Published in final edited form as:

Soc Sci Med. 2013 November ; 97: . doi:10.1016/j.socscimed.2013.08.019.

The socioeconomic health gradient across the life cycle: what role for selective mortality and institutionalization?

Steeff Baeten^a, Tom Van Ourti^{b,*}, and Eddy van Doorslaer^{a,b}

^aInstitute for Health Policy and Management, Erasmus University Rotterdam

^bErasmus School of Economics, Erasmus University Rotterdam, Tinbergen Institute, and NETSPAR

Abstract

Several studies have documented the now fairly stylized fact that health inequalities by income differ across the age distribution: in cross-sections the health gap between rich and poor tends to widen until about age 50 and then declines at higher ages. It has been suggested that selective mortality and institutionalization could be important factors driving the convergence at higher ages. We use eight waves of a health survey linked to four registries (on mortality, hospitalizations, (municipal) residence status and taxable incomes) to test this hypothesis. We construct life cycle profiles of health for birth year/gender/income groups from the health surveys (based on 128,689 observations) and exploit the registries to obtain precise estimates of individual probabilities of mortality and institutionalization using a seven year observation period for 2,521,122 individuals. We generate selection corrected health profiles using an inverse probability weighting procedure and find that attrition is indeed not random: older, poorer and unhealthier individuals are significantly more likely not to survive the next year and to be admitted to an institution. While these selection effects are very significant, they are not very large. We therefore reject the hypothesis that selective dropout is an important determinant of the differential health trajectories by income over the life course in the Netherlands.

Keywords

SES health gradient; selective mortality; institutionalization; inverse probability weighting; life cycle profiles; the Netherlands

Introduction

The evolution of health disparities by socioeconomic status (SES) over the life cycle has been studied by scholars from a variety of disciplines (Beckett 2000, Lynch 2003, Case and Deaton 2005, Smith 2007, House et al. 2005, Sacker et al. 2005, Herd 2006, Mirowsky and Ross 2008). Two hypotheses are generally put forward on how this SES health gradient evolves over the life cycle: *cumulative advantage* and *ageas-leveler*. Both hypotheses assume that health deteriorates with age and that the rate of deterioration is steeper for low than high SES individuals up to late middle age. Proponents of cumulative advantage argue

© 2013 Elsevier Ltd. All rights reserved.

*Corresponding author: P.O. Box 1738, 3000 DR Rotterdam, the Netherlands vanourti@ese.eur.nl.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

that health differences between SES groups keep widening until the late stages of life, while age-as-leveler states that health disparities start converging from late middle age onwards. A potential explanation for such a converging trend is that the biological determinants of ageing (and thus health deterioration) start dominating the influence of SES after late-middle age (Herd 2006), with most prominence in the literature so far being given to explanations based on cohort effects (Lynch 2003) and selective mortality (Beckett, 2000).

In this paper we focus on the role of selective mortality and institutionalization – mortality differences and differences in institutionalization rates between high and low SES groups – in shaping the life course profile of the SES health gradient. Selective mortality seems a likely candidate as an explanation for the converging SES health gradient after late-middle age as relative mortality differences have been found to peak around ages 50-60 and to fall again at higher ages (van Kippersluis et al. 2010). When a high SES individual is more likely to survive than a low SES individual – even when both are equally (un)healthy – then at older ages, the high SES group will include a greater share of unhealthy individuals than the low SES group. Selective mortality is a general phenomenon that might occur in any population, but selective institutionalization can only arise when institutionalized individuals have not been sampled. When the institutionalized are excluded, as is the case in most surveys, one can treat both forms of selection alike from a methodological perspective, i.e. when the low SES and unhealthy are more likely to move into an institution than their equally unhealthy high SES counterparts, the SES health gradient would start falling around those ages at which individuals move to institutions like elderly and nursing homes. Both types of selective dropout may therefore contribute to the apparent narrowing of the health gradient above a certain age in cross-sectional evidence.

