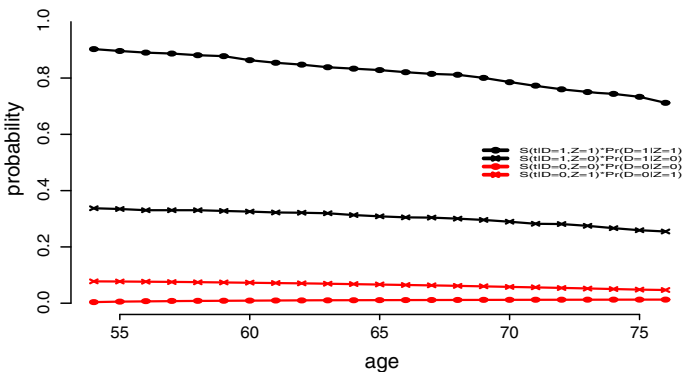
$$S(t|D = 0, Z = 0) \Pr(D = 0|Z = 0) \geq S(t|D = 0, Z = 1) \Pr(D = 0|Z = 1)$$

Thus, a figure with these four curves  $S(t|\cdot) \times \Pr(D|Z)$  serves as a graphical test of the validity of these assumptions regarding the instrument, as defined by the threshold before and after the reform of 1947. Figure 5 depicts the four curves and shows that the two inequalities hold (for the age interval with sufficient observations: ages 54–76).

**Table 5** Estimated educational gain in life-expectancy at age 18 principal strata model, adjustment for Never takers

PH	MPH <sup>1</sup>
Removing never takers	
<i>Educational gain</i>	
-1.07 (2.48)	0.03 (2.35)
Assuming never takers continue	
<i>Educational gain</i>	
-1.96 (2.38)	-0.88 (2.29)

The bandwidth around 1–4–1933 included in the estimation is based on the Randomization test and is 12 (all), 14 (males only) or 13 (females only) years. The window of included births is twice the bandwidth. Based on AIC the order of the local polynomial trend in month–year of birth (from 1–4–1933) is determined. +  $p < 0.05$ , \*\*  $p < 0.01$   
<sup>1</sup>Gamma unobserved heterogeneity



**Fig. 5** Graphical joint test of exclusion restriction and monotonicity of using the cut-off birth of the 1947 reform as an instrument

## 6 Conclusion

We investigate the educational gain in life-expectancy using data for England and Wales from the Health and Lifestyle Survey. For causal identification of the educational gain we propose 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. It defines complier types (always takers, compliers and never takers) for educational attainment, that depend on the policy reform.

A simple Gompertz mortality rate model suggests that staying in school beyond age 15 years significantly increases life-expectancy. However estimates of causal effects obtained from the principal strata method indicate that the total educational gain is not statistically significant. We conducted a range of robustness tests, allowing for

additional covariates, smaller bandwidths around the threshold and ruling out never-takers and did not find substantial changes in the estimated results. This reinforces earlier evidence that shows only 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) of education on mortality. Our empirical application shows that this finding stands up to a rigorous analysis of the mortality hazard based on non-linear duration analysis and in the British educational system.

**Data availability** The data that support the findings of this study are available from the corresponding author upon request.

## Declarations

**Conflict of interest** The authors have no relevant financial or nonfinancial interests to disclose.

**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

## Appendix A: Likelihood

Based on the assumption of a Gompertz<sup>6</sup> functional form of the baseline hazard it is very easy to derive the likelihood function contribution of individual  $i$ .

$$L_i = \prod_{Z_i} \prod_{D_i} \left[ \lambda_i(t_i | Z_i, D_i)^{\delta_i} \times S(t | Z_i, D_i) \right]^{I(Z_i, D_i)}$$

with  $S(t | Z_i, D_i)$  is the survival rate at age  $t$  for an individual with  $Z_i, D_i$ , with

$$S(t | Z = 0, D = 0) = p^n \exp\left(-\int_0^t e^{\alpha_n s + \gamma_n + \beta' X} ds\right) + p^c \exp\left(-\int_0^t e^{\alpha_0 s + \gamma_0 + \beta' X} ds\right) \quad (\text{A.1})$$

$$S(t | Z = 1, D = 0) = p^n \exp\left(-\int_0^t e^{\alpha_n s + \gamma_n + \beta' X} ds\right) \quad (\text{A.2})$$

$$S(t | Z = 0, D = 1) = p^a \exp\left(-\int_0^t e^{\alpha_a s + \gamma_a + \beta' X} ds\right) \quad (\text{A.3})$$

$$S(t | Z = 1, D = 1) = p^a \exp\left(-\int_0^t e^{\alpha_a s + \gamma_n + \beta' X} ds\right)$$

