TI 2013-0 44/5
Tinbergen Institute Discussion Paper

Education and Health: The Role of Cognitive Ability

Govert Bijwaard[1]
Hans van Kippersluis[2,4]
Justus Veenman[2,3]

[1] *NIDI, The Hague, and IZA, Bonn;*
[2] *Erasmus School of Economics, Erasmus University Rotterdam;*
[3] *ERCOMER, Utrecht/Rotterdam;*
[4] *Tinbergen Institute, Amsterdam/Rotterdam, and NETSPAR, Tilburg.*

Tinbergen Institute Amsterdam
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 1600

Tinbergen Institute Rotterdam
Burg. Oudlaan 50
3062 PA Rotterdam
The Netherlands
Tel.: +31(0)10 408 8900
Fax: +31(0)10 408 9031

Duisenberg school of finance
Gustav Mahlerplein 117
1082 MS Amsterdam
The Netherlands
Tel.: +31(0)20 525 8579

Education and Health: the Role of Cognitive Ability*

Govert Bijwaard[†] Hans van Kippersluis[‡] Justus Veenman[§]

March 15, 2013

Abstract

We aim to disentangle the relative contributions of (i) cognitive ability, and (ii) education on health and mortality using a structural equation model suggested by Conti et al. (2010). We extend their model by allowing for a duration dependent variable, and an ordinal educational variable. Data come from a Dutch cohort born around 1940, including detailed measures of cognitive ability and family background at age 12. The data are subsequently linked to the mortality register 1995-2011, such that we observe mortality between ages 55 and 75. The results suggest that the treatment effect of education (i.e. the effect of entering secondary school as opposed to leaving school after primary education) is positive and amounts to a 4 years gain in life expectancy, on average. Decomposition results suggest that the raw survival differences between educational groups are about equally split between a 'treatment effect' of education, and a 'selection effect' on basis of cognitive ability and family background.

Keywords: Education, Cognitive Ability, Mortality, Structural Equation Model, Duration model
JEL Codes : C41, I14, I24

*Van Kippersluis gratefully acknowledges funding from the National Institute on Aging (NIA) under grant R01AG037398 and from NETSPAR under the project "Income and health, work and care across the life cycle II". The authors acknowledge access to linked data resources (DO 1995-2011) by Statistics Netherlands (CBS). We thank Mars Cramer and Mirjam van Praag for help in accessing the original data, and are grateful to seminar participants from the Empirical Health Economics conference in Munich, Multistate Event History Analysis in Hangzhou, and Erasmus University Rotterdam, for their comments.

[†] Netherlands Interdisciplinary Demographic Institute (NIDI), The Hague, the Netherlands and IZA, Bonn

[‡] Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands, Tinbergen Institute, Rotterdam, The Netherlands, and Netspar, Tilburg, The Netherlands

[§] Erasmus School of Economics, Erasmus University Rotterdam, The Netherlands, and ERCOMER, Utrecht/Rotterdam

1 Introduction

Disparities in health and life expectancy across educational groups are striking and pervasive, and are considered one of the most compelling and well established facts in social science research (Mazumder, 2012). Even in an egalitarian country such as the Netherlands, with a very accessible health care system, the difference in life expectancy between the university educated and those who finished only primary school is 6 to 7 years (CBS, 2008). It is commonly assumed that a large part of this association derives from the causal effect of education on health outcomes. An abundant list of possible mechanisms was proposed, among which occupational demands, health behavior, and the ability to process information are the most commonly mentioned (Ross and Wu, 1995; Cutler and Lleras-Muney, 2008).

Yet, the association between education and health could also stem from (i) 'reverse causality', in which childhood ill-health constrains educational attainment (Behrman and Rosenzweig, 2004; Case et al. 2005), and (ii) confounding 'third factors' such as ability, parental background and time preference that influence both education and health outcomes (Fuchs, 1982; Auld and Sidhu, 2005; Deary, 2008).

Studies based on natural experiments in education, such as changes in compulsory schooling laws, overcome the difficulty of separating the direct causal effect of education from third factor effects. The estimates based on these studies point towards a small effect (Lleras-Muney, 2005; Oreopoulos, 2006; Van Kippersluis et al. 2011; Meghir et al. 2012), or even insignificant effect of education on health and mortality (Arendt, 2005; Albouy and Lequien, 2008; Mazumder, 2008; Braakmann, 2011; Gathmann et al. 2012; Clark and Royer, 2013). This suggests that confounding factors may well play an important role in shaping the strong association between education and health.

Surprisingly little research in economics has investigated the contribution of early childhood abilities and childhood social background in shaping the association between education and health.[1] Some recent economic studies report associations between childhood cognitive and non-cognitive abilities, and health outcomes at ages 30-40 using the British Cohort Study (Murasko, 2007), the U.K. National Child Development Study (Carneiro et al. 2007), the U.S. National Longitudinal Study of Youth 1979 (Elias, 2004; Auld and Sidhu, 2005; Kaestner and Collison, 2011), or the Dutch "Brabant data" (Cramer, 2012). It is established that cognitive ability and some non-cognitive factors such as self-esteem are associated with health outcomes. Nonetheless, hardly anything is known about (i) the relative impact of education and childhood abilities on health outcomes, and in turn (ii) how much of the association between education and health is explained by these cognitive and non-cognitive abilities.

A notable contribution to the literature is a recent series of papers by Conti and Heckman (2010), and Conti et al. (2010; 2011) who, using the British Cohort Study, estimate a structural equation model in which the interdependence between education,

[1] See Gottfredson (2004) for an excellent overview of the epidemiological literature.

health, and two latent factors capturing cognitive and non-cognitive skills is explicitly modeled. The authors show that for most health outcomes around half of the association between education and health is driven by cognitive and non-cognitive abilities and early childhood social background. The other half is interpreted as the causal effect of education on health.

While the series of papers by Conti and co-authors provided a significant contribution to the literature, there are two notable limitations. First, the health outcomes are measured at age 30, an age at which health differences by education may not have fully materialized. In fact, disparities in health and mortality seem to peak around middle-age (Cutler and Lleras-Muney, 2008). Secondly, the health measures are all self-reported, which may bias the estimates since education and abilities are related to subjective health perceptions (Bago d'Uva et al. 2008).

In this paper we aim to disentangle the effects of education and cognitive ability on health outcomes by using and extending the structural equation model of Conti and Heckman (2010). We will use the so-called "Brabant data" - a representative cohort of primary school sixth graders in the Dutch province of Noord-Brabant - that has detailed information on cognitive ability and social background measured back in 1952. Three follow-up surveys in 1957, 1983 and 1993 contain information on education, employment, and self-reported health. We have linked these data to the mortality register 1995-2011, such that the impact on mortality can be analyzed.

The contribution of this paper is threefold. First, we study the relative impact of cognitive ability and education on mortality, as an *objective* health indicator. The second contribution is that, in contrast to existing studies that measure health outcomes at ages 30-40, we observe mortality during ages 55-75 - an age-span which has been shown to exhibit the largest relative health disparities (Cutler and Lleras-Muney, 2008). Finally, we extend the structural equation model by Conti et al. (2010) by allowing for (i) a duration dependent variable (mortality) and (ii) an ordinal independent variable, as education typically is not measured as a binary indicator.

The results suggest that the treatment effect of education (i.e. the effect of entering secondary school as opposed to leaving school after primary education) is positive and amounts to a 4 years gain in life expectancy, on average. Therefore, even after controlling for cognitive ability, family social class, and a range of other background variables, education remains very important in determining mortality. For most ages, the treatment effect of education explains around half of the raw differences in mortality across educational groups, in line with the findings of Conti et al. (2010).

This paper is structured as follows. Section 2 presents the Brabant data including the available register data from Statistics Netherlands, section 3 presents the structural equation model that we will use to disentangle the relative contributions of cognitive ability and education on health outcomes. Section 4 presents the results and section 5 discusses them.

2 Data and descriptive statistics

The data are from a Dutch cohort born between 1937 and 1941. Very detailed information about individual intelligence, social background and school achievement is available for 5,823 individuals. The survey was held in the spring and summer of 1952 among pupils of the sixth (last) grade of primary schools in the Dutch province of Noord-Brabant, and hence is referred to as the "Brabant data". One-fourth of the province population was sampled; mainly by including every fourth child from the schools' list of pupils.[2] Hartog (1989) investigated the data and found no reason to doubt randomness. A selective dropout of pupils before participating in the data collection does not exist, as primary school was compulsory and reinforcement of school attendance was strict (Dronkers, 2002).

Follow-up surveys took place in 1957, 1983 and 1993.[3] In 1957 only a sub-sample - those who scored above-average on six tests - of the original cohort was interviewed about the school careers between 1952 and 1957 to particularly investigate school career choices of the most intelligent half of the cohort. In 1983 and 1993 attempts were made to trace all initial respondents of the Brabant-cohort to investigate labour market behavior, with overall response rates of around 45 percent. The sample is reduced to 2,998 individuals who have measurements in 1952 and in either 1983 or 1993, or both.[4]

The Brabant data are subsequently linked to administrative records from Statistics Netherlands. The basis for this linkage is identifying information on ZIP code, date of birth, and sex, provided in 1993 by Dutch municipalities. The administrative records are available since 1995. Because of the two-year discrepancy only 86 percent of the 2,998 individuals could be traced in the municipality register in 1995, leaving us with a working sample of 2,579 individuals. Administrative records include the mortality register and the municipality register for the years 1995-2011 inclusive. The mortality register is used to identify drop out due to death in the follow-up period. Demographics (age, sex, and nationality) are obtained from the municipality register.

Dependent variables: Our outcome variable is *Mortality*, which is identified from the mortality register in the period 1995-2011. Given that most pupils are born around 1940, this implies that we follow mortality from age 55 until 75.[5] In our sample, 409 individuals, or 16 percent, died during the period 1995-2011. Close to 50 percent died from cancer, 25 percent from cardiovascular diseases, and 8 percent from respiratory diseases such as

[2] Some schools had school years beginning in April rather than in September. For these schools, half the pupils of half the schools were included in the sample, which yielded 369 observations on a total of 5,823 (Hartog, 1989).

[3] Mathijsen and Sonnemans (1958), Hartog and Pfann (1985), Van Praag (1992), and Hartog, Jonker and Pfann (2002). The complete questionnaire is included in Van Praag (1992) 'Brabantse zesdeklassers, 1952-2010'.

[4] In section 4.3 it is verified that selective attrition does not affect our results.

[5] Of the Dutch population 1940 cohort, only 6.8 percent died between the ages of 12 and 55 – Human Mortality Database, University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany). Available at www.mortality.org or www.humanmortality.de (data downloaded on July 30, 2012).

COPD and pneumonia. External causes such as accidents comprise only two percent, as do mental disorders (e.g. dementia), diseases of the digestive system (e.g. liver cirrhosis) and diseases of the nervous system (e.g. Parkinson).

Independent variables: Our main independent variable of interest is *Education*, here defined as the highest level of education attended, in three categories:

1. *Lower Education*, including those who attended at most (extended)[6] primary school.

2. *Lower Vocational Education*, including those who attended at most lower vocational education ("LBO") such as the lower agricultural school or lower polytechnic schools.

3. *At least General Secondary School*, including those who attended at least lower general secondary school ("(M)ULO" or "MAVO"), higher general secondary school ("HBS", "HAVO", "VWO". "Atheneum", or "Gymnasium"), and Higher Vocational Education or University.

Education is retrieved mainly from the 1983 and 1993-survey variables on the highest level of education attended. The maximum of the two defines *Education*, and where missing we update our educational variable with information from the 1957 survey for a sub-set of the sample.

Table 1 presents descriptive statistics and shows that 14 percent did not continue school after primary school forming the *Lower Education* category, 35 percent only attended *Lower Vocational Education*, and the other 51 percent attended at least *General Secondary School*. Figures 1 and 2 show the Kaplan-Meier survival curves separately for the three education categories, and for a binary indicator of education with threshold at *Lower Education*. It is clear that the largest survival differences are between those with only primary school and those above primary school, and that the difference grows with age to around ten percentage points near age 75.

Our second independent variable is *Cognitive Ability*. In the Brabant data there are two separate measurements for cognitive ability, both measured at age 12: (i) the Raven Progressive Matrices Test, and (ii) a Vocabulary test (picking synonyms).[7] The timing of the intelligence test at age 12 avoids possible reverse causality from education to intelligence (Deary and Johnson, 2010) and allows measuring the clean impact of childhood cognitive ability.

[6] At the time, pupils had to stay in school for at least 8 years, or until they reached the age of 14. Since regular primary school only consisted of 6 grades, some schools offered an additional 2-year extended primary school ("vglo").

[7] The data also contain the so-called LO-IV test, which consists of six sub-tests: regularities in series of numbers, analogies in figures, analogies in words, and similarities between concepts (equal, not-equal, cause). Since the quality of this test has been questioned (Hartog et al. 2002, p. 5) we will not use this test in our analyses. There is also information on grades for specific courses (Dutch language, mathematics (arithmetics), history, physics, geography, health sciences, and traffic), but since these are not clean measures of cognitive ability and are relative to others in one's classroom, we choose not to use these grades.

The IQ p.m. ('progressive matrices') test focuses on mathematical ability and is a replication of the British Progressive Matrices test, designed by Raven (1958). It is considered to be a 'pure' measurement of problem solving abilities, as it does not require any linguistic or general knowledge (Dronkers, 2002). Hence, the Raven test is supposed to measure cognitive abilities or analytic intelligence (Carpenter et al. 1990). In this sense, the test can be compared to Spearman's g test (1927). The term g refers to the determinants of the common variance within intelligence tests, being the core issue of intelligence measurement (Carpenter et al. 1990).