In this paper we add to the existing literature on this phenomenon in five respects. First of all, we propose a new method to control for selective attrition by combining an approach examining health levels at an aggregated group level (Deaton and Paxson 1998) with inverse probability reweighting at the individual level. This combination of methods has not featured in the literature on selective attrition and the SES health gradient before, but has been used in related fields (e.g. van Kippersluis et al. 2009). The aggregated approach consists of transforming the set of repeated cross-sections into a panel dataset of age-groups, thereby allowing for the identification of life cycle effects. Inverse probability weights (IPWs) are used to correct for selective attrition (see e.g. Jones et al. 2006, Weuve et al. 2012, Tchetgen Tchetgen et al. 2012a,b). They are derived from an individual's predicted mortality/institutionalization conditional on past individual characteristics such as age, SES and health status. It thus, and unlike the imputation method of Beckett (2000), considers the experiences of *both* survivors/non-attriters and decedents/attriters.

Secondly, we allow attrition to depend on the interaction between prior SES and prior health status. Again, this is a crucial feature and a major advantage of our approach, because selective mortality will matter most for the SES health gradient if and when the association between mortality and ill-health differs *across* SES groups. Instead, all prior studies have relied simply on interactions between age and SES in explaining differential health trajectories.

Third, our analysis has greater statistical power to study selective attrition as we link cross-sectional surveys to administrative registry data for a 7 year follow-up period. This allows for a rather precise estimation of the mortality and institutionalization probabilities due to the much larger sample size than what is typically available in surveys.

Fourth, we test and correct for selection effects due to residential long term care. While others, such as Beckett (2000), have typically studied the effect of *overall* dropout (including

mortality and institutionalization), we can separate these. The Netherlands is a particularly interesting country to study this phenomenon since Dutch elderly have relatively higher utilization rates of residential long term care than elderly in other OECD countries (OECD 2011): in 2009, almost 10 percent of the 65+ population was living in a residential long term care institution (Statistics Netherlands 2012).

Finally, we are the first to study the importance of selective mortality for the life cycle profile of the SES health gradient in the Netherlands. With the exception of one Canadian (Prus 2007) and two UK studies (Sacker et al. 2005, McMunn et al. 2009), all other studies have been based on US data.

Previous approaches to correct for selection

The combined influence of selective mortality, institutionalization and other sources of sample attrition was analyzed first by Beckett (2000) for the US. Exploiting changes between the 1982-1984 and 1992 waves of the NHANES I Epidemiologic Followup Study, she identified the overall effect of any type of attrition on the SES health gradient. She confirmed that health by SES differences widen till late middle age and converge thereafter, but also found the convergence not to be the result of attrition. Noymer (2001) criticized her approach for relying on counterfactual health levels of deceased that may be overestimated because they were imputed from the experience of the survivors only. Another criticism was that the counterfactual health levels differed by age, SES and their interaction, but did not account for the interaction between initial health and initial SES. It was therefore not too surprising that Beckett (2000) found no effect of selective mortality on converging SES health differences at older ages. In later work, Beckett and Elliott (2001) partially addressed both concerns by estimating mortality models on the sample of survivors and deceased and by conditioning on initial health status and its interaction with initial SES, but did not use these models to analyze the effect of selective mortality on the SES health gradient.

Another influential approach – mainly credited to Lynch (2003) – consists of analyzing SES-differences in the life cycle profile of health using a random coefficients model. It has generally found evidence in favor of converging SES health differences at older ages being driven to some extent by selective mortality (Sacker et al. 2005, Herd 2006, Mirowsky and Ross 2008). A crucial feature of this approach concerns the splitting of overall health variation into variation *between* and *within* individuals. The life cycle behavior of the SES health gradient is then derived from the within-individual variation only and, in contrast to Beckett's (2000) approach, the effect of selective mortality (and attrition) on the life cycle profile of the SES health gradient is only implicitly revealed, and imputation of counterfactual health levels for the deceased (or other attritors) is not required. The main intuition is that purging within-individual health variation from between-variation should eliminate the effect of selective mortality (and attrition), but this will only hold if the deceased/attritors are "missing at random". This assumption has recently been criticized by Petrie et al. (2011). They argue that "missing at random" contradicts the idea that death itself is an indication that health deteriorates more rapidly for the deceased than for the survivors, and will thus lead to a lower bound of the SES health gradient. They instead propose to impute a(n absolute) zero health level to the deceased individuals which is the opposite of 'missing at random' and will inevitably give rise to an upper bound of the SES health gradient.