<sup>6</sup> It is straightforward to derive the likelihood for other known baseline hazards.

$$+ p^c \exp\left(-\int_0^t e^{\alpha_1 s + \gamma_1 + \beta' X} ds\right) \tag{A.4}$$

and for complier type  $P = \{a(lways), n(ever), c(omplier)\}$ ,  $p^f = \Pr(P = f)$ .<sup>7</sup>  
 The hazard rate  $\lambda(t|Z = z, D = d, M_1, M_2) = -\partial \log S(t|Z = z, D = d, M_1, M_2)/\partial t$ .

### Appendix B: Derivation of causal quantities

Given:

$$\begin{aligned} & Y(D(1)) - Y(D(0), M_1(D(0)), M_2(D(0))) \\ &= [Y(1) \cdot D(1) + Y(0) \cdot (1 - D(1))] \\ & - [Y(1) \cdot D(0) + Y(0) \cdot (1 - D(0))] \\ &= [Y(1) - Y(0)] \cdot [D(1) - D(0)] \end{aligned}$$

This implies that:

$$\begin{aligned} & E[Y(D(1)) - Y(D(0))] \\ &= E[Y(1) - Y(0) | D(1) - D(0) = 1] \cdot \Pr[D(1) - D(0) = 1] \end{aligned}$$

and the Local Average Treatment Effect is given by:

$$\text{LATE} = E[Y(1) - Y(0) | P = c] \tag{B.1}$$

with  $P$  is the complier type and  $c$  are the compliers.

### Appendix C: Additional tables: AIC of models

See the Table 6.

<sup>7</sup> Note that it is possible to allow the probabilities to depend on observed exogenous variables,  $X_c$ , which may be different from  $X$ .

**Table 6** AIC<sup>C</sup> of estimated models

Order of polynomial <sup>2</sup>	PH			MPH <sup>1</sup>		
	(0)	(1)	(2)	(0)	(1)	(2)
<i>Basic Gompertz</i>						
All individuals	8391.373	8386.9991	8385.3197	8387.5901	8385.9161	8385.4967
Males only	9495.9663	9491.0957	9494.1523	9497.141	9493.0957	9496.1523
Females only	8888.0169	8890.3409	8892.7039	8889.6907	8892.3192	8894.7038
<i>Principal strata model</i>						
All individuals	14150.175	14143.92	14136.675	14143.741	14140.113	14136.107
Males only	8150.6707	8149.0513	8151.8742	8151.4815	8151.0512	8153.856
Females only	8029.1129	8027.8399	8026.9966	8030.2656	8029.5231	8028.9822
<i>Principal strata model with covariates<sup>3</sup></i>						
All individuals	14115.797	14109.535	14102.969	14107.823	14104.317	14100.3
Males only	8148.618	8146.6706	8149.6942	8148.384	8148.1248	8151.1401
Females only	8040.2234	8038.545	8037.822	8040.3806	8039.4194	8039.3741

The bandwidth around 1–4–1933 included in the estimation is based on the Randomization test and is 12 (all), 14 (males only) or 13 (females only) years. The window of included births is twice the bandwidth.

<sup>+</sup>  $p < 0.05$ , <sup>\*\*</sup>  $p < 0.01$

<sup>1</sup>Gamma unobserved heterogeneity

<sup>2</sup>Order of a local polynomial trend in month–year of birth (from 1–4–1933)

<sup>3</sup>Including covariates (region and sex, only for all individuals)

## Appendix D: Estimated coefficients

See the Tables 7, 8, 9, 10, 11 and 12.