Table 1 shows that the ability test designed by Raven has an average of 102, with standard deviation of 13 while the vocabulary test is 101, on average, with standard deviation 13. The correlation between the Raven test and the vocabulary test is 0.38. This suggests that while there seems to be some overlap between what the two tests measure, they additionally measure some idiosyncratic part of cognitive ability. Therefore, we will use both measurements to build a comprehensive latent factor of cognitive ability. In a robustness check we solely use the Raven test to see whether the results differ.

Control variables: Apart from a fairly standard set of demographic control variables such *Age*, whether *Male*, and *Birth Rank*, we also have information about the social and school environment of the individuals. Most of these variables are reported by the School Principal. *Family Social Class* is measured in three categories from lowest to highest depending on father's occupation.[8] We additionally know whether the child had to work in the parent's farm or company, defining the binary indicator *Child Works*.

Available information regarding the school includes *School Type*, being either Roman Catholic, Protestant, or other (including Montessori, Dalton, and Public schools), and we know the *Number of Teachers*, which we divide into three categories (1-4 teachers, 5-8 teachers, and 9-12 teachers). *Repeat* defines the number of classes that children had to repeat. Further, we know the *Teacher's Advice* regarding further education of the child, divided into four categories: (i) continue in primary school until compulsory schooling age, (ii) lower vocational education, (iii) lower secondary education, or (iv) higher secondary education. Finally, we know the *Preference of the Parents* concerning the education of the pupil, divided into (i) work in family company, (ii) paid work without vocational education, (iii) paid work with vocational education, (iv) only vocational education, or (v) general secondary education.

We have no information about childhood health status, which prevents us from investigating the possibility of reverse causality from health to education in our sample. The sample is comprised of pupils that made it to the final grade of primary school.

[8] We classify lower administrative workers, agricultural workers, industrial workers, other lower workers, and disabled into the Lowest Social Class. If the School Principal considered the family "antisocial", the family is also classified into the Lowest Social Class. Intermediary personnel, self-employed farmers, self-employed craftsmen, and the retired are categorized into the Intermediate Social Class (following Cramer, 2012). Teachers, executives and academics are classified into the Highest Social Class. In case father's occupation is missing, we use father's education for individuals in the 1957 survey. Father's education is classified into 3 levels, which we directly translate into the three social classes. We use mother's education in case the father died or was not present in the household.

Hence, pupils with severe health problems impairing going to school in the first place will not be in our sample. Moreover, in the 1983 wave of the survey male respondents were asked whether they served in the military, which was compulsory in the Netherlands at the time for all males turning age 18. The main reason for disqualification for the military is health problems.[9] Since the fraction of individuals having served in the military is almost identical across educational levels, this provides some indirect evidence that health differences across educational levels were minimal during teenage years. We furthermore refer to Conti et al. (2010) who showed that in their sample childhood health was not an important determinant of educational choice. The lack of information on childhood health should therefore not be a major source of concern.

The bottom panel of table 1 includes descriptive statistics for the control variables. Males (58 percent) are overrepresented because of specific efforts in 1983 and 1993 to contact males, as researchers were mainly interested in labor market behavior and a lot of women did not participate in paid labor at that time. Close to half of the families belong to the lowest social class, and only 3 percent to the highest. In 28 percent of the cases, the child had to work at least sometimes in the family business or farm.

More than 75 percent of the schools are Roman Catholic, corresponding well to the predominant Catholicism of the Noord-Brabant province. One out of five school are Protestant, and the remaining five percent is divided between special and public schools. The average number of teachers per school is close to 7, which is one more than the number of grades. 64 percent of all pupils did not repeat any grade, 27 percent repeated one grade, and 9 percent repeated two or more grades.

For almost one out of four children, the teacher thinks it's best to stay in primary school, while the teacher advises lower vocational school for 38 percent of the children. Around a fourth is advised to go to lower secondary school, leaving only 10 percent of whom the teacher thinks higher secondary school is the best fit. More than 30 percent of the parents want their child to start working after primary school, with 27 percent preferring their child to follow a vocational training first, and 41 percent of the parents advise their child to enter into general secondary school.

3 Methodology

Our empirical approach is an extension of the structural equation framework developed by Conti et al. (2010), which is rooted in the framework suggested by Carneiro et al. (2003). It allows a flexible way of modeling the interrelationships between abilities, education and health outcomes. We first present the standard model, after which we will present our two extensions. Finally, we explain how we estimate the treatment effect of education.

[9] Other reasons were exemption owing to one's brother's service, grounds of conscience, or personal indispensability (e.g. Van Schellen and Nieuwbeerta, 2007)

3.1 Basic structural equation model

The standard model presented in Conti et al. (2010) consists of three parts: (i) a binary educational choice depending on latent abilities and other covariates, (ii) potential outcomes depending on the choice of education, latent abilities, and other covariates, and (iii) a measurement system for the latent abilities.

The binary indicator for education D_i is defined as 1 if individuals took any education beyond the compulsory schooling age, and 0 if not:

$$D_i = \begin{cases} 1 & \text{if } D_i^* \geq 0 \\ 0 & \text{otherwise} \end{cases} \tag{1}$$

where we assume D_i^* is an underlying latent utility which is continuous and linear, and depends on latent abilities θ, and observed characteristics X^D:

$$D_i^* = \gamma X_i^D + \alpha_D \theta_i + \upsilon_{iD} \tag{2}$$

with υ_D an error term independent of X^D and θ. We assume that υ_D is normally distributed, which implies that we have a probit model for the educational choice. We fix the variance at 1 since the variance is not identified in a probit model.

The second part is the potential outcomes part, in which there are two potential outcomes Y_{i1} and Y_{i0} where the former is the outcome in case the individual chose to pursue education beyond what is compulsory, and the latter is the outcome in case the individual dropped out of school right after the compulsory schooling age. In case of a continuous outcome, the observed outcome Y_i can be written as

$$Y_i = D_i Y_{i1} + (1 - D_i)\, Y_{i0} \tag{3}$$

where both Y_{i1} and Y_{i0} depend on latent ability θ, and on observed characteristics X^Y:

$$Y_{i1} = \beta_1 X_i^Y + \alpha_1 \theta_i + \nu_{i1} \tag{4}$$

$$Y_{i0} = \beta_0 X_i^Y + \alpha_0 \theta_i + \nu_{i0} \tag{5}$$

with (ν_0, ν_1) independent of X^Y and θ, and both follow a normal distribution with variance σ_1^2 and σ_0^2, respectively.

The final part of the model is the measurement equation, where one or more measurements implicitly define the latent ability θ:

$$M_i = \delta X_i^M + \alpha_M \theta_i + \upsilon_{iM} \tag{6}$$

with υ_M independent of X^M and θ. We assume that υ_M is normally distributed with variance σ_M^2.

If a standard normal distribution for the latent ability θ is assumed, the model is identified and can be estimated by Maximum Likelihood on basis of Gaussian quadrature approximation or simulation methods.

3.2 Allowing for a duration dependent variable

While the basic model of Conti et al. (2010) is extremely useful in disentangling the relative contributions of education and abilities on continuous and binary health outcomes, the model does not allow for a duration outcome like survival till death.

In our extended model, the first part is the same, defining a binary educational choice as in (1) and (2), placing the cut-off at *Lower Education* (primary school). Hence, in our model individuals face the choice of quitting after primary education ($D = 0$), or enrolling into secondary education ($D = 1$). Depending on the educational choice, we have two potential outcomes Y_{i1} and Y_{i0} where the former is the outcome in case the individual chose to pursue more education after primary school, and the latter is the outcome in case the individual chose not to pursue any further education after primary school. The measurement equation for latent ability is also the same as in Conti et al. (2010), and defined by (6), where we have two measurements for latent cognitive ability.

The main difference between our model and the Conti et al. (2010) model, is that for a duration outcome like mortality it is more common to define the potential outcomes in terms of the hazard (or intensity) that the outcome of interest occurs.[10] The observed hazard is

$$\lambda(t_i) = D_i\lambda^{(1)}(t_i) + \big(1 - D_i\big)\lambda^{(0)}(t_i) \tag{7}$$

which is equivalent to

$$\lambda(t_i) = \lambda^{(1)}(t_i)^{D_i} \cdot \lambda^{(0)}(t_i)^{1-D_i} \tag{8}$$

with $\lambda^{(1)}(t_i)$ being the hazard rate for an individual with education level beyond primary school ($D_i = 1$), and $\lambda^{(0)}(t_i)$ being the hazard rate for an individual with an education level equal to primary school ($D_i = 0$). We assume a Gompertz proportional hazard model for the two potential hazards, which has been shown to be an accurate representation of mortality between the ages of 30 and 80 (e.g. Gavrilov and Gavrilova, 1991). Both potential hazards depend on the latent skill θ,[11] and observed characteristics X^Y:

$$\lambda^{(0)}(t_i|X^Y, \theta) \;=\; \exp\big(a_0 t_i + \beta_0 X_i^Y(t_i) + \alpha_0\theta_i\big) \tag{9}$$

$$\lambda^{(1)}(t_i|X^Y, \theta) \;=\; \exp\big(a_1 t_i + \beta_1 X_i^Y(t_i) + \alpha_1\theta_i\big) \tag{10}$$

The effect of the latent skill on the hazard is captured by α_0 and α_1. The corresponding potential survival rates are simply

$$S^{(0)}(t_i|X^Y, \theta) \;=\; \exp\Big(-\int_0^t \lambda^{(0)}(s_i|X^Y, \theta)ds\Big) \tag{11}$$

$$S^{(1)}(t_i|X^Y, \theta) \;=\; \exp\Big(-\int_0^t \lambda^{(1)}(s_i|X^Y, \theta)ds\Big) \tag{12}$$

[10] We can use a duration model with potential outcomes because the endogenous education choice is determined before mortality plays a major role: mortality can be largely ignored for young ages. If the education choice would still play a role during higher mortality rates the model should take dynamic selection into account. Then a 'timing-of-events' model could be a better model, see Abbring and Van den Berg (2003).

[11] The latent skill in the hazard is similar to including unobserved heterogeneity in the hazard.

An important feature of duration data is that for some individuals we only know that he or she survived up to a certain time (often the end of the observation window). In this case an individual is (right) censored, $\Delta_i = 0$, and we use the survival function instead of the hazard in the likelihood function. Another feature of duration data is that only individuals are observed having survived up to a certain age. In our case, mortality follow-up is only available from age 55 onwards. In this case the individuals are left-truncated, and we need to condition on survival up to the age of first observation, t_{i0}.

The likelihood contribution of individual i in our duration model is[12]

$$L_i^{(j)} = \lambda^{(j)}(t_i)^{\Delta_i} S^{(j)}(t_i)/S^{(j)}(t_{i0}), \qquad j = 0, 1 \tag{14}$$

With left-truncated data the distribution of the latent skill among the survivors (up to the left-truncation time) changes. When only individuals are observed that have survived until age t_{i0} the likelihood contribution is

$$\begin{aligned} L_i \;=\; \int & \Big[\Phi\big(\gamma X_i^D + \alpha_D\theta\big) \cdot \lambda^{(1)}(t_i|X^Y,\theta)^{\Delta_i} S^{(1)}(t_i|X^Y,\theta)/S^{(1)}(t_{i0}|X^Y,\theta)\Big]^{D_i} \\ & \times \Big[\Phi\Big(-\gamma X_i^D - \alpha_D\theta\Big) \cdot \lambda^{(0)}(t_i|X^Y,\theta)^{\Delta_i} S^{(0)}(t_i|X^Y,\theta)/S^{(0)}(t_{i0}|X^Y,\theta)\Big]^{1-D_i} \\ & \times \frac{1}{\sigma_M}\phi\Big(\frac{M_i - \delta X_i^M - \alpha_M\theta}{\sigma_M}\Big)\, dH(\theta|T > t_{i0}) \end{aligned} \tag{15}$$

with the distribution of the latent skills conditional on survival up to t_{i0}

$$dH(\theta|T > t_{i0}) = \frac{\Big[\Phi\big(\gamma X_i^D + \alpha_D\theta\big) S^{(1)}(t_{i0}|X^Y,\theta) + \Phi\big(-\gamma X_i^D - \alpha_D\theta\big) S^{(0)}(t_{i0}|X^Y,\theta)\Big] h(\theta)}{\int\Big[\Phi\big(\gamma X_i^D + \alpha_D\theta\big) S^{(1)}(t_{i0}|X^Y,\theta) + \Phi\big(-\gamma X_i^D - \alpha_D\theta\big) S^{(0)}(t_{i0}|X^Y,\theta)\Big] h(\theta)\, d\theta} \tag{16}$$

with $h(\theta)$ is a normal distribution with variance $\sigma_\theta^2 = 1$. The maximum likelihood estimation of the parameters involves the calculation of an integral that does not have an analytical solution. However, Gaussian quadrature can approximate this one dimensional integral very well.