While the importance of selective mortality and institutionalization for the SES health gradient has not been analyzed for the Netherlands before, van Kippersluis et al. (2010) do provide some evidence for the age-as-leveler hypothesis using self-reported health and different indicators for SES. They show that the health decline with age – as observed in a

cross-section – becomes less steep and even reverses after the age of 55 for low SES individuals. While the underlying mechanisms leading to this pattern have not been well understood, it has also been reported for the US by Smith (2004). Van Kippersluis et al. (2010) also find that cohort effects in the SES health gradient are absent in the Netherlands, thereby confirming earlier findings with other Dutch survey data (van Kippersluis et al. 2009).

Data and Methods

Data, variables and sample

Our primary source of self-assessed health data consists of representative samples of non-institutionalized Dutch individuals taken from eight annual cross-sectional Surveys of Living Conditions held between 1998 and 2005 (SLC hereafter). The main variables for our analysis are ill-health status and income. We define poor self-reported health as the bottom three (out of 5) response categories on the question “How good is your health in general?” (“very good”, “good”, “it’s ok”, “poor”, “very poor” in 2002-2005 and “very good”, “good”, “it’s ok”, “sometimes good, sometimes poor”, “poor” in 1998-2001), and use income quartiles as a measure of SES. We constructed the income quartiles from current disposable after-tax household income (corrected for the number of adults and children living in the household using the Statistics Netherlands equivalence scale (Siermann et al. 2004) and expressed in constant 1997 prices). We restrict our analysis to individuals aged 37 (very few die or become institutionalized before this age) to 84 in 1998 and follow these until 2005. The restriction at age 84 (followed up until 91) is a compromise between small cell sizes at advanced ages and the need to include sufficiently old individuals to study the effect of institutionalization (and mortality). This leaves a sample of 128,689 observations to construct the ill-health profiles. As the SLC is cross-sectional, we cannot observe (reasons for) attrition directly but we can do this by adding external information from four linked administrative data sources, i.e. (1) the Regional Income Survey (RIS), (2) the cause of death registry (CoD), (3) the Municipal Population Register (MPR) and (4) the Hospital Admissions registration (HA).

The RIS provides a longitudinal sample of annual income and tax records for one third of the Dutch population (more than 5 million individuals) and includes the same income information as in the SLC. We use all individuals in the RIS in 1998 in the relevant age range ($n=2,521,122$) and model their survival until 2005. The CoD registry, which provides date of death (if any) for all Dutch citizens, is used to construct a binary indicator of whether the individual was alive at 31 December of year t . The MPR includes information on the residential status of all individuals living in the Netherlands, and defines an ‘institutional household’ as a nursing home, a revalidation center, a psychiatric hospital, a long term care institution for the mentally ill or a prison. We use not living in an ‘institutional household’ as an indicator of whether the individual resides in a private household in year t . Since this indicator is not defined for individuals that died in 1999-2005, we remove these individuals from the analysis that corrects for selective institutionalization.

A correction for selective mortality should ideally account for differential survival patterns for those with similar levels of self-reported poor health, but different income levels (or vice versa). Ideally we would like to use the prior level of the same self-reported health variable as a predictor of future mortality or institutionalization. However, this is not possible because we only observe current (not past) self-reported health status in the SLC. While HA does not include the same self-reported health variable, it does record all hospital admissions in the Netherlands for all years $t=1998, \dots, 2005$. We construct an indicator of whether there was an overnight hospital stay during the previous calendar year and use it as a measure of prior health.

Construction of life cycle profiles of health

We use an aggregated group-approach (as in e.g. van Kippersluis et al. 2010) to construct life cycle profiles of health using a procedure involving 4 steps.

First, we categorize each cross-section of the SLC in groups defined by income quartile, sex and year of birth. Income quartiles are determined separately for each cross-section to rule out the impact of income growth and to avoid small cell sizes. For the same reason, we subdivide the income quartiles into only 12 birth-year intervals – born 1914-1917, 1918-1921, to 1958-1961.

Second, the average level of ill-health of each gender/birth-year/income group in each cross-section is computed as the proportion of respondents reporting to be in poor self-reported health.