**Table 7** Estimated coefficients in Gompertz models

Order of polynomial <sup>2</sup>	PH			MPH <sup>1</sup>		
	(0)	(1)	(2)	(0)	(1)	(2)
Schooling	−0.542** (0.073)	−0.439** (0.080)	−0.403** (0.082)	−0.685** (0.107)	−0.556** (0.117)	−0.497** (0.117)
$\tau_1$		−0.015 (0.011)	−0.097 <sup>+</sup> (0.041)		−0.023 (0.014)	−0.109 <sup>+</sup> (0.049)
$\tau_1^2$		−0.034 <sup>+</sup> (0.016)	−0.025 (0.056)		−0.020 (0.019)	−0.012 (0.061)
$\tau_2$			−0.007 <sup>+</sup> (0.003)			−0.008 <sup>+</sup> (0.004)
$\tau_2^2$			0.001			0.000

**Table 7** continued

Order of polynomial <sup>2</sup>	PH			MPH <sup>1</sup>		
	(0)	(1)	(2)	(0)	(1)	(2)
Constant	-9.123** (0.299)	-8.596** (0.346)	-8.741** (0.356)	-10.043** (0.520)	-9.432** (0.614)	-9.418** (0.635)
$\alpha$	0.079** (0.004)	0.070** (0.005)	0.071** (0.005)	0.097** (0.009)	0.086** (0.010)	0.083** (0.011)
$\ln(\sigma_V)$				-0.276 (0.448)	-0.540 (0.593)	-0.767 (0.760)

The bandwidth around 1–4–1933 included in the estimation is based on the Randomization test and is 12 (all), 14 (males only) or 13 (females only) years. The window of included births is twice the bandwidth. Local trend in month–year of birth (from 1–4–1933);  $\tau_1$ , before and  $\tau_2$ , after. <sup>+</sup> $p < 0.05$ , <sup>\*\*</sup> $p < 0.01$

<sup>1</sup> Gamma unobserved heterogeneity

<sup>2</sup> Order of a local polynomial trend in month–year of birth (from 1–4–1933)

**Table 8** Estimated coefficients in Gompertz models, males

Order of polynomial <sup>2</sup>	PH			MPH		
	(0)	(1)	(2)	(0)	(1)	(2)
Schooling	-0.521** (0.092)	-0.391** (0.100)	-0.371** (0.102)	-0.580** (0.121)	-0.391** (0.100)	-0.371** (0.102)
$\tau_1$		-0.022 (0.012)	-0.036 (0.044)		-0.022 (0.012)	-0.036 (0.044)
$\tau_1^2$		-0.033 (0.018)	-0.069 (0.064)		-0.033 (0.018)	-0.069 (0.064)
$\tau_2$			-0.001 (0.003)			-0.001 (0.003)
$\tau_2^2$			0.003 (0.005)			0.003 (0.005)
Constant	-9.247** (0.364)	-8.543** (0.432)	-8.544** (0.441)	-9.648** (0.591)	-8.543** (0.432)	-8.544** (0.441)
$\alpha$	0.084** (0.005)	0.072** (0.006)	0.072** (0.006)	0.092** (0.010)	0.072** (0.006)	0.072** (0.006)
$\ln(\sigma_V)$				-1.380 (1.148)	-12.789 (555.815)	-13.151 (708.510)

<sup>1</sup> Gamma unobserved heterogeneity

<sup>2</sup> Order of a local polynomial trend in month–year of birth (from 1–4–1933)

The bandwidth around 1–4–1933 included in the estimation is based on the Randomization test and is 12 (all), 14 (males only) or 13 (females only) years. The window of included births is twice the bandwidth. local trend in month–year of birth (from 1–4–1933);  $\tau_1$ , before and  $\tau_2$ , after. <sup>+</sup> $p < 0.05$ ; <sup>\*\*</sup> $p < 0.01$