3.3 Allowing for an ordered discrete educational choice

While Conti et al. (2010) define education as a binary variable, usually education is available in more than two categories with a natural ordering of the alternative education

[12] Note that to accommodate a time-varying covariate that changes at some discrete time, say at $t_{i1}, t_{i2}, \ldots, t_{iJ}$, is very similar to repeated left-truncation, with the observation 'censored' ($\Delta_{ij} = 0$) for each discrete time point. Thus, the likelihood contribution is

$$L_i = \prod_{j=1}^{J} \lambda^{(j)}(t_{ij})^{\Delta_{ij}} S^{(j)}(t_{ij})/S^{(j)}(t_{ij-1}), \qquad j = 0, 1 \tag{13}$$

levels. We extend the standard model to account for this type of ordinal independent variable, where the starting point is, again, an index model with a single latent variable given as in (2). Assume there are K education levels and define D_i as the indicator of education that takes value k if the individual has reached education level k:

$$D_i = k \qquad \text{if } \zeta_{k-1} < D_i^* \leq \zeta_k \tag{17}$$

where $\zeta_0 = -\infty$ and $\zeta_K = \infty$. Then, assuming normally distributed υ_D, we have an ordered probit model:

$$\Pr\Big[D_i = k\Big] = \Phi\big(\zeta_k - \gamma X_i^D - \alpha_D \theta\big) - \Phi\big(\zeta_{k-1} - \gamma X_i^D - \alpha_D \theta\big) \tag{18}$$

with $(K-1)$ additional threshold parameters, ζ_k. Each education level now has a potential hazard $\lambda^{(k)}$, with observed hazard:

$$\lambda(t_i) = \sum_{k=1}^{K} I_{ik} \lambda^{(k)}(t_i) \tag{19}$$

where $I_{ik} = I\big(D_i = k\big)$ is an indicator function (1 if $D_i = k$ and zero otherwise). All $\lambda^{(k)}(\cdot)$ are Gompertz hazards and depend on exogenous characteristics X^Y and on the unobserved latent ability, i.e.,

$$\lambda^{(k)}(t_i | X^Y, \theta) = \exp\big(a_k t_i + \beta_k X_i^Y(t_i) + \alpha_k \theta_i\big) \tag{20}$$

The likelihood becomes

$$\begin{aligned} L_i \;=\; & \int \prod_{k=1}^{K} \Bigg[\Big\{ \Phi\big(\zeta_k - \gamma X_i^D - \alpha_D \theta_i\big) - \Phi\big(\zeta_{k-1} - \gamma X_i^D - \alpha_D \theta_i\big) \Big\} \\ & \qquad \cdot \lambda^{(k)}(t_i | X^Y, \theta)^{\Delta_i} S^{(k)}(t_i | X^Y, \theta) / S^{(k)}(t_{i0} | X^Y, \theta) \Bigg]^{I_{ik}} \\ & \qquad \times \frac{1}{\sigma_M} \phi\Big(\frac{M_i - \delta X_i^M - \alpha_M \theta_i}{\sigma_M}\Big)\, dH(\theta | T > t_{i0}) \end{aligned} \tag{21}$$

with the distribution of the latent skills conditional on survival up to t_{i0}

$$dH(\theta | T > t_{i0}) = \frac{\prod_{k=1}^{K} \Big\{ \Phi\big(\zeta_k - \gamma X_i^D - \alpha_D \theta_i\big) - \Phi\big(\zeta_{k-1} - \gamma X_i^D - \alpha_D \theta_i\big) \Big\} S^{(k)}(t_{i0} | X^Y, \theta) h(\theta)}{\int \prod_{k=1}^{K} \Big\{ \Phi\big(\zeta_k - \gamma X_i^D - \alpha_D \theta_i\big) - \Phi\big(\zeta_{k-1} - \gamma X_i^D - \alpha_D \theta_i\big) \Big\} S^{(k)}(t_{i0} | X^Y, \theta) h(\theta)\, d\theta} \tag{22}$$

3.4 Estimating the effects of ability and education

The individual-specific treatment effect of education is $Y_{i1} - Y_{i0}$, where the script 1 refers to the potential outcome corresponding to post-primary education, and the script 0 refers to the potential outcome with only primary education. However, we are usually not interested in the individual specific treatment effect, but rather in the average treatment effect over some population. For survival data, this can be done in many ways, and we choose to define the treatment effect in terms of the survival function. Using the estimated parameters, we define the Average Treatment Effect (ATE) at all ages t as follows:

$$ATE(t) = \int\int E\left[S^{(1)}(t) - S^{(0)}(t)|X = x, \theta = f\right] dF_{X,\theta}(x, f) \quad (23)$$

where $S^{(1)}(t)$ denotes the survival time up to a age t for individuals with at least secondary education ($D = 1$), $S^{(0)}(t)$ is the survival time up to age t for those with primary school only ($D = 0$), X are the covariates, and θ is the value of the latent cognitive ability. We integrate over the joint distribution of the covariates and the latent ability, $F_{X,\theta}(x, f)$. Note that the treatment effect is conditional on surviving to the initial age, which is 55 in our case.

The difference in the Kaplan-Meier survival curves can be interpreted as the *unconditional* survival difference between the two levels of educational attainment. The treatment effect of education is the *conditional* survival difference between the two levels of educational attainment, where conditioning is on basis of cognitive ability, family background and the other control variables in our model.

The ATE estimates the average effect over the whole population. However, the effect may well be very different for individuals in different parts of the education distribution. Therefore, we additionally define the Average Treatment Effect on the Treated (ATET), i.e. the average effect for those with $D = 1$, and the Average Treatment Effect on the Untreated (ATEU), i.e. the average effect for those with $D = 0$ as follows:

$$ATET(t) = \int\int E\left[Y_1(t) - Y_0(t)|X = x, \theta = f, D = 1\right] dF_{X,\theta|D=1}(x, f) \quad (24)$$

$$ATEU(t) = \int\int E\left[Y_1(t) - Y_0(t)|X = x, \theta = f, D = 0\right] dF_{X,\theta|D=0}(x, f) \quad (25)$$

Unfortunately, the integrals cannot be solved analytically, as the dimension of the covariates X is too large. Hence in order to illustrate the treatment effects we resort to simulation. This procedure consists of three steps:

1. We determine the distribution of all included variables – separately for the whole sample, and separately for those with $D = 0$ and $D = 1$.

2. We draw 10,000 individuals on basis of the empirical distribution of the covariates and compute the conditional hazard rates using the estimated coefficients of equations (9) and (10), conditional on the value of the latent skill.

3. For every conditional hazard rate we determine the unconditional survival function for every age from 55 to 100 on basis of equations (11) and (12), and by integrating out the latent skill through Gaussian quadrature methods.

We repeat these steps 100 times and for each simulation round we draw a vector of parameter estimates assuming that the estimated coefficients are normally distributed around the point estimates with a variance-covariance matrix equal to the estimated one.

With this information, we can compute the fraction of individuals that is still alive at a certain age for the two educational groups. This defines the treatment effect of education, since we condition on cognitive ability and the other covariates. The simulations also allow us to compute life expectancy separately for the two educational groups.

In order to illustrate the importance of the treatment effect, we decompose the unconditional survival differences from the Kaplan-Meier curves in Figures 1 and 2 into the treatment effect (the conditional survival differences) and a residual, which we interpret as a selection effect on basis of cognitive ability and the other factors.

For the ordinal education measure the procedure is very similar. We have three potential hazards and three possible survival functions, one corresponding to each educational level. Although there are more possibilities now to define the treatment effects of education, we choose to focus on two binary treatment effects of the particular educational level to the educational level directly preceding it. Hence, we estimate two different treatment effects: (i) lower vocational education compared to primary education only, and (ii) at least general secondary education compared to vocational education.

4 Results

In this section we present our main results regarding the effect of cognitive ability and education on mortality. Our baseline specification is the survival model with a binary education variable and two measurements for cognitive ability. We estimate the model by maximizing the likelihood in (15), and present the results in section 4.1. Then we generalize the model by allowing for an ordinal educational variable including three levels, the likelihood of which is presented in (21), and results of which are presented in section 4.2.

The set of included observed characteristics does not differ in both cases. Exogenous factors influencing the outcome, X_y in (9) and (10), include male, whether the child is working, family social class, and birth rank. Factors *additionally* influencing the measurements of cognitive ability, X_m in (6), include school type and the number of teachers at school. Finally, on top of the exogenous variables affecting the outcome and intelligence, *additional* factors influencing the educational choice, X_d in (2), include the teacher's advice, whether a grade was repeated, and the preference of the parents.

4.1 Binary education variable - two measurements

This section presents the results for a model with a binary educational choice, where the cut-off for education is placed at primary school. Hence, $D = 0$ refers to individuals only having finished primary school, while $D = 1$ implies that the individual attended at least secondary school (corresponding to levels 2 and 3 in the definition of *Education* in section 2). We include both intelligence measurements, the Raven test and the Vocabulary test, to build a comprehensive factor for cognitive ability.

Table 2 contains the parameter estimates of the model. The first column shows that our latent factor of cognitive ability strongly influences the educational choice, as expected. Figure 3 illustrates the importance of cognitive ability for the probability of enrolling in secondary school ($D = 1$). The probability of entering secondary school is already beyond 0.6 for those with the lowest cognitive skills, and gradually increases towards one for those with the highest cognitive skills.

Females were less likely to enter secondary school, as are children who had to work in the family business during primary school. Family social class is a strong predictor of education, with children from the higher social classes significantly more likely to enter secondary school. Children who went to protestant or other schools, as compared to those who went to catholic schools, were more likely to enter secondary school. Strong predictors of educational choice are the teacher's advice and the preference of the parents. Children who repeated one or more grades were less likely to enter secondary school.

Interestingly, columns 2 and 3 show that on both measurements of cognitive ability girls did slightly better, and children from higher social classes had higher scores. School characteristics such as the school type and the number of teachers also relate to the test scores.

The final two columns of the table present the determinants of mortality across the two educational groups. While the point estimates of the effect of cognitive ability on mortality are negative as expected, the effects do not reach statistical significance at the 10 percent level, although the p-values are extremely close to the 10 percent cut-off. Males have a higher hazard of dying compared to females, although the effect is only statistically significant among the higher educated.

The coefficients in Table 2 allow to compute the treatment effects of education, as described in section 3.4. Figure 4 shows the ATE, ATET, and ATEU for all age groups from 55 to 75 years of age. The effect of entering secondary school on mortality, after controlling for family background and cognitive ability, is positive and increases with age. The effect sizes can be interpreted as percentage point differences in the survival probability at a certain age. Hence, around age 70 the treatment effect of entering secondary school is a 2 percentage point increase in the survival probability. Note that the confidence intervals are fairly wide, such that the treatment effects only reach statistical significance at higher ages.

If we extrapolate the estimated survival functions outside of our observed age window, the simulations allow computing the estimated differences in life expectancy. This provides an alternative summary measure of the treatment effects. The life expectancy of those only finishing primary school is 82.86, compared with 87.15 for those having finished at least secondary school, a difference that is statistically significant. This implies that the average treatment effect of secondary school is more than 4 years in life expectancy. The average treatment effect on the treated is also slightly larger than 4, while the average treatment effect on the untreated is even close to a 5 years gain in life expectancy.

Using the estimated treatment effects, it is possible to decompose the unconditional differences in the Kaplan-Meier survival curves from Figure 2 into a treatment effect (ATE)[13] of education and a selection effect on basis of cognitive ability and family background variables. Figure 5 shows that at early ages mortality differentials are mainly due to selection effects, while after age 60 the importance of education increases. For most ages, the treatment effect of education is responsible for around half of the unconditional differences in survival across educational groups, which is in line with the findings by Conti and Heckman (2010).

To gauge the importance of cognitive ability in the selection effect, we additionally ran all models without the latent factor for cognitive ability. The results show that the treatment effects are larger in a model without cognitive ability.[14] This is an indication that cognitive ability plays an important role in the selection effect. It is tempting to decompose the selection effect into a selection due to cognitive ability and a selection on other characteristics. The selection on other characteristics can be computed as the difference between (i) the unconditional difference from the observed Kaplan-Meier survival rate of the two education levels and (ii) the treatment effect from the model without cognitive skills. The selection on cognitive ability can then be easily computed as the difference between the total selection effect and the part of the selection effect attributed to other characteristics. This is illustrated in Figure 6, which shows that cognitive ability explain the largest part of the selection effect. We have to emphasize, however, that this interpretation should be taken with care, as this is not a formal test of the importance of cognitive ability.

4.2 Ordinal education, two measurements

This section presents the results for a model with an ordinal educational choice, corresponding to levels 1, 2 and 3 in the definition of *Education* in section 2. We include both measures for cognitive ability, the Raven test as well as the Vocabulary test, to build a comprehensive factor.

The coefficient estimates of the exogenous variables are very similar to the ones presented for the binary educational variable, and hence are not presented here. Figure 7 shows the relationship between cognitive ability and the probabilities of entering the three levels of education. For those with the lowest cognitive skills quitting school after primary school ($D = 0$) or entering lower vocational educational ($D = 1$) are the most

[13]The corresponding graphs for the ATET and ATEU are very similar.

[14]Results are available upon request.

likely alternatives with each a probability of around 0.4. The likelihood of both choices decreases with a higher level of cognitive ability, with the probability of quitting primary school decreasing sharper than the probability of entering lower vocational education. In contrast, while less than one out of five of those with the lowest cognitive ability enter general secondary education, this probability increases almost linearly towards one with increasing cognitive ability.

Figure 8 presents the Average Treatment Effects for the three different educational levels. Results for the ATET and ATEU are very similar. It is clear that there is a large, but insignificant, treatment effect of lower vocational school (level 2) compared to only having finished primary school (level 1). At age 75, those only finishing primary school are around four percentage points more likely to die than those who entered vocational school. The treatment effect of general secondary school compared to lower vocational school is practically zero. This clearly indicates that the largest difference is between those having finished primary school and those beyond primary school. Hence, the dichotomization in the previous subsection seems justified.