Third, the previous steps provide us with a panel at the group level. For each gender/income category, we have 96 observations of average ill-health (i.e. 8 cross-sections times 12 birth-year intervals). Variation over time *within* birth year groups is used to estimate life cycle ill-health profiles, while variation *between* birth year groups allows to identify cohort effects. In step 3, we run a separate OLS regression per gender/income category on the 96 observations and model average ill-health as a function of birth year and age. By using dummies to represent birth year and age, one can separately identify life cycle and birth year effects in a very flexible way. We use 12 dummies for birth year and 26 dummies for age and find that the birth year dummies were jointly insignificant in 7 of the 8 OLS regressions (i.e. one per income quartile, separately for males and females). While additional assumptions are required to separately identify life cycle and cohort effects beyond a range of 8 years – the SLC's cover the period 1998-2005 –, the absence of cohort effects finding is in line with earlier work by van Kippersluis et al. (2010) who report no or very small cohort effects in the Netherlands between 1983 and 2000. Therefore, the analysis is based on regressions of average ill-health on age dummies only.

Fourth, we show the OLS-predicted average ill-health levels across the life cycle in a graph for each gender/income classification. We use a second degree Gaussian weighting kernel to smooth these lines over the life cycle. To avoid an overload of information, we only present the ill-health life cycle profiles for the lowest and highest income quartile.

Inverse probability reweighting of life cycle profiles of health

We adjust for selective mortality and institutionalization in the Netherlands by combining the grouping approach with IPWs (Jones et al. 2006, Weuve et al. 2012, Tchetgen Tchetgen et al. 2012a,b). The grouping approach reveals how average ill-health of each gender/birth year/income group changes when its members grow older, but it does not disentangle whether it is driven by changes in its composition (e.g. selective mortality and institutionalization) or by changes in the health status of its members. Using IPWs we can correct for these compositional changes (during 7 years) and thus remove the effect of selective dropout from the profiles. Any selective dropout prior to 1998 is not addressed by the IPW's, as no pre-1998 data were available.

The IPW of an observation in year t measures the inverse of this observation's probability to belong to the age range that was 37 to 84 in 1998. Hence, in the first cross-section of 1998, all observations have an IPW equal to 1 since those in the age range 37-84 constitute the group that we follow over time. From 1999 to 2005, observations' IPWs will deviate from 1 when there is selective dropout.

Correction for selective dropout in our aggregated group approach is obtained by including these IPWs into the stepwise approach explained above. In the first step, we construct income quartiles that are equal in size in terms of IPWs, and further subdivide based on birth-years. Next, we calculate the IPW-weighted average levels of ill-health of each gender/birth-year/income group. The intuition is that groups of individuals that are *initially* similar to those that drop out are inflated in order to keep the original size and composition of the gender/birth year/income group in 1998. The third and fourth step are similar to the approach without correction for selective dropout. The resulting ill-health life cycle profiles reflect the evolution of the (hypothetical) population without selective dropout.

Estimation of inverse probability weights

IPW construction for selective mortality and institutionalization is similar, but is done separately to examine whether attrition effects of mortality and institutionalization differ by age, income and prior hospitalization. In the remainder of this section we therefore term both forms of selective dropout as selective survival.

We do not use the SLC, but estimate IPWs from the RIS as it follows individuals over time and since its large sample enables very precise estimation of the IPWs. When the same individual is observed repeatedly in a longitudinal setting, one can explicitly model selective survival using a binary dependent variable model. The traditional IPW-approach models the probability that each individual present in the first wave ($t=1$) is still present in a future wave t as a function of explanatory variables in the first wave ($t=1$). The same model (on all individuals in wave 1) is run for each future wave $t=2, \dots, T$. IPWs in wave t are obtained as the inverse of the predicted probabilities resulting from the binary dependent variable model for wave t . As in Beckett (2000) and Lynch (2003), this IPW-approach thus assumes that the unexplained part of selective survival is ‘missing at random’, but – in contrast to the imputation approach – it uses variation from *both* attritors and non-attritors; and unlike the random coefficient model, it allows for a more flexible and less parametric correction for selective mortality. An alternative approach – discussed in Wooldridge (2002) – is to model yearly survival probabilities when selective survival is an absorbing state. In practice one regresses survival in wave t as a function of explanatory variables in the previous wave $t-1$, conditional on those individuals that have survived up to the year $t-1$. In other words, this is equivalent to a procedure of ‘cumulative’ binary dependent variable models that first estimates survival from year 1 to 2, next from year 2 to 3, and so on until year T . The IPWs for year t are then obtained as the product of the IPWs of the binary dependent variable models for year 2 until year t . The main advantage of this approach is that the dynamics in the explanatory variables are used to predict survival. This should lead to a better explanation of actual survival, and thus makes the assumption of ‘missing at random’ more plausible.