**Table 9** Estimated coefficients in Gompertz models, females

order of polynomial <sup>2</sup>	PH			MPH <sup>1</sup>		
	(0)	(1)	(2)	(0)	(1)	(2)
Schooling	-0.452** (0.100)	-0.399** (0.109)	-0.384** (0.111)	-0.479** (0.116)	-0.408** (0.124)	-0.385** (0.125)
$\tau_1$		-0.003 (0.014)	-0.066 (0.052)		-0.004 (0.015)	-0.066 (0.052)
$\tau_1^2$		-0.025 (0.021)	0.032 (0.073)		-0.024 (0.023)	0.032 (0.074)
$\tau_2$			-0.005 (0.004)			-0.005 (0.004)
$\tau_2^2$			-0.004 (0.006)			-0.004 (0.006)
Constant	-10.010** (0.411)	-9.672** (0.492)	-9.815** (0.506)	-10.322** (0.699)	-9.772** (0.840)	-9.825** (0.818)
$\alpha$	0.088** (0.005)	0.084** (0.007)	0.084** (0.007)	0.094** (0.012)	0.085** (0.014)	0.084** (0.013)
$\ln(\sigma_V)$				-1.376 (1.797)	-2.630 (6.819)	-4.962 (65.552)

The bandwidth around 1–4–1933 included in the estimation is based on the Randomization test and is 12 (all), 14 (males only) or 13 (females only) years. The window of included births is twice the bandwidth. local trend in month–year of birth (from 1–4–1933);  $\tau_1$ , before and  $\tau_2$ , after. <sup>+</sup>  $p < 0.05$ , <sup>\*\*</sup>  $p < 0.01$

<sup>1</sup>Gamma unobserved heterogeneity

<sup>2</sup>Order of a local polynomial trend in month–year of birth (from 1–4–1933)

**Table 10** Estimated coefficients in principal strata models

Order of polynomial <sup>2</sup>	PH			MPH <sup>1</sup>		
	(0)	(1)	(2)	(0)	(1)	(2)
$\gamma_n$	-9.225** (0.349)	-8.544** (0.396)	-8.503** (0.387)	-10.497** (0.605)	-9.758** (0.681)	-9.420** (0.714)
$\gamma_{a_1}$	-9.361** (0.346)	-8.914** (0.373)	-9.182** (0.395)	-10.716** (0.636)	-10.210** (0.700)	-10.116** (0.733)
$\gamma_{a_2}$	-9.736** (0.308)	-9.169** (0.355)	-9.361** (0.373)	-11.153** (0.616)	-10.511** (0.693)	-10.314** (0.722)
$\gamma_{a_3}$	-9.873** (0.334)	-9.301** (0.378)	-9.472** (0.394)	-11.384** (0.665)	-10.744** (0.746)	-10.499** (0.775)
$\gamma_0$	-9.064** (0.329)	-8.659** (0.355)	-8.982** (0.387)	-10.262** (0.572)	-9.815** (0.633)	-9.795** (0.661)
$\gamma_1$	-9.580**	-8.818**	-8.806**	-10.890**	-10.071**	-9.721**

**Table 10** continued

Order of polynomial <sup>2</sup>	PH			MPH <sup>1</sup>		
	(0)	(1)	(2)	(0)	(1)	(2)
	(0.281)	(0.364)	(0.375)	(0.568)	(0.672)	(0.705)
$\tau_1$		-0.020 (0.012)	-0.137** (0.044)		-0.034 <sup>+</sup> (0.017)	-0.151** (0.054)
$\tau_1^2$		-0.043 <sup>+</sup> (0.017)	-0.093 (0.061)		-0.026 (0.021)	-0.076 (0.067)
$\tau_2$			-0.010** (0.004)			-0.011 <sup>+</sup> (0.004)
$\tau_2^2$			0.006 (0.005)			0.005 (0.006)
Logit coefficients of $\pi = \Pr(P = p)$ <sup>3</sup>						
$\beta_1$	0.162 (0.129)	0.162 (0.129)	0.160 (0.129)	0.162 (0.129)	0.162 (0.129)	0.162 (0.129)
$\beta_2$	1.190** (0.107)	1.190** (0.107)	1.188** (0.107)	1.190** (0.107)	1.191** (0.107)	1.190** (0.107)
$\beta_3$	0.114 (0.120)	0.115 (0.120)	0.112 (0.120)	0.115 (0.120)	0.115 (0.120)	0.115 (0.120)
$\beta_c$	1.997** (0.111)	1.997** (0.111)	1.995** (0.111)	1.997** (0.111)	1.998** (0.111)	1.997** (0.111)
$\alpha$	0.079** (0.004)	0.071** (0.005)	0.071** (0.005)	0.102** (0.010)	0.091** (0.010)	0.086** (0.011)
$\ln(\sigma_V)$				-0.025 (0.189)	-0.106 (0.220)	-0.269 (0.319)