The estimated life expectancy of those only having finished primary school is 84.71, compared to 87.76 for those having finished lower vocational school, and 86.99 for those with at least general secondary school. This implies that the treatment effect between level 1 on the one hand and level 2 and 3 on the other hand is around 3 years of life expectancy, which is slightly smaller but reasonably close to the treatment effects estimated when using a dichotomous classification of education. It also shows that the treatment effect of lower vocational education (level 2) compared to general secondary education (level 3) in terms of life expectancy is negative, but very small.

Finally, if we decompose the unconditional survival differences between the three educational groups into a treatment effect (ATE) and a selection effect, we obtain Figure 9. This graph shows that the treatment effect of primary to vocational education is positive and becomes larger than the selection effect from age 70 onwards, in line with the findings of the dichotomous indicator for education. The treatment effect from vocational to higher education is negligible. The selection effect here is negative rendering the raw differences between vocational and higher educated small.

4.3 Robustness checks

While mortality is an objective, and in some sense "the ultimate", health outcome, the influence of education and cognitive ability may differ depending on the health outcome used. In the 1993 wave of our Brabant survey, hence around age 53 for our sample, a subjective assessment of one's health was asked to the respondents in five categories, i.e. "poor", "sometimes good, sometimes bad", "fair", "good", and "very good". We estimated the model described in section 3, now allowing for an ordinal dependent variable variable, to check robustness to our main outcome measure, and to compare our results to the literature.[15]

[15] All results not presented and details of the models used in this section are available upon request.

We estimated the treatment effects of education on the probability to report any of the five categories. In the model using the binary educational variable, the average treatment effects for categories "poor" and "sometimes good, sometimes bad" are -0.03 and -0.08 respectively. The average treatment effects for the categories "fair" and "very good" are very close to zero, while the average treatment effect on the probability of reporting to be in "good" health is large and amounts to a 15 percentage points increase. The results are similar when we focus on the ATET or ATEU.

Interestingly, when using the three educational categories instead of the binary indicator, it becomes clear that the treatment effects for the lower four categories (that is, up to "good" health) is entirely driven by the difference between lower education and lower vocational education, as in the case of the mortality outcomes. However, there is a relatively large average treatment effect on the probability of reporting "very good" health when moving from lower vocational education to general secondary education.

The results are largely in line with using mortality as an outcome, where the largest differences were also between primary and lower vocational education. When comparing our results to the literature, we confirm the findings of Hartog and Oosterbeek (1998) that both education and cognitive ability affect self-reported health, and that part of the effect of education is due to cognitive ability. Conti and Heckman (2010) used a binary indicator for "poor health" and found that half of the raw differences in poor health is due to a treatment effect of education and the other half was selection. We found that the treatment effect plays a larger role in explaining the raw differences in health levels in a model with five health levels and binary education, see Figure 10. For the model with ordinal education the importance of the treatment effect is less pronounced, see Figure 11.

Since the sample size is somewhat small we chose not to present all results separately by gender. Yet, since both educational choices and survival are obviously dependent on gender, we ran all models separately for males and females. Figures 12 and 13 show the raw Kaplan-Meier survival curves for males and females separately. While survival is larger for females, strong disparities in survival across educational groups exist for both males and females. The treatment effects are slightly larger for females than for males. However, the relative importance of the treatment effect derived from the decomposition of the raw survival differences into a treatment and a selection effect, is higher for males (see Figure 14 and 15).

Even though the initial sample in 1952 was found to be representative for the Dutch population at that time, more than half of the sample is lost between 1952 and our observation period that starts in 1995. This could lead to an attrition bias, if attrition is non-random. Unfortunately, we do not have access to the original data files such that we cannot investigate attrition directly. However, Hartog (1989) investigated the non-response for the 1983 survey and found no attrition bias in a wage analysis (see also Vermunt, 1988).[16] Since the sample in 1983 has been shown to be representative, we reran

[16] Following Hartog, (1989) we investigated whether the attrition between 1993 and 1995 was related to observed characteristics. Literally all explanatory variables including education, family background, and intelligence were not related to attrition. The only exception was self-reported health; a worse health

all analyses on just the respondents that were observed in 1983 and found no substantial changes in the results. This suggests that selective attrition does not affect our results.

One could argue that the Raven progressive matrices test is a purer measurement of cognitive ability and should be used independently from the vocabulary test. We ran all analyses for both the binary and ordinal educational classification using only the Raven test as a measure of cognitive ability, and the results were very similar.

We additionally varied the observed characteristics in the model. First by including additional variables among the exogenous variables such as family size, number of children, additional school characteristics (e.g. whether restricted to girls, restricted to boys, or mixed), and whether both parents were still alive. These variables were not statistically significant in any of the models, and did not alter the results. Second, we also checked robustness to excluding individuals with item non-response on some of the observed characteristics. In the main analysis we decided to include separate categories for the missing values in order to maximize the sample size. When excluding individuals with item non-response the results remain similar. Results for both of these analyses are available upon request.

5 Discussion

This paper decomposes survival differences across educational groups into a 'treatment effect' of education, and a 'selection effect' on basis of cognitive ability and other background variables. We extend the structural equation model of Conti et al. (2010) and estimate the model on basis of a Dutch cohort born around 1940. The treatment effect of education (i.e. the effect of entering secondary school as opposed to leaving school after primary education) is responsible for around half of the raw differences in survival, and corresponds to a 4 years gain in life expectancy.

Even though we analyze mortality between ages 55 and 75 rather than self-reported health at age 30, our findings are in line with the results presented by Conti et al. (2010). Due to this striking similarity in findings, irrespective of the health measures and samples used, two tentative conclusions regarding the education-health gradient are emerging. First, even after controlling for cognitive ability, family social class, and a range of other background variables, education remains important in determining mortality. This strongly suggests that at least part of the educational differences in health outcomes is due to a genuine, causal effect of education on health. Second, at least half of the raw association between education and health is due to confounding 'third factors', of which cognitive ability proved very important in our analysis, while Conti et al. (2010) stress the importance of non-cognitive factors.

A limitation of our data is the absence of direct measurements of non-cognitive ability. We do however observe the teacher's advice regarding secondary education of the child, which is a function of both the cognitive and non-cognitive skills of the pupil. Hence, one

status increased the probability of attrition between 1993 and 1995.

could argue that while controlling for cognitive ability, the teacher's advice could be a proxy for non-cognitive skills. When allowing the teacher's advice to influence mortality directly, on top of being a determinant of educational choice, our results are similar. This gives some comfort in claiming that the lack of a direct measurement of non-cognitive ability does not alter our main conclusions. Strictly speaking, we cannot rule out that specific non-cognitive factors could still influence both education and health, such that the 'treatment effect' of education cannot be interpreted literally as – and is likely to be an upper bound to – the causal effect of education on mortality.

A fruitful avenue for future research would be analyzing the robustness of our and Conti et al. (2010)'s results by investigating mortality and other health outcomes using a more elaborate set of non-cognitive abilities. In doing so, the literature could benefit from our extended structural equation model that allows for a duration dependent variable like mortality, and an ordinal independent variable such as educational attainment.

References

Abbring, Jaap, and Gerard van den Berg. 2003. "The Nonparametric Identification of Treatment Effects in Duration Models." *Econometrica* 71(5): 1491-1517.

Albouy, Valerie, and Laurent Lequien. 2008. "Does compulsory education lower mortality?" *Journal of Health Economics* 28(1): 155-168.

Arendt, Jacob Nielsen. 2005. "Does Education Cause Better Health? A Panel Data Analysis Using School Reforms for Identification." *Economics of Education Review* 24(2): 149-60.

Auld, M. Christopher, and Nirmal Sidhu. 2005. "Schooling, Cognitive Ability and Health." *Health Economics* 14(10): 1019-1034.

Behrman, Jere R., Mark R. Rosenzweig. 2004. "Returns to Birthweight." *The Review of Economics and Statistics* 86(2): 586-601.

Bago d'Uva, Teresa, Owen O'Donnell, and Eddy van Doorslaer. 2008. "Differential health reporting by education level and its impact on the measurement of health inequalities among older Europeans." *International Journal of Epidemiology* 37(6): 1375-1383.

Braakmann, Nils. 2011. "The causal relationship between education, health and health related behaviour: Evidence from a natural experiment in England." *Journal of Health Economics* 30(4): 753-763.

Carneiro, Pedro, Karsten T. Hansen, and James J. Heckman. 2003. "Estimating distributions of treatment effects with an application to the returns to schooling and measurement of the effects of uncertainty on college choice." *International Economic Review* 44(2): 361-422.

Carneiro, Pedro, Claire Crawford, and Alissa Goodman. 2007. "The Impact of Early Cognitive and Non-Cognitive Skills on Later Outcomes." CEE DP 92.

Carpenter, Patricia A., Marcel A. Just, and Peter Shell. 1990. "What one intelligence test measures: a theoretical account of processing in the Raven Progressive Matrices Test." *Psychological Review* 97(3): 404-431.

Case Anne, Angela Fertig, and Christine Paxson. 2005. "The lasting impact of childhood health and circumstance" *Journal of Health Economics* 24(2): 365-389.

Centraal Bureau voor de Statistiek, CBS. 2008. "Hoogopgeleiden leven lang en gezond" in: Gezondheid en zorg in cijfers 2008, CBS.

Clark, Damon, and Heather Royer. 2013. "The Effect of Education on Adult Mortality and Health: Evidence from Britain." *American Economic Review*, forthcoming.

Conti, Gabriella, James J. Heckman. 2010. "Understanding the Early Origins of the Education-Health Gradient: A Framework That Can Also Be Applied to Analyze Gene-Environment Interactions." *Perspectives on Psychological Science* 5(5): 585-605.

Conti, Gabriella, James J. Heckman, and Sergio Urzua. 2010. "The Education-Health Gradient." *American Economic Review Papers and Proceedings* 100: 234-238.

Conti, Gabriella, James J. Heckman, and Sergio Urzua. 2011. "Early Endowments, Education, and Health." Unpublished manuscript, University of Chicago, Department of Economics.

Cramer, Jan S. 2012. "Childhood Intelligence and Adult Mortality, and the Role of Socio-Economic Status." Tinbergen Institute Discussion Paper 2012-070/4.

Cutler, David, and Adriana Lleras-Muney. 2008. "Education and Health: Evaluating Theories and Evidence." In: House James S, Schoeni Robert F, Kaplan George A, Pollack Harold., editors. Making Americans Healthier: Social and Economic Policy as Health Policy. New York: Russell Sage Foundation.

Deary, Ian. 2008. "Why do intelligent people live longer?" *Nature* 456: 175-176.

Deary, Ian, and Wendy Johnson. 2010. "Intelligence and education: causal perceptions drive analytic processes and therefore conclusions." *International Journal of Epidemiology* 39: 1362-1369.

Dronkers, Jaap. 2002. "Bestaat er een samenhang tussen echtscheiding en intelligentie?", *Mens & Maatschappij* 77(1): 25-42.

Elias, Julio J. 2004. "The Effects of Ability and Family Background on Non-Monetary Returns to Education." Ph.D. dissertation. University of Chicago.

Fuchs, Viktor R. 1982. "Time Preference and Health: An Exploratory Study.", in V. Fuchs (ed.) Economic Aspects of Health (Chicago: The University of Chicago Press).

Gathmann, Christina, Hendrik Jurges, and Steffen Reinhold. 2012. "Compulsory Schooling Reforms, Education and Mortality in Twentieth Century Europe.", CESifo Working Paper Series No. 3755.

Gavrilov, Leonid A., and Natalia S. Gavrilova. 1991. The Biology of Life Span: A Quantitative Approach. New York: Harwood Academic Publisher, ISBN 3-7186-4983-7.

Gottfredson, Linda. 2004. "Intelligence: is the Epidemiologists' elusive "Fundamental Cause" of Social Class Inequalities in Health?" *Journal of Personality and Social Psychology* 86(1): 174-199.

Hartog, Joop. 1989. "Survey non-response in relation to ability and family background: structure and effects on estimated earnings functions." *Applied Economics* 21: 387-395.

Hartog, Joop, and Gerard Pfann. 1985. Vervolgonderzoek Noord-Brabantse zesdeklassers 1983, Verantwoording van hernieuwde gegevensverzameling onder Noordbrabantse zesdeklassers van 1952 ['Follow-up research among pupils of the 6th grade of primary schools in Noord-Brabant in 1983, A justification of renewed data collection among pupils of the 6th grade of primary schools in Noord-Brabant in 1952'], Amsterdam: University of Amsterdam.

Hartog, Joop, and Hessel Oosterbeek. 1998. "Health, wealth and happiness: why pursue a higher education? *Economics of Education Review*, 17(3): 245-256.

Hartog, Joop, Nicole Jonker, and Gerard Pfann. 2002. "Documentatie Brabant data ['Documentation Brabant data']", Amsterdam: Netherlands Institute for Scientific Information Services.

Kaestner, Robert, and Kevin Callison. 2011. "Adolescent Cognitive and Non-Cognitive correlates of Health." *Journal of Human Capital* 5(1): 29-69.

Lleras-Muney, Adriana. 2005. "The Relationship Between Education and Adult Mortality in the United States." *Review of Economic Studies* 72: 189-221.

Mathijssen, Mathias A.J.M., and G.J.M. Sonnemans. 1958. "Schoolkeuze en schoolsucces bij VHMO en ULO in Noord-Brabant ['Choice of school and school achievement in tertiary education in Noord-Brabant']", Tilburg: Zwijssen.