We model yearly survival probabilities using the approach of ‘cumulative’ binary choice models by estimating 14 logit models for survival up to year t as a function of income and hospitalization in year $t-1$, and age in 1998. Age is included using one-year age dummies, and income is defined in absolute rather than relative terms. If we were to use a relative income concept (as we do for the ill-health profiles), survival from year $t-1$ to t would depend on the income quartiles in year $t-1$, but also on potential bias in these income quartiles arising from selective mortality between 1998 and $t-1$. This would require a correction of the income quartiles in year $t-1$. Instead an absolute income concept remains invariant over time and is therefore not prone to this additional form of selectivity bias. We define four income groups i.e. €0-€2,500; €2,500-€7,500; €7,500-€30,000 and €30,000 and higher. We do not include interactions between age and hospitalization since these proved jointly insignificant, suggesting that the selectivity impact of hospital stays is similar for younger and older individuals. We do include interactions between each income group

and hospitalization, but for the interaction between age and income we combine some age dummies to allow for the limited variation in mortality at relatively young ages: younger than 64, 65-69, 70-74 and one-year age dummies above 75. We also tested but found no statistical evidence for a three-way interaction between age, income and hospitalization. IPWs for each year t are obtained as the product of the IPWs of the logit models for years 2 to t . The precision of the resulting IPWs is high owing to the large sample size.

Linking estimated inverse probability weights to the SLC

The final step projects the IPWs estimated for every individual in the RIS onto the individuals in the SLC. This would be straightforward if every individual in the SLC would also be included in the RIS. However, as the RIS is a representative sample of around one third of the total Dutch population, while the SLC is a much smaller sample of the same Dutch population, we do not have an exact link to the RIS for (around) two thirds of SLC respondents. Therefore, an additional procedure was required to assign the IPWs from RIS to *all* individuals in the SLC. For every gender and year $t=1999, \dots, 2005$, we ran an *auxiliary* OLS regression (on all RIS individuals that survived until year t) of the *predicted* probability to survive up to year t (obtained from the logit models) on age, hospitalization in year $t-1$, and income in year t , and impose interactions between income and hospitalization, and income and age. We use current income as we do not observe previous year's income for the individuals that are in the SLC but not in the RIS. We also apply a logit transformation to the dependent variable before running the auxiliary OLS regressions since predicted probabilities always lie between 0 and 1. The estimated coefficients are then applied to all individuals in the SLC – which is justified due to the ‘missing-at-random’ assumption underlying IPWs –, and (after retransforming) provide us with the IPWs used to correct the ill-health life cycle profiles for selective mortality. As a sensitivity test, we also constructed selective-mortality-corrected ill-health life cycle profiles without this additional procedure by using only those individuals that are included in both the RIS and the SLC. While this smaller subsample results in less precise estimates, the resulting estimates generally confirm the results based on the IPWs obtained from the additional procedure.

Results

We start with the life cycle profiles of health across income groups before accounting for selective dropout. Next, we present the results of the selection models for mortality and institutionalization, and discuss the results of the auxiliary regressions to link the survival probabilities to the SLC. We then present the life cycle profile of health across income groups corrected for selective mortality and institutionalization.

Life cycle profiles of health by income without correcting for selective dropout

We first show the baseline life cycle ill-health profiles for different income quartiles without correcting for selective mortality and institutionalization. Summary statistics of the individual data are provided in table 1, and the baseline scenario of the lowest and highest income quartile, obtained from the individual data using the aggregated group approach explained in section “construction of life cycle profiles of health”, is presented in figure 1. We find that average health decreases for both income quartiles and genders from age 40 onwards, but more steeply for the lowest male quartile. Around ages 50-55, health of the lowest income quartile improves until around ages 60-65 while health levels of the highest income quartile evolve at a more constant rate, and the converging trend is more pronounced for males than females. At more advanced ages, the profiles are less smooth due to fewer observations in the birth year intervals. We also observe that average health among the highest income quartile is better for men than women, while the life cycle profiles of the lowest income quartiles cross twice.