The bandwidth around 1–4–1933 included in the estimation is based on the Randomization test and is 12 (all), 14 (males only) or 13 (females only) years. The window of included births is twice the bandwidth. local trend in month–year of birth (from 1–4–1933);  $\tau_1$ , before and  $\tau_2$ , after. <sup>+</sup>  $p < 0.05$ , **\*\***  $p < 0.01$

<sup>1</sup>Gamma unobserved heterogeneity

<sup>2</sup>Order of a local polynomial trend in month–year of birth (from 1–4–1933)

<sup>3</sup> $\Pr(P = p) = e^{\beta_p I(P=p)} / (1 + e^{\beta_c I(P=c)} + e^{\beta_1 I(P=a_1)} + e^{\beta_2 I(P=a_2)} + e^{\beta_3 I(P=a_3)})$  with  $\beta_n = 0$

**Table 11** Estimated coefficients in principal strata models, males

Order of polynomial <sup>2</sup>	PH			MPH <sup>1</sup>		
	(0)	(1)	(2)	(0)	(1)	(2)
$\gamma_n$	-9.704** (0.492)	-8.985** (0.561)	-8.883** (0.568)	-10.259** (0.731)	-8.984** (0.308)	-8.978** (0.916)
$\gamma_{a_1}$	-9.218** (0.422)	-8.736** (0.465)	-8.736** (0.473)	-9.780** (0.696)	-8.736** (0.210)	-8.830** (0.852)
$\gamma_{a_2}$	-9.736** (0.377)	-9.144** (0.450)	-9.108** (0.458)	-10.349** (0.699)	-9.143** (0.121)	-9.210** (0.892)
$\gamma_{a_3}$	-9.763** (0.403)	-9.146** (0.477)	-9.105** (0.483)	-10.401** (0.735)	-9.145** (0.212)	-9.212** (0.935)
$\gamma_0$	-8.958** (0.392)	-8.514** (0.434)	-8.526** (0.445)	-9.470** (0.635)	-8.514** (0.091)	-8.613** (0.791)
$\gamma_1$	-9.624** (0.339)	-8.845** (0.467)	-8.720** (0.483)	-10.189** (0.640)	-8.845** (0.142)	-8.818** (0.881)
$\tau_1$		-0.019 (0.013)	-0.034 (0.047)		-0.019 <sup>+</sup> (0.008)	-0.036 (0.049)
$\tau_1^2$		-0.034 (0.020)	-0.090 (0.069)		-0.034 (0.020)	-0.087 (0.072)
$\tau_2$			-0.001 (0.003)			-0.001 (0.003)
$\tau_2^2$			0.005 (0.005)			0.005 (0.005)
Logit coefficients of $\pi = \Pr(P = p)$ <sup>3</sup>						
$\beta_1$	0.004 (0.181)	0.009 (0.181)	0.010 (0.181)	0.007 (0.181)	0.009 (0.181)	0.010 (0.181)
$\beta_2$	1.103** (0.148)	1.109** (0.148)	1.109** (0.148)	1.106** (0.148)	1.109** (0.148)	1.109** (0.148)
$\beta_3$	0.086 (0.165)	0.091 (0.165)	0.092 (0.165)	0.088 (0.165)	0.091 (0.165)	0.092 (0.165)
$\beta_c$	1.989** (0.153)	1.996** (0.153)	1.996** (0.153)	1.992** (0.153)	1.996** (0.153)	1.996** (0.153)
$\alpha$	0.082** (0.005)	0.073** (0.006)	0.073** (0.006)	0.092** (0.011)	0.073** (0.001)	0.074** (0.013)
$\ln(\sigma_V)$				-0.589 (0.481)	-11.442 (41.169)	-1.596 (3.754)