Mazumder, Bhaskar. 2008. ""Does Education Improve Health: A Reexamination of the Evidence from Compulsory Schooling Laws." *Economic Perspectives* 33(2).

Mazumder, Bhaskar. 2012. "The Effects of Education on Health and Mortality" *Nordic Economic Policy Review* 2012: 261-301.

Meghir, Costas, Martin Palme, and Emilia Simeonova. 2012. "Education, health and mortality: Evidence from a social experiment." NBER Working Paper 17932.

Murasko, Jason E. 2007. "A lifecourse study on education and health: The relationship between childhood psychosocial resources and outcomes in adolescence and young adulthood" *Social Science Research* 36(4): 1348-1370.

Oreopoulos, Philip. 2006."Estimating Average and Local Average Treatment Effects of Education when Compulsory School Laws Really Matter" *American Economic Review* 96(1): 152-175.

Raven, John C. 1958. Mill Hill Vocabulav Scale, 2nd ed. London: H. K. Lewis.

Ross, Catherine E., and Chia-ling Wu. 1995. "The links between education and health" *American Sociological Review* 60(5): 719-745.

Spearman, Charles. 1927. The Abilities of Man: Their Nature and Measurement. New York: Macmillan.

Statistics Netherlands. 2008. "The Dutch Standard Classification of Education, SOI 2006", by Roel Schaart, Mies Bernelot Moens en Sue Westerman 5 June 2008.

Van Kippersluis, Hans, Owen O'Donnell, and Eddy van Doorslaer. 2011. "Long Run Returns to Education: Does Schooling lead to an Extended Old Age?" *Journal of Human Resources* 46(4): 695-721.

Van Praag, Mirjam. 1992. "Zomaar een dataset: 'Noordbrabantse zesde klassers', Een presentatie van 15 jaar onderzoek ['Just a dataset: Pupils of the 6th grade of primary schools in Noord-Brabant, A presentation of a 15 years research project']" Amsterdam: University of Amsterdam.

Van Schellen, Marieke, and Paul Nieuwbeerta. 2007. "De invloed van de militaire dienstplicht op de ontwikkeling van crimineel gedrag." *Mens & Maatschappij* 82(1): 5-27.

Vermunt, Jeroen K. 1988. "Loglineaire modellen met latente variabelen en missing data" Tilburg: University of Tilburg.

Tables

Table 1: Descriptive Statistics of the Brabant Data sample

Variable	Average	Standard Deviation	Number of Observations
Dependent Variables			
Mortality	0.16	0.35	2,579
Independent Variables			
Education			
Lower Education	0.14	0.34	2,537
Lower Vocational Education	0.34	0.48	2,537
At least General Secondary School	0.51	0.35	2,537
Raven p.m. test	102.04	13.28	2,579
Vocabulary test	101.42	12.87	2,579
Control Variables			
Male	0.58	0.49	2,579
Birth Rank	2.50	2.55	2,412
Family Social Class			
Lowest Social Class	0.53	0.50	2,409
Middle Social Class	0.44	0.50	2,409
Highest Social Class	0.03	0.16	2,409
Child Works	0.28	0.45	2,256
School Religion			
Roman-Catholic	0.76	0.43	2,518
Protestant	0.19	0.40	2,518
Special	0.03	0.17	2,518
Public	0.02	0.13	2,518
Number of Teachers	6.92	2.47	2,452
Repeat			
No Repetition of Grade	0.64	0.48	2,462
Repeated Once	0.27	0.45	2,462
Repeated Twice or More	0.09	0.28	2,462
Teacher's Advice			
Continue Primary School	0.24	0.43	2,429
Lower Vocational Education	0.38	0.48	2,429
Lower Secondary Education	0.24	0.43	2,429
Higher Secondary Education	0.14	0.20	2,429
Preference of the Parents			
Work in Family Company	0.13	0.33	2,200
Paid Work without Vocational Education	0.20	0.28	2,200
Paid Work with Vocational Education	0.27	0.44	2,200
General Secondary Education	0.41	0.49	2,200

Notes: Author's calculations on basis of the Brabant Data linked to the municipality register and the mortality register.

Table 2: Duration model - Binary education variable, two measurements for ability

Outcome	Education D	Raven Test M_1	Vocabulary Test M_2	Hazard $\lambda^{(0)}$	Hazard $\lambda^{(1)}$
Cognitive Ability					
α	0.36^{***}	9.63^{***}	10.37^{***}	-0.33	-0.20
Constant term					
c	2.13^{***}	3.59^{***}	4.69^{***}	-11.68^{***}	-10.71^{***}
a				0.11^{***}	0.09^{***}
Control variables					
Male	-0.25^{***}	-0.93^{*}	-0.87^{*}	0.33	0.66^{***}
Child is working - base is "No"					
Yes	-0.29^{***}	-3.84^{***}	-7.14^{***}	0.34	0.15
Missing	-0.32^{***}	-1.07	2.52^{***}	-0.87	0.12
Family social class - base is "Low"					
Middle	0.42^{***}	2.55^{***}	2.24^{***}	-0.34	0.00
High	0.42	4.16^{***}	4.68^{***}	-0.34	0.44
Missing	-0.54^{***}	-4.39^{***}	-7.63^{***}	-0.67	0.17
Birthrank - base is "First"					
Second	-0.15	0.53	-0.02	-0.17	-0.00
Third or Fourth	-0.09	-0.22	-2.70^{***}	0.09	-0.19
Fifth or higher	-0.09	-3.02^{***}	-4.52^{***}	0.09	-0.26^{*}
Missing	0.11	-0.63	0.47	1.13^{*}	-0.65^{*}
School religion - base is "Catholic"					
Protestant	0.31^{***}	0.62	2.59^{***}		
Other	0.42^{**}	5.19^{***}	7.32^{***}		
Number of teachers - base is "5-8 teachers"					
≤ 4	-0.16	-3.81^{***}	-3.16^{***}		
$9-12$	0.05	0.37	0.42		
Missing	0.33	0.81	0.66		
Teacher's advice - base is "Lower vocational school"					
Continued primary school	-0.22^{**}				
Lower general secondary school	0.42^{**}				
Higher general secondary school	0.39				
Missing	-0.56^{**}				
Repeat grade - base is "None"					
Once	-0.30^{***}				
Twice or more	-0.74^{***}				
Missing	0.74^{*}				
Preference of the parents - base is "Only vocational education"					
Work in own company	-0.78^{***}				
Work without education	-1.29^{***}				
Work with education	-0.88^{***}				
General secondary school	-0.28				
Missing	-0.85^{***}				

* p-value < 0.1, ** p-value < 0.05, *** p-value < 0.01

Notes: Author's calculations on basis of the Brabant Data linked to the municipality register and the mortality register.

Figures

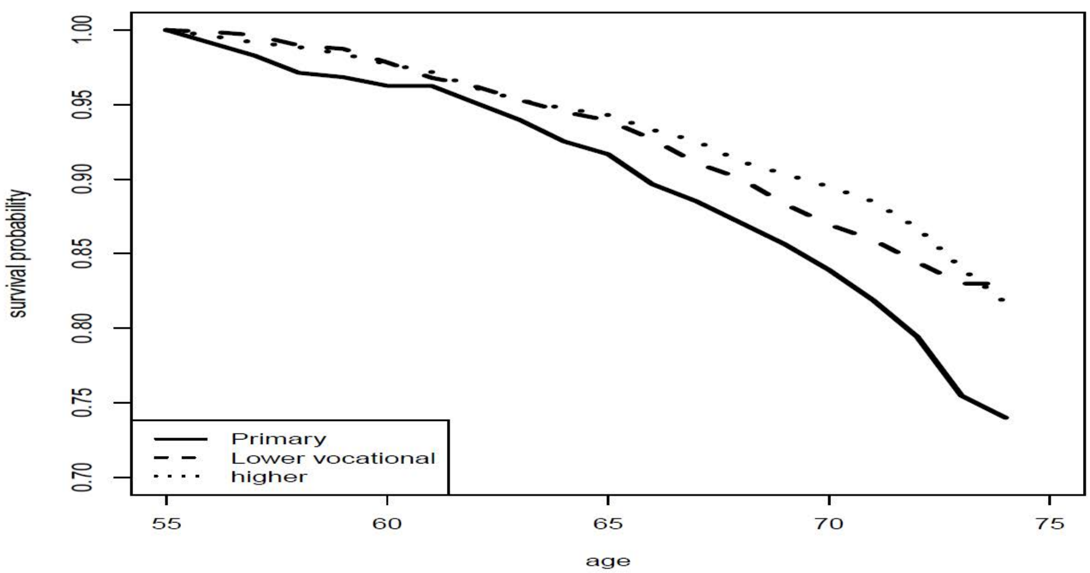

Figure 1: Kaplan-Meier Survival function by education level in three categories

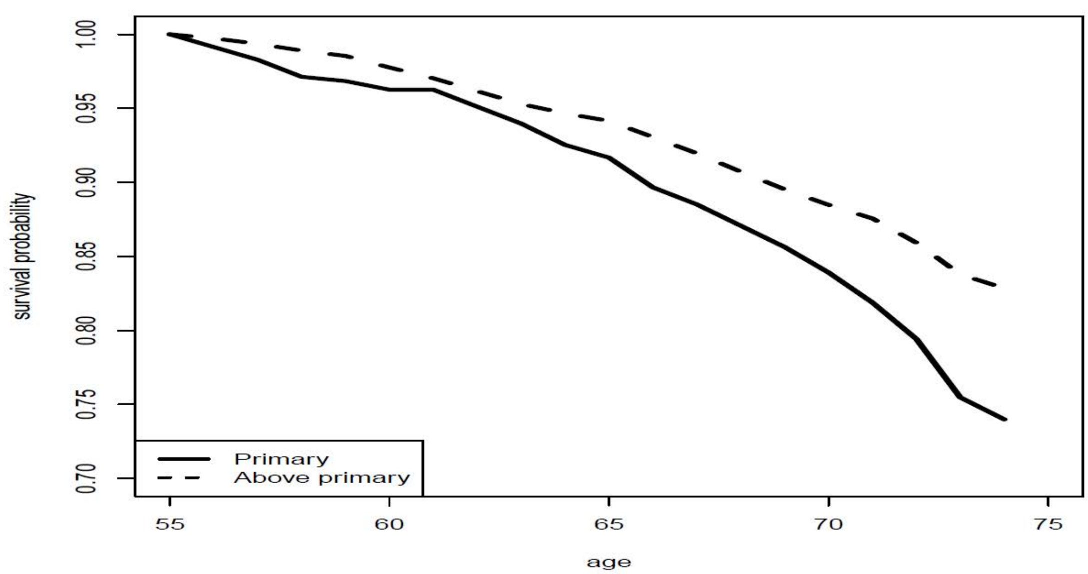

Figure 2: Kaplan-Meier Survival function by education level in two categories

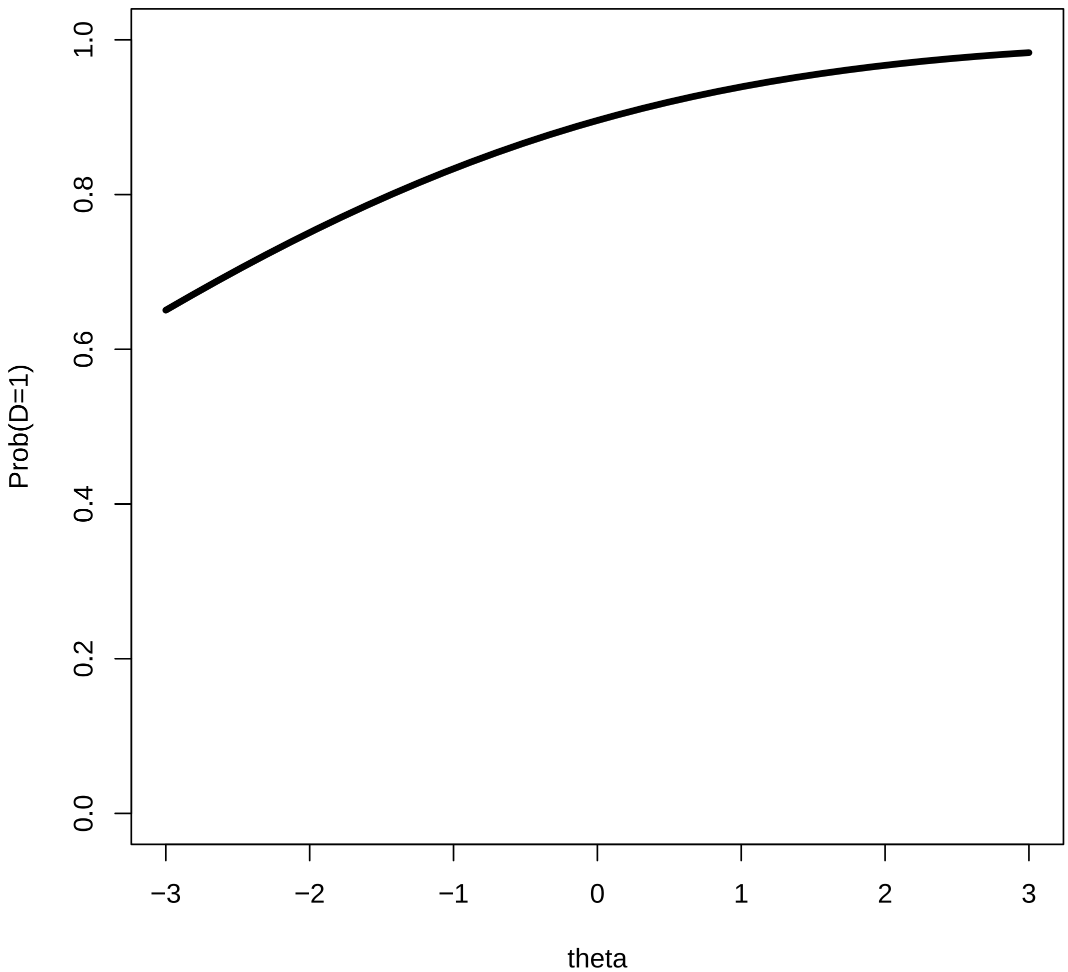

Figure 3: *Relationship between cognitive ability and the binary measure for education.*