Selective mortality and institutionalization

Table 2 provides summary statistics of the RIS data (with linkage to CoD, MPR and HA) which we use to obtain IPWs. We see that after one year (see column 1998) 1.18 percent of the 1,249,737 men and 0.72 percent of the 1,271,385 women of the 1998 RIS sample has dropped out due to mortality and this accumulates to 9.14 percent and 6.47 percent respectively in 2005. The corresponding 7-year dropout rates for institutionalization are 0.71 percent for males and 1.51 percent for females. We do not present detailed results of all (28) selection models (available upon request from authors), but discuss how these translate into IPWs after application of the auxiliary regressions.

The 14 ‘cumulative’ logit models for mortality show that older, poorer and hospitalized individuals have a significantly lower probability to survive the next year. This is in line with *a priori* expectations, but the interactions between age and income, and between income and hospitalization are more important for understanding the process of selective mortality. Our estimates suggest that income has a stronger positive survival effect for older individuals. We also find a hospitalization to particularly reduce survival chances of individuals in the lowest income group. This is confirmed by figure 2 which visualizes the magnitude of these interaction effects on the IPWs. By gender, it plots the relationship between age and the IPW’s in 1999 for 4 income groups interacted with having been hospitalized or not. The greater impact of hospitalization on mortality for those in the lowest income group is clear from the larger difference in IPW’s between those hospitalized or not. For 70 year old males, for example, the one-year survival difference between the hospitalized and non-hospitalized is 50% larger in the lowest than in the highest income group. The patterns in figure 2 show, however, that the selective mortality weights are not large, certainly not for those under age 70.

We find smaller effect sizes of income, age and hospitalization on institutionalization (than on mortality) and an insignificant interaction between income and hospitalization. Accordingly, figure 3 shows smaller IPW values compared to those of mortality in figure 2, and smaller absolute differences between the lines defined by income and hospitalization. Nevertheless, we still observe evidence of selective institutionalization: the effect of having been hospitalized seems to matter most for the lowest income group, even though the interaction between income and hospitalization was insignificant.

Corrected life cycle profiles of health by income

In this section, we demonstrate the effect of correcting the ill-health life cycle profiles in figure 1 for selective mortality and institutionalization. Given the findings in figures 2 and 3, we would expect the ill-health life cycle profile of the lowest quartile to shift upward after correcting for selective dropout, and that of the highest income quartile to remain the same (or slightly shift down). This would imply that the convergence of the ill-health levels of the rich and poor would turn out weaker. For selective mortality, we also expect (a) the effect to be larger at higher ages –especially for the lowest quartile – because of the interaction between age and income; and (b) the effect to be larger for the lowest quartile since this group has poorer health (as proxied by hospitalization) which leads (for the same income) to a lower survival probability. Given the lower absolute values of the IPWs for institutionalization and the insignificance of the interaction between income and hospitalization, we expect a more limited correction of the ill health life cycle profiles.

Figure 4a (males) and 4b (females) reproduce the ill-health life cycle profiles reported in figure 1 and compare these with the profiles resulting after correction for selective mortality. Remember that the IPW corrections matter for the construction of both the income quartiles and the average ill-health levels per income quartile. For males, we find an effect in the

expected direction from age 50 onward for the lowest income quartile, except around the age of 63. We hardly see any effect for the highest income quartile. For men the maximal effect is reached around the age of 80 where it increases the gap between the lowest and highest income quartiles by around 13 percent (i.e. the difference between the uncorrected (dashed) lines is 10.5 percentage points while after correction the difference amounts to 11.9 percentage points, which is a 13.3% increase). For females, the effects are more limited and basically show that selective mortality is unimportant to explain income differences in the evolution of female health. For women the maximal effect occurs at age 87: when the gap between the line for the highest and the one for the lowest income quartile is widened by a relative increase of 10.3%.