The bandwidth around 1–4–1933 included in the estimation is based on the Randomization test and is 12 (all), 14 (males only) or 13 (females only) years. The window of included births is twice the bandwidth. local trend in month–year of birth (from 1–4–1933);  $\tau_1$ , before and  $\tau_2$ , after. <sup>+</sup>  $p < 0.05$ , <sup>\*\*</sup>  $p < 0.01$

<sup>1</sup> Gamma unobserved heterogeneity

<sup>2</sup> Order of a local polynomial trend in month–year of birth (from 1–4–1933)

<sup>3</sup>  $\Pr(P = p) = e^{\beta_p I(P=p)} / (1 + e^{\beta_c I(P=c)} + e^{\beta_1 I(P=a_1)} + e^{\beta_2 I(P=a_2)} + e^{\beta_3 I(P=a_3)})$  with  $\beta_n = 0$

**Table 12** Estimated coefficients in principal strata models, females

Order of polynomial <sup>2</sup>	PH			MPH <sup>1</sup>		
	(0)	(1)	(2)	(0)	(1)	(2)
$\gamma_n$	-9.935** (0.445)	-9.294** (0.515)	-9.409** (0.520)	-10.490** (0.778)	-9.690** (0.889)	-9.498** (0.910)
$\gamma_{a_1}$	-10.647** (0.498)	-10.175** (0.540)	-10.460** (0.569)	-11.266** (0.871)	-10.624** (0.985)	-10.559** (1.009)
$\gamma_{a_2}$	-10.832** (0.446)	-10.259** (0.514)	-10.498** (0.540)	-11.442** (0.823)	-10.695** (0.943)	-10.593** (0.963)
$\gamma_{a_3}$	-11.261** (0.495)	-10.694** (0.556)	-10.930** (0.581)	-11.925** (0.904)	-11.173** (1.032)	-11.035** (1.056)
$\gamma_0$	-10.432** (0.485)	-9.987** (0.525)	-10.314** (0.562)	-10.996** (0.811)	-10.400** (0.920)	-10.403** (0.938)
$\gamma_1$	-10.583** (0.401)	-9.826** (0.517)	-9.906** (0.535)	-11.151** (0.760)	-10.231** (0.900)	-9.996** (0.927)
$\tau_1$		-0.016 (0.015)	-0.123 <sup>+</sup> (0.056)		-0.020 (0.018)	-0.124 <sup>+</sup> (0.058)
$\tau_1^2$		-0.043 (0.023)	-0.050 (0.078)		-0.039 (0.025)	-0.049 (0.079)
$\tau_2$			-0.008 <sup>+</sup> (0.004)			-0.008 <sup>+</sup> (0.004)
$\tau_2^2$			0.002 (0.007)			0.002 (0.007)
Logit coefficients of $\pi = \Pr(P = p)$ <sup>3</sup>						
$\beta_1$	0.317 (0.168)	0.319 (0.168)	0.322 (0.168)	0.318 (0.168)	0.321 (0.168)	0.323 (0.168)
$\beta_2$	1.347** (0.141)	1.350** (0.141)	1.353** (0.141)	1.349** (0.141)	1.351** (0.141)	1.353** (0.141)
$\beta_3$	0.187 (0.159)	0.189 (0.159)	0.193 (0.159)	0.188 (0.159)	0.191 (0.159)	0.193 (0.159)
$\beta_c$	2.057** (0.147)	2.060** (0.147)	2.063** (0.147)	2.059** (0.147)	2.061** (0.147)	2.064** (0.147)
$\alpha$	0.093** (0.006)	0.085** (0.007)	0.086** (0.007)	0.103** (0.013)	0.092** (0.014)	0.088** (0.015)
$\ln(\sigma_V)$				-0.415 (0.567)	-0.614 (0.903)	-1.359 (4.171)