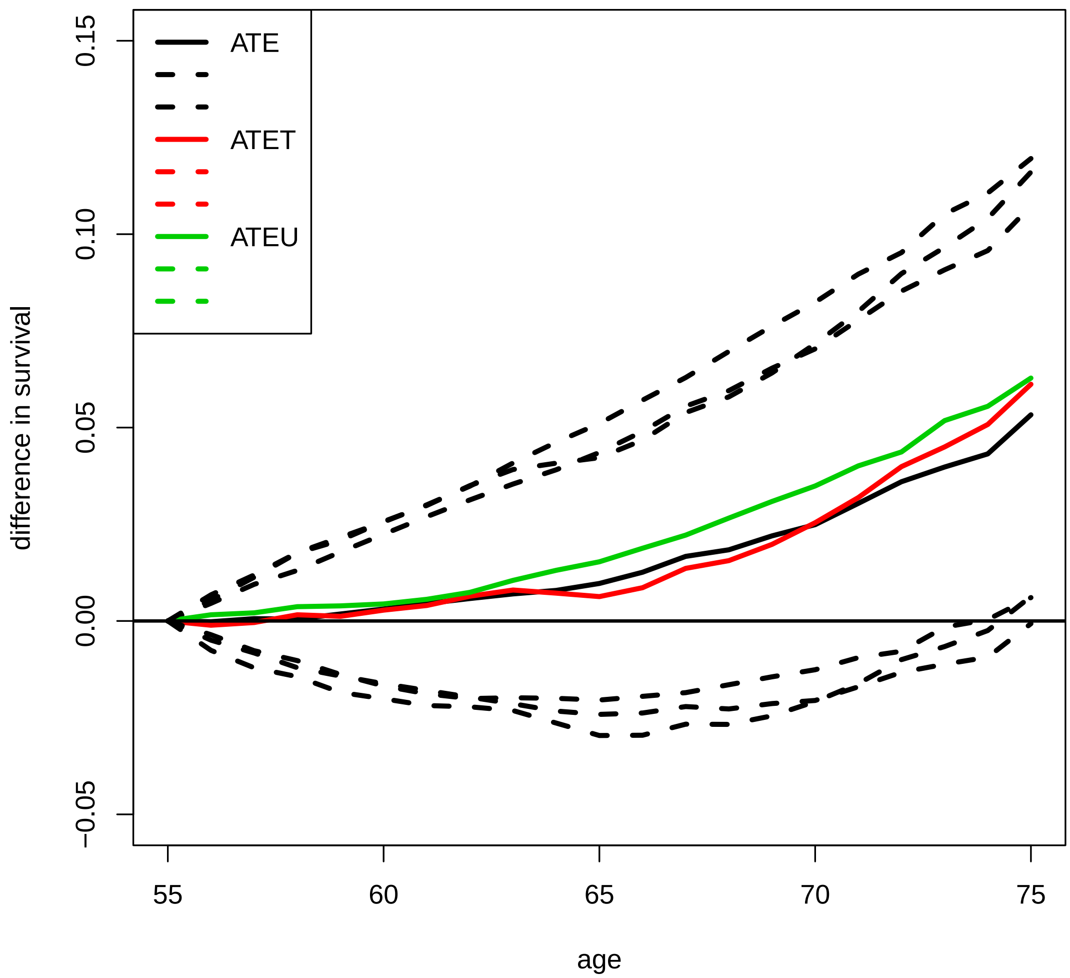

Figure 4: *Treatment effects by age, binary education variable, two measurements for cognitive ability. Dashed lines indicate the 90 percent confidence intervals.*

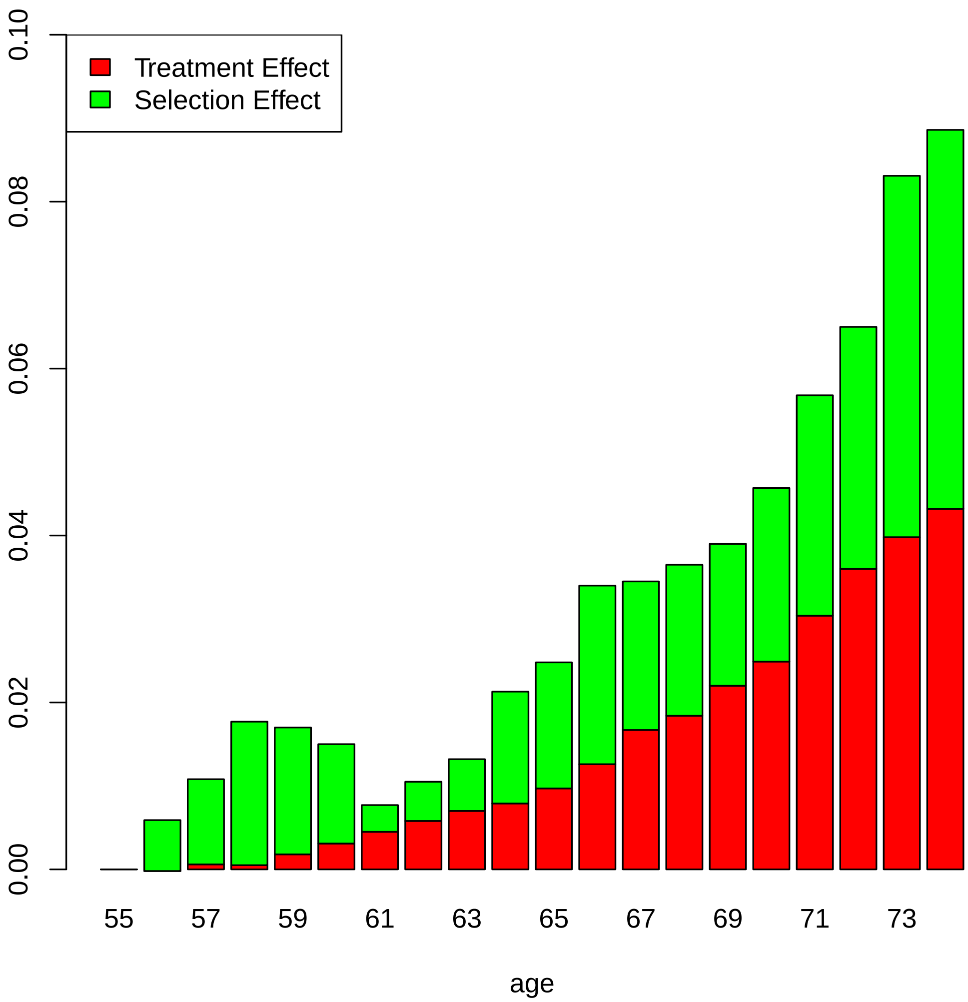

Figure 5: Decomposition of observed difference in the Kaplan-Meier Survival function in a treatment (ATE) and selection effect, with binary education variable and two measurements for cognitive ability

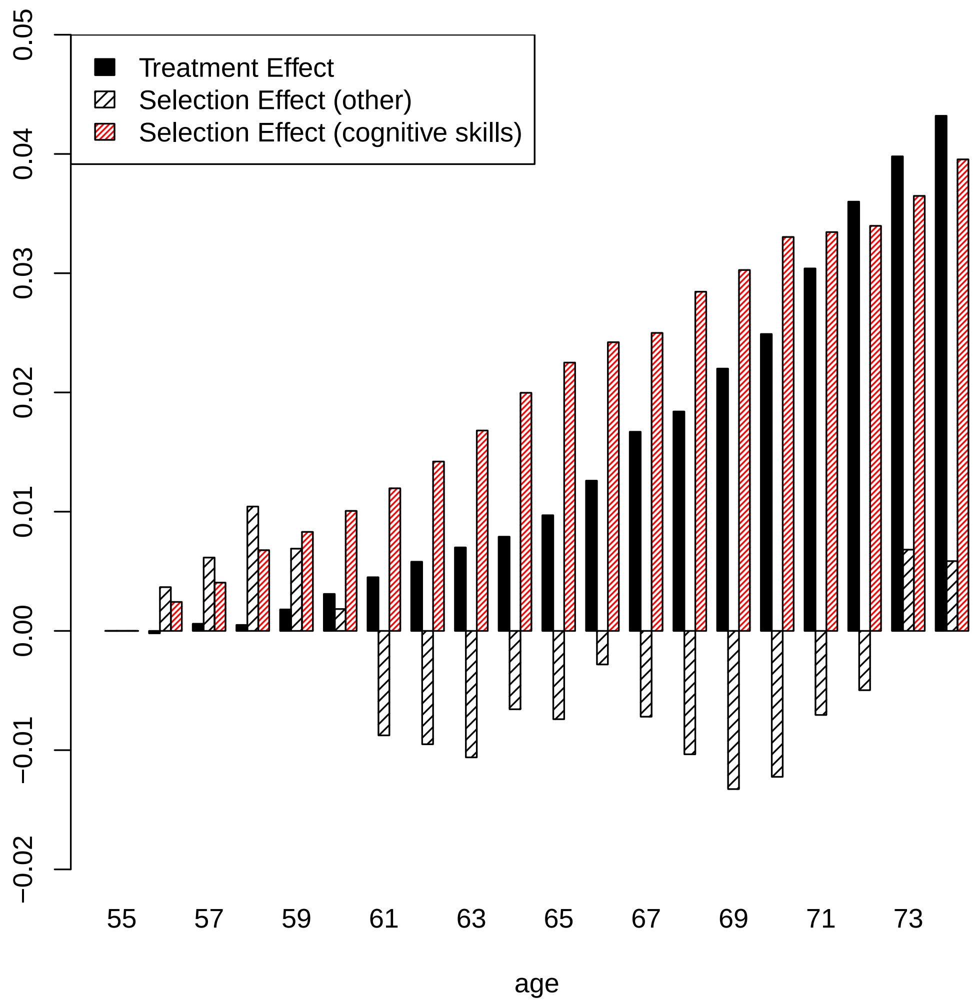

Figure 6: *Decomposition of observed difference in the Kaplan-Meier Survival function in a treatment (ATE) and selection effect due to cognitive skills and other selection effects, with binary education variable and two measurements for cognitive ability*

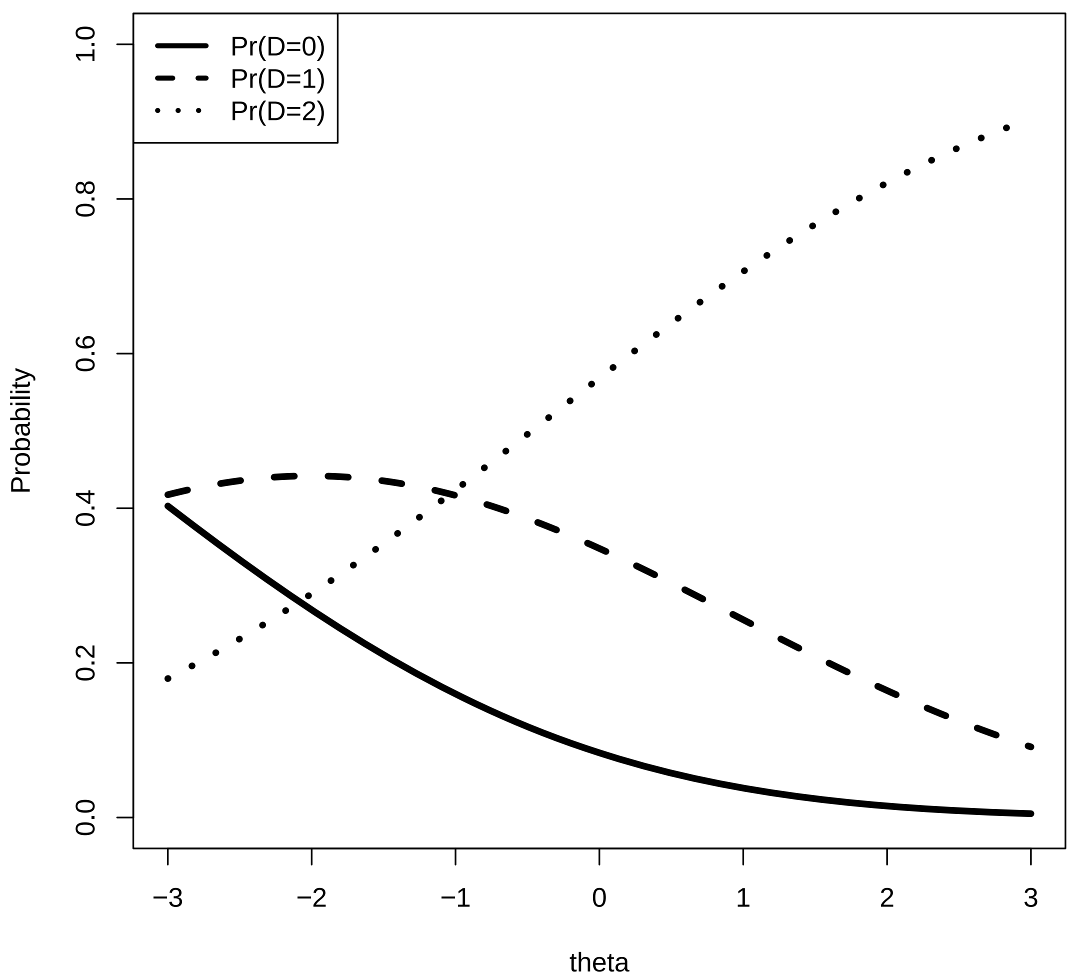

Figure 7: *Relationship between cognitive ability and the ordinal measure for education.*