Figure 5a (males) and 5b (females) show the corresponding results for selective institutionalization. The effects are generally much smaller and tell us that selective institutionalization is an unimportant phenomenon which is most likely explained by the low number of institutionalizations compared to the number of deaths (see table 1) and the universal coverage for long term care in the Netherlands. A small effect is visible around the age of 75-80 for women in the highest income quartile, who seem to improve health after correction. This change is about 2.6% and occurs at the age at which many Dutch women move into nursing homes (see also figure 3b). However, this effect is not visible at older ages, which could indicate that women quickly die after entering a nursing home between 75 and 80 or just be a result of the small sample size at older ages.

Discussion and conclusion

This paper focuses on one of the possible explanations for a well-documented but not well understood pattern in the life cycle profiles of health by SES: differences in health across SES groups tend to diverge from young adulthood onwards and to widen until late middle age, after which these health differences start converging. Selective mortality and institutionalization could be responsible for the convergence at higher ages. We test for such selectivity and estimate to what extent it may explain the convergence of health trajectories by SES at higher ages.

The issue has been investigated by others, but we adopt a reweighting approach to correct for selection effects that overcomes some of the earlier deficiencies. We use a series of repeated cross-section surveys covering the period 1998-2005 to construct longitudinal aggregate data by birth year/gender/income groups to obtain life cycle profiles of health by income. Linked registry data are used to estimate individual probabilities of mortality and institutionalization in the seven year period for more than 2.5 million individuals to obtain precise estimates. Any selection biases due to mortality and institutionalization in the life cycle profiles by income are then corrected using inverse probability (re)weighting. The combination of four essential features distinguishes our approach from earlier ones: (a) the IPWs depend on the experience of both survivors/non-attriters and decedents/attriters; (b) we allow attrition to depend on the interaction between prior income and prior health status; (c) we allow for dynamics in (the effect of) the explanatory variables underlying the IPWs which makes the assumption of 'missing at random' more plausible; (d) we do not only study the possible confounding effects of selective mortality but also of selective institutionalization.

Our findings are as follows. First of all, we find evidence of both selective mortality and institutionalization at higher ages. Attrition is not random: older, poorer and unhealthier (i.e. hospitalized) individuals are significantly less likely to survive the next year and more likely to be admitted to a long term care institution. While all of these selection effects are statistically significant, they are not very large. For example, the probability of dying in the

first year is at the oldest ages 4 percentage points higher for the poor than the rich, but this difference is raised by less than 2 percentage points when poor health is accounted for. These are relatively small effects given that the probability of dying at the oldest ages within a year after a hospitalization is higher than 20%.

Second, and more importantly, we allow the selectivity to be not only health- and age-related but differentially so by income groups. We do this by including interactions between lagged health, age and income and find that high income has a stronger positive survival effect for older individuals. We also find that lagged health decreases survival chances most for the lowest income group. While these effects are relatively large, their absolute impact on survival is limited. If sufficiently large, this heterogeneity in effects may help explain the observed lifetime patterns.

Third, correcting for selective mortality and institutionalization adjusts the life cycle patterns of health by age and income in the expected direction, and increases the gap between rich and poor with maximally 13 percent. This reduces the convergence of the ‘uncorrected’ ill-health life cycle profiles between poor and rich. Among males, the effect is visible from the age of 50 onwards, but most pronounced around the age of 80. The same is generally true for females but the effects are much smaller and only emerge above age 75. Corrections for selective institutionalization are more modest at all ages mainly due to the lower average institutionalization than death rates and the smaller size of the income effects, but potentially also due to the institutional arrangements of long-term care in the Netherlands. As government provision and private providers of nurse and hospice care for old age might refrain individuals from being institutionalized, the impact of institutionalization may have different effects in countries where the organization of long-term care differs.