The bandwidth around 1–4–1933 included in the estimation is based on the Randomization test and is 12 (all), 14 (males only) or 13 (females only) years. The window of included births is twice the bandwidth. local trend in month–year of birth (from 1–4–1933);  $\tau_1$ , before and  $\tau_2$ , after. <sup>+</sup>  $p < 0.05$ , **\*\***  $p < 0.01$

<sup>1</sup>Gamma unobserved heterogeneity

<sup>2</sup>Order of a local polynomial trend in month–year of birth (from 1–4–1933)

<sup>3</sup> $\Pr(P = p) = e^{\beta_p I(P=p)} / (1 + e^{\beta_c I(P=c)} + e^{\beta_1 I(P=a_1)} + e^{\beta_2 I(P=a_2)} + e^{\beta_3 I(P=a_3)})$  with  $\beta_n = 0$

## References

- Albouy V, Lequien L (2009) Does compulsory education lower mortality? *J Health Econ* 28(1):155–168
- Angrist JD, Imbens GW, Rubin DB (1996) Identification of causal effects using instrumental variables (with discussion). *J Am Stat Assoc* 91:444–472
- Basu A, Jones AM, Rosa Dias P (2018) Heterogeneity in the impact of type of schooling on adult health and lifestyle. *J Health Econ* 57:1–14
- Bijwaard GE (2009) Instrumental variable estimation for duration data. In: Engelhardt H, Kohler H-P, Führenkrantz-Prskawetz A (eds) *Causal analysis in population studies: concepts, methods, applications*. Springer, Berlin, pp 111–148
- Bijwaard GE, Jones AM (2019) An IPW estimator for mediation effects in hazard models: with an application to schooling, cognitive ability and mortality. *Empir Econ* 57(1):129–175
- Bijwaard GE, van Kippersluis H, Veenman J (2015) Education and health: the role of cognitive ability. *J Health Econ* 42:29–43
- Bijwaard GE, van Poppel F, Ekamper P, Lumey LH (2015) Gains in life expectancy associated with higher education in men. *PLoS ONE* 10:e0141200
- Bijwaard GE, Myrskylä M, Tynelius P, Rasmussen F (2017) Educational gains in cause-specific mortality: accounting for cognitive ability and family-level confounders using propensity score weighting. *Soc Sci Med* 184:49–56
- Bijwaard GE, Myrskylä M, Tynelius P (2019) Education, cognitive ability and cause-specific mortality: a structural approach. *Popul Stud* 73(2):217–232
- Calonico S, Cattaneo MD, Farrell MH, Titiunik R (2019) Regression discontinuity designs using covariates. *Rev Econ Stat* 101(3):442–451
- Cattaneo MD, Titiunik R (2022) Regression discontinuity designs. *Annu Rev Econ* 14:821–851
- Cattaneo MD, Frandsen BR, Titiunik R (2015) Randomization inference in the regression discontinuity design: an application to party advantages in the US Senate. *J Causal Inference* 3(1):1–24
- Cattaneo MD, Titiunik R, Vazquez-Bare G (2017) Comparing inference approaches for RD designs: a reexamination of the effect of Head Start on child mortality. *J Policy Anal Manage* 36(3):643–681
- Clark D, Royer H (2013) The effect of education on adult mortality and health: evidence from Britain. *Am Econ Rev* 103(6):2087–2120
- Cuzick J, Sasieni P, Myles J, Tyrer J (2007) Estimating the effect of treatment in a proportional hazards model in the presence of non-compliance and contamination. *J R Stat Soc B* 69:565–588
- Fletcher JM (2015) New evidence of the effects of education on health in the US: compulsory schooling laws revisited. *Soc Sci Med* 127:101–107
- Frangakis CE, Rubin DB (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 CE, Rubin DB (2002) Principal stratification in causal inference. *Biometrics* 58:21–29
- Gavrilov LA, Gavrilova NS (1991) *The biology of life span: a quantitative approach*. Harwood Academic Publisher, New York
- Grossman M (2006) Education and non-market outcomes. In: Hanushek E, Welch F (eds) *Handbook of the economics of education*, vol 1, Chapter 10. Elsevier, Amsterdam, pp 577–633
- Grossman M (1972) On the concept of health capital and the demand for health. *J Polit Econ* 80(2):223–255
- Imbens GW (2016) Regression discontinuity designs in the econometrics literature. *Obs Stud* 3(2):147–155
- Imbens GW, Lemieux T (2008) Regression discontinuity designs: a guide to practice. *J Econom* 142:615–635
- Imbens GW, Rubin DB (1997) Bayesian inference for causal effects in randomized experiments with noncompliance. *Ann Stat* 25:305–327
- Johnston DW, Lordan G, Shields MA, Suziedelyte A (2015) Education and health knowledge: evidence from UK compulsory schooling reform. *Soc Sci Med* 127:92–100
- Jones AM, Rice N, Rosa Dias P (2011) Long-term effects of school quality on health and lifestyle: evidence from comprehensive schooling reforms in England. *J Hum Cap* 5:342–376
- Jürges H, Kruk E, Reinhold S (2013) The effect of compulsory schooling on health: evidence from biomarkers. *J Popul Econ* 26:645–672
- Kitagawa T (2015) A test for instrument validity. *Econometrica* 83:2043–2063
- Lancaster T (1990) *The econometric analysis of transition data*. Cambridge University Press, Cambridge