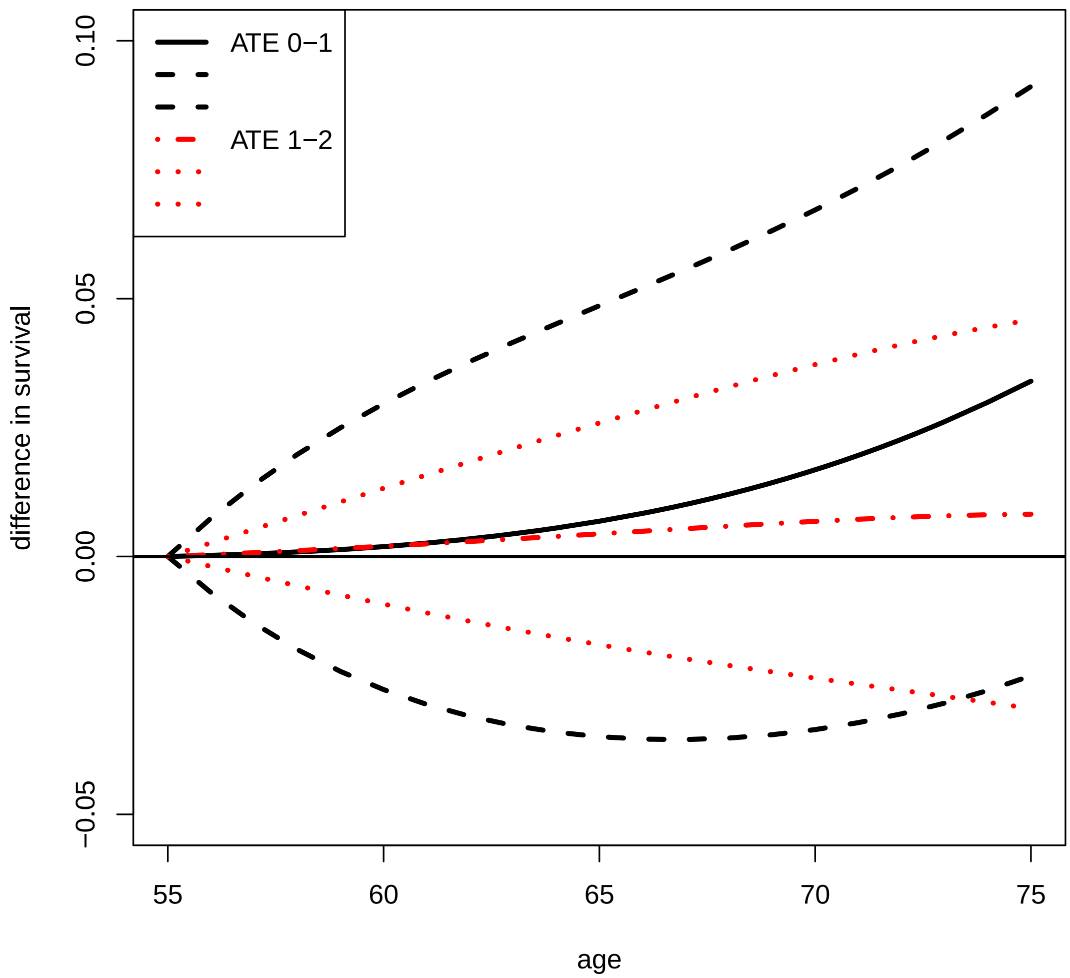

Figure 8: *Treatment effects by age, ordinal education variable, two measurements for cognitive ability. Dashed lines indicate the 90 percent confidence intervals.*

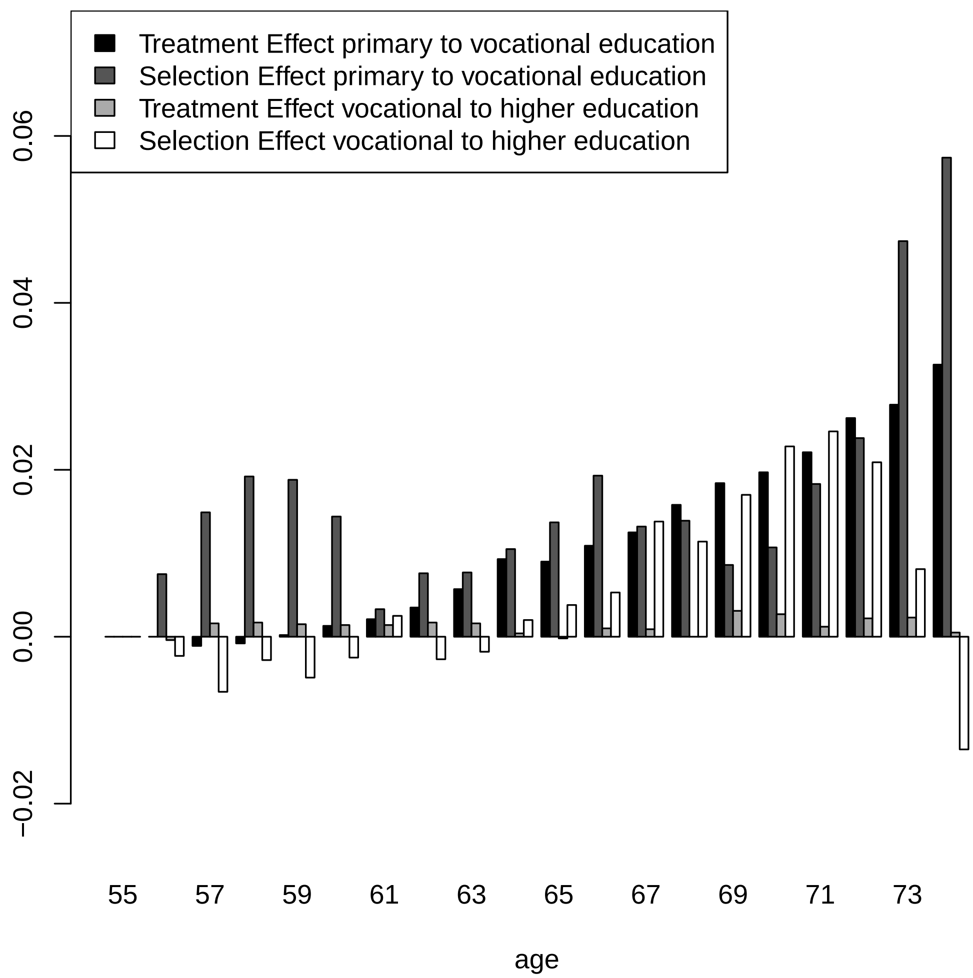

Figure 9: Decomposition of observed difference in the Kaplan-Meier Survival function in a treatment (ATE) and selection effect, with ordinal education variable and two measurements for cognitive ability

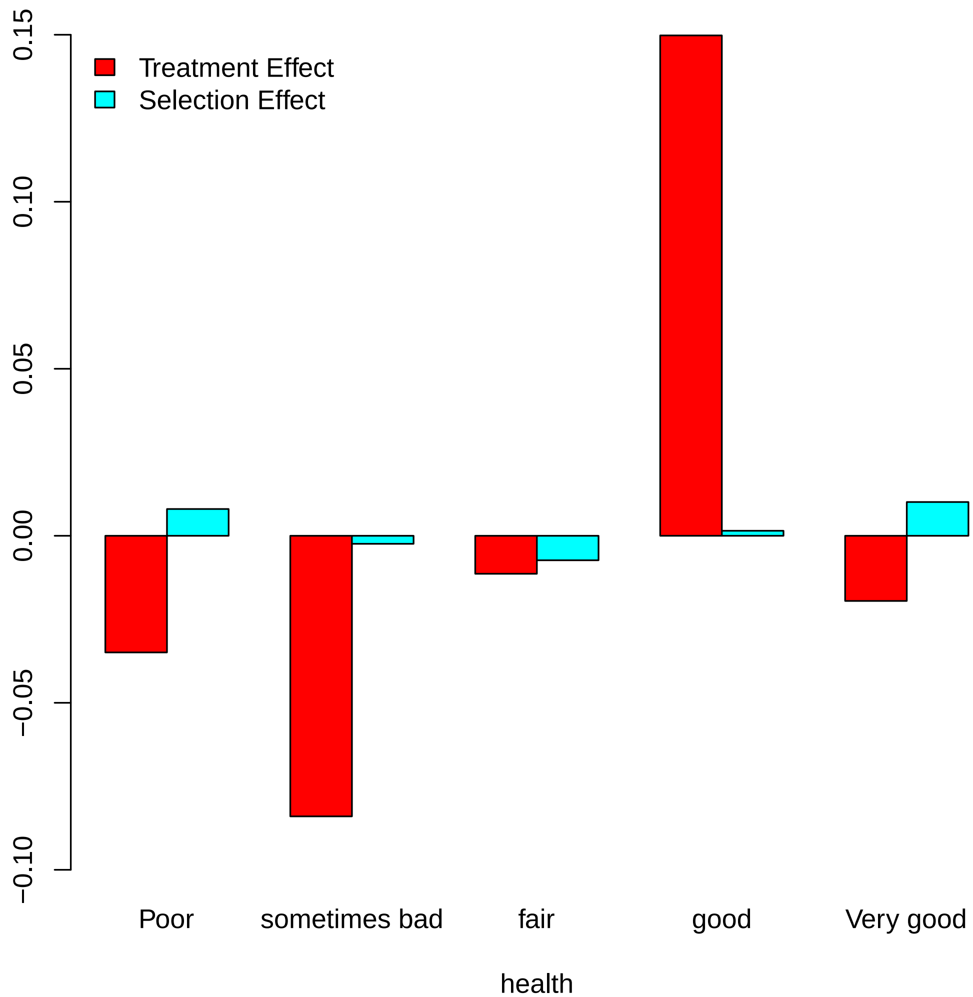

Figure 10: *Decomposition of observed difference in the self-reported health in a treatment (ATE) and selection effect, with binary education variable and two measurements for cognitive ability*

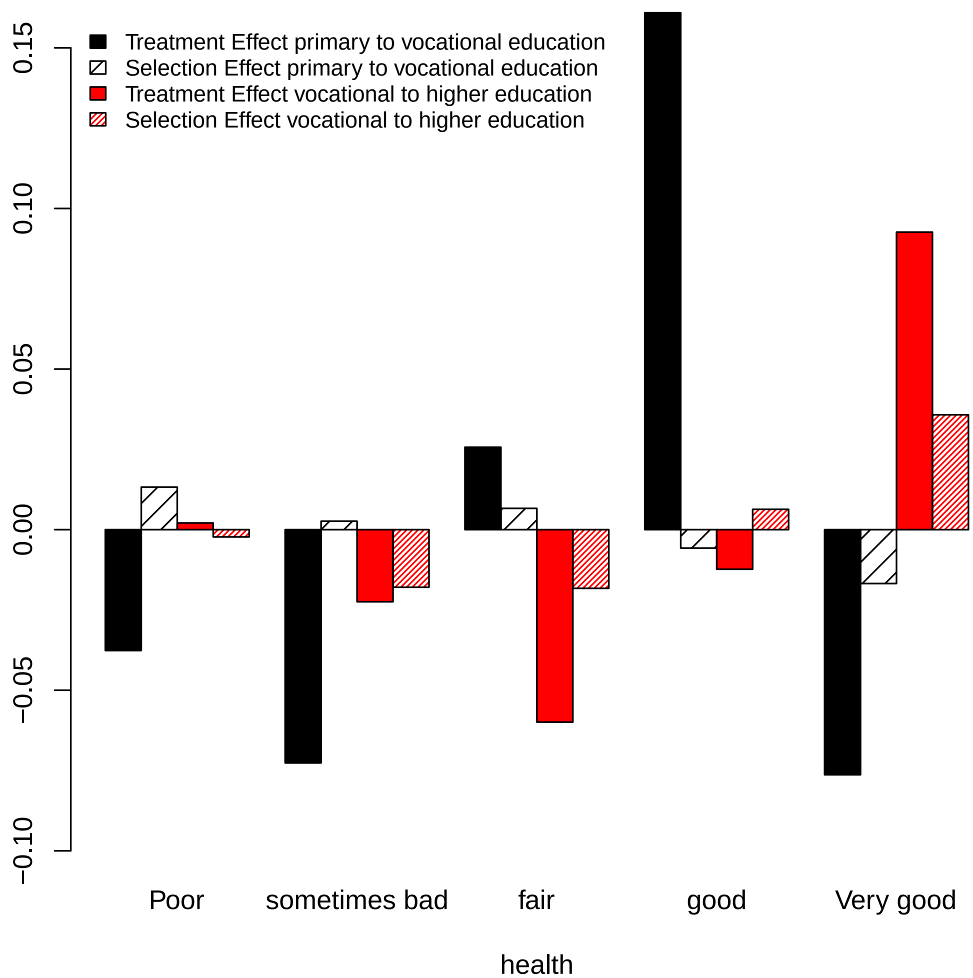

Figure 11: *Decomposition of observed difference in the self-reported health in a treatment (ATE) and selection effect, with ordinal education variable and two measurements for cognitive ability*