- Lee DS (2008) Randomized experiments from non-random selection in US House elections. *J Econom* 142(2):675–697
- Lee DS, Lemieux T (2010) Regression discontinuity designs in economics. *J Econ Lit* 48:281–355
- Lenart A (2014) The moments of the Gompertz distribution and the maximum likelihood of its parameters. *Scand Actuar J Inst Actuar* 3:255–277
- Li F, Mattei A, Mealli F (2015) Evaluating the causal effect of university grants on student dropout: evidence from a regression discontinuity design using principal stratification. *Ann Appl Stat* 9(4):1906–1931
- Lin H, Li Y, Jiang L, Li G (2014) A semiparametric linear transformation model to estimate causal effects for survival data. *Can J Stat* 42:18–35
- Lleras-Muney A (2005) The relationship between education and adult mortality in the United States. *Rev Econ Stud* 72:189–221
- Mazumder B (2008) Does education improve health: a reexamination of the evidence from compulsory schooling laws. *Fed Reserve Bank Chicago Econ Perspect* 33(2):2–16
- Mazumder B (2012) The effects of education on health and mortality. *Nord Econ Policy Rev* 1:261–301
- McCartney G, Collins C, Mackenzie M (2013) What (or who) causes health inequalities: Theories, evidence and implications? *Health Policy* 113:221–227
- Mealli F, Mattei A (2012) A refreshing account of principal stratification. *Int J Biostat* 8:1–19
- Meghir C, Palme M, Simeonova E (2018) Education, cognition and health: evidence from a social experiment. *Am Econ J Appl Econ* 10(2):234–256
- Missov TI (2013) Gamma-Gompertz life expectancy at birth. *Demogr Res* 28:259–270
- Mourifié I, Wan Y (2017) Testing local average treatment effect assumptions. *Rev Econ Stat* 99(2):305–313
- Terza J, Basu A, Rathouz P (2008) Two-stage instrumental variable methods: addressing endogeneity in health econometric modeling. *J Health Econ* 27:531–543
- Van den Berg GJ (2001) Duration models: specification, identification, and multiple duration. In: Heckman J, Leamer E (eds) *Handbook of econometrics*, vol 5, Chapter 55. North-Holland, Amsterdam, pp 3381–3460
- Van Kippersluis H, O'Donnell O, van Doorslaer E (2011) Long run returns to education: Does schooling lead to an extended old age? *J Hum Resour* 46(4):695–721
- Wan F, Small D, Bekelman JE, Mitra N (2015) Bias in estimating the causal hazard ration when using two-stage instrumental variable methods. *Stat Med* 34:2235–2265
- Wan F, Small D, Mitra N (2018) A general approach to evaluating the bias of 2-stage instrumental variable estimators. *Stat Med* 37:1997–2015
- Zhang JL, Rubin DB, Mealli F (2009) Likelihood-based analysis of causal effects of job-training programs using principal stratification. *J Am Stat Assoc* 104(485):166–176

**Publisher's Note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.