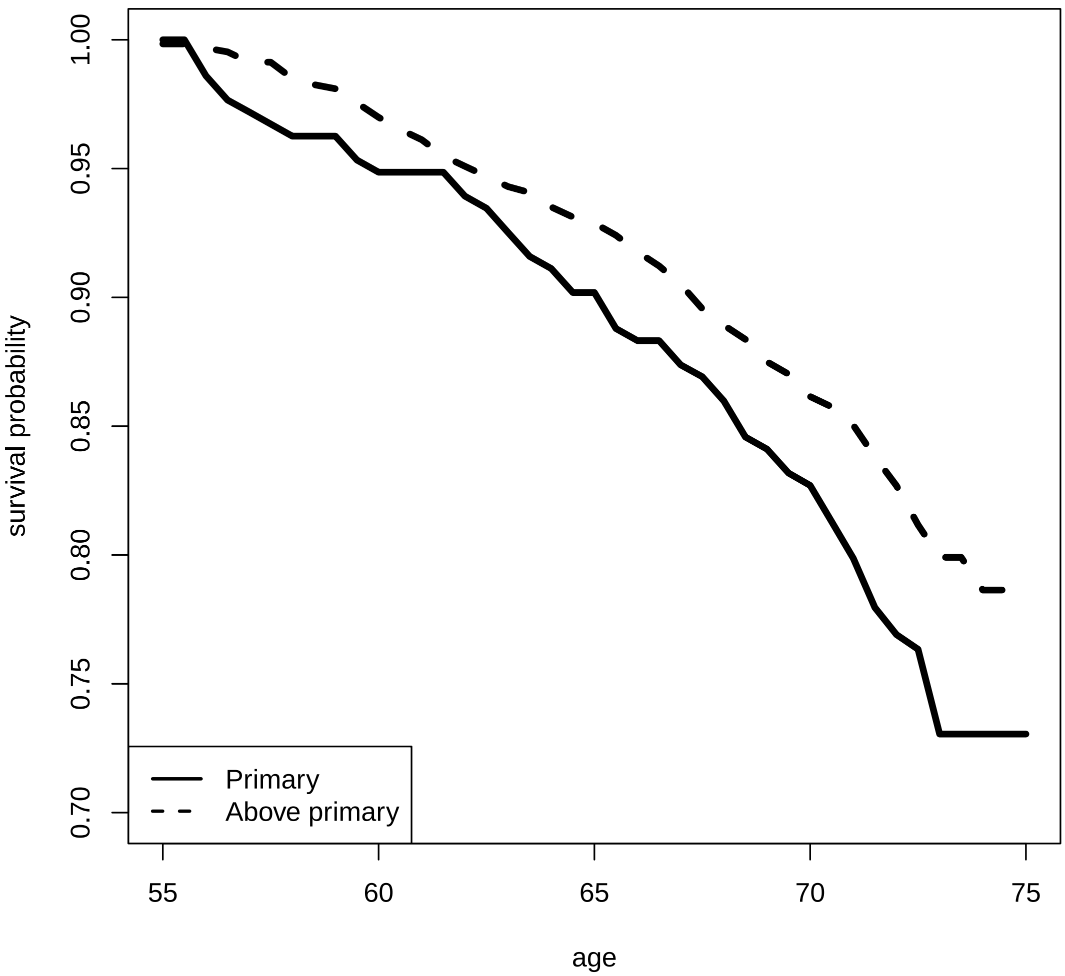

Figure 12: Kaplan-Meier Survival function by education level in two categories, males.

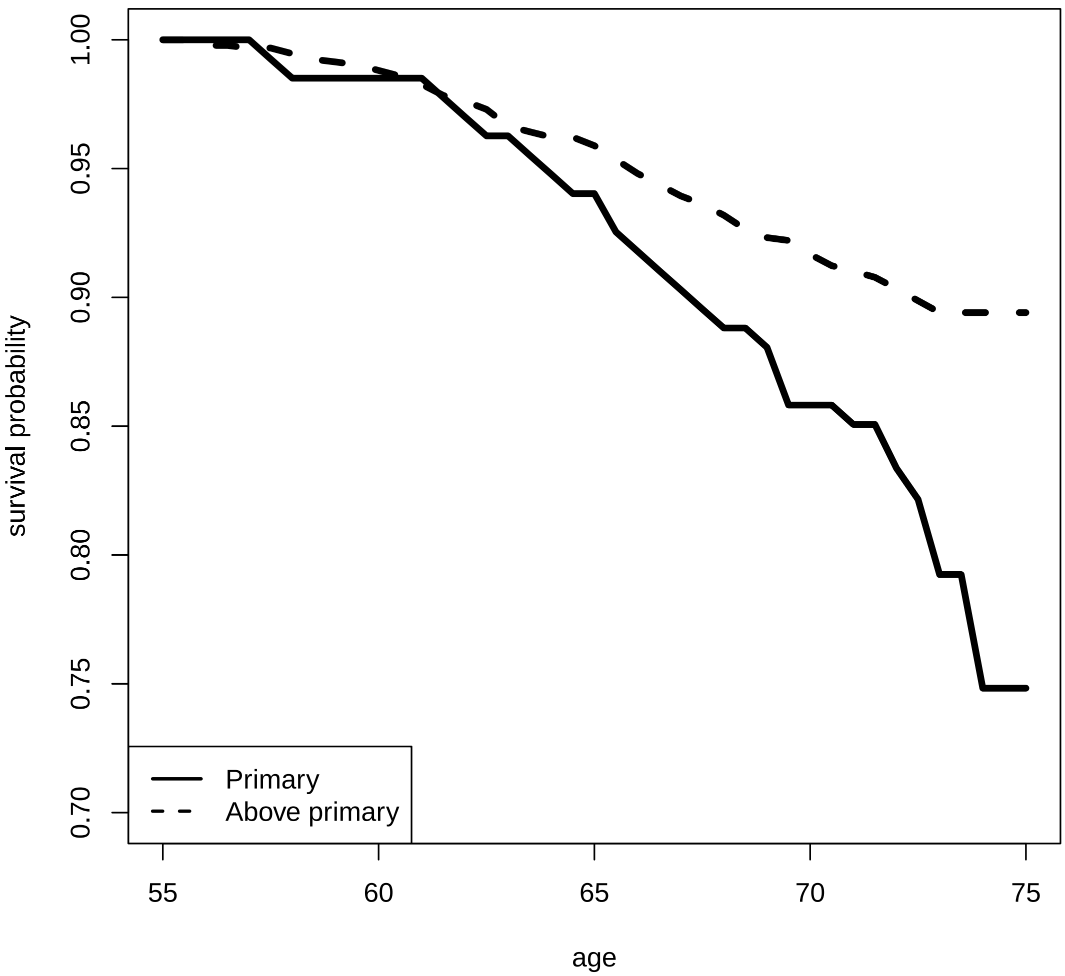

Figure 13: Kaplan-Meier Survival function by education level in two categories, females.

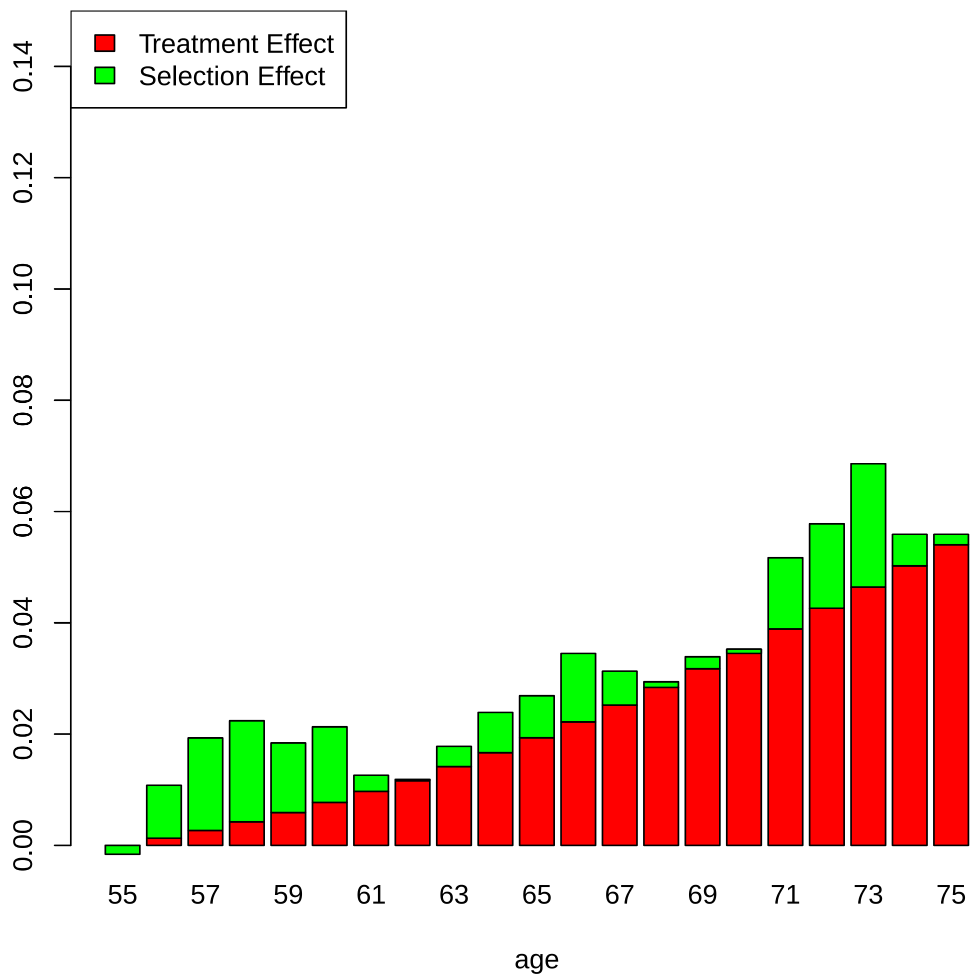

Figure 14: *Decomposition of observed difference in the Kaplan-Meier Survival function in a treatment (ATE) and selection effect, with ordinal education variable and two measurements for cognitive ability, males.*

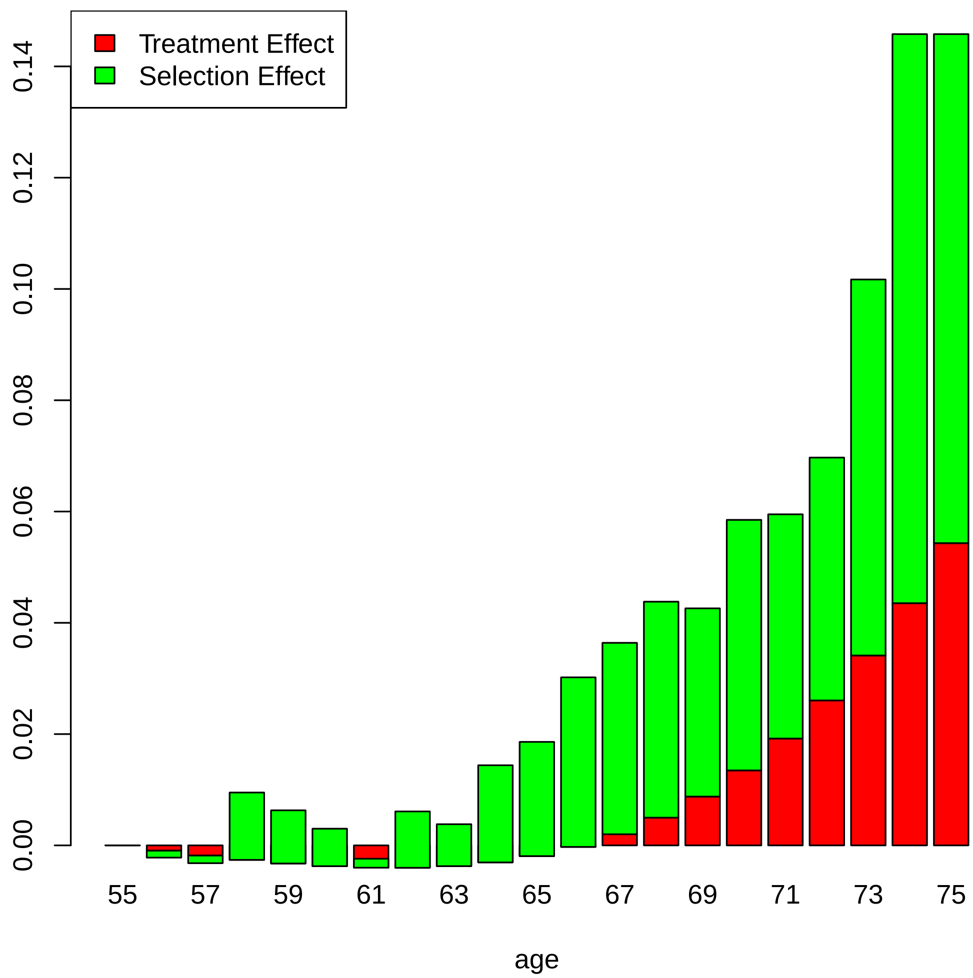

Figure 15: *Decomposition of observed difference in the Kaplan-Meier Survival function in a treatment (ATE) and selection effect, with ordinal education variable and two measurements for cognitive ability, females.*