One legitimate concern that can be raised with respect to our analysis is that there is reporting heterogeneity by age in measures of self-reported health. It has been demonstrated that – conditioning on a more extensive and robust health measure like the Canadian Health Utility Index – older individuals are less likely to report themselves in poor health (Lindeboom and van Doorslaer, 2004). This may cause the observed decline in self-reported health to underestimate the decline as observed in other, more objective measures, as demonstrated by e.g. van Kippersluis et al (2009), also on Canadian data. Our concern here, however, is with the *differential* life cycle health patterns by income. There is no evidence that the age adjustment of health reporting occurs differentially by SES. In fact, Lindeboom and van Doorslaer (2004) cannot reject the hypothesis that self-reported health reporting is equal by income and education, after controlling for age and gender. This makes it unlikely that reporting heterogeneity is the cause of the convergence pattern (or of the shape of the life cycle profile of the lowest income quartile around ages 55-60)..

All in all, these findings imply that – despite the dropout being selective by both health and income – the healthy survivor phenomenon – and therefore the age as leveler hypothesis – cannot be the main explanation of the first diverging and then converging pattern in the low-to-high income difference in health by age. Other mechanisms must be responsible for its occurrence. One possibility is retirement. The relatively limited empirical evidence suggests that (early) retirement (temporarily) has a positive causal effect on the self-reported health of older Europeans (Coe and Zamarro, 2011) but it is unlikely that this effect can explain the observed health improvements in the cross-sectional profiles. Another possibility is reverse causality: rather than the health of those with high incomes deteriorating faster with age, it could be that lower health reduces income through individuals dropping out or reducing their participation in the labor force. Our data do not permit a proper test of this hypothesis, but recent work by García-Gómez et al. (2013) on labor exits in the Netherlands shows that an acute hospital admission lowers employment and personal income without subsequent

recovery, and the impact is larger at the bottom end of the income distribution. Their findings suggest that reverse causality – from health to income – is likely to be another important mechanism driving the income-health association which is worthwhile exploring to further unravel the mechanisms behind the hump shaped pattern of the life cycle pattern of the SES health gradient.

Acknowledgments

The authors thank Statistics Netherlands (CBS) for granting access to linked data resources (RIO 1998-2005, POLS Basis 1998-2005, LMR 1998-2005, and DO 1998-2005). This paper derives from the NETSPAR funded project “Health and income, work and care across the life cycle II”, and we also acknowledge support from the National Institute on Ageing, under grant R01AG037398. We have benefited from the comments and suggestions of Teresa Bago d’Uva, three anonymous referees, and participants of seminars at Erasmus University Rotterdam, and the LowLands Health Economics Study Group in Egmond aan Zee. We also thank Claudine de Meijer for help with the institutional background of residential long term care use in the Netherlands. The usual caveats apply and all remaining errors are our responsibility.

References

- Beckett M. Converging Health Inequalities in Later Life-An Artifact of Mortality Selection? *Journal of Health and Social Behavior*. 2000; 41(1):106–119. [PubMed: 10750325]
- Beckett M, Elliott M. Mortality and sample selection: reply to Noymer. *Journal of Health and Social Behavior*. 2001; 42(3):328–331.
- Case, A.; Deaton, A. Broken down by work and sex: How our health declines. In: Wise, DA., editor. *Analyses in the Economics of Aging*. Chicago University Press; Chicago: 2005. p. 185-205.
- Coe NB, Zamarro G. Retirement effects on health in Europe. *Journal of Health Economics*. 2011; 30(1):77–86. [PubMed: 21183235]
- Deaton A, Paxson C. Aging and Inequality in Income and Health. *The American Economic Review*. 1998; 88(2):248–253.
- García-Gómez P, van Kippersluis H, O’Donnell O, van Doorslaer E. Effects of health on own and spousal employment and income using acute hospital admissions. *Journal of Human Resources*. 2013 forthcoming.
- Herd P. Do Functional Health Inequalities Decrease in Old Age?: Educational Status and Functional Decline Among the 1931-1941 Birth Cohort. *Research on Aging*. 2006; 28(3):375–392.
- House JS, Lantz PM, Herd P. Continuity and change in the social stratification of aging and health over the life course: evidence from a nationally representative longitudinal study from 1986 to 2001/2002 (Americans’ Changing Lives Study). *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences*. 2005; 60(Special Issue 2):S15–S26.
- Jones A, Koolman X, Rice N. Health related non-response in the British Household Panel Survey and European Community Household Panel: using inverse-probability-weighting estimators in non-linear models. *Journal of Royal Statistical Society A*. 2006; 169(3):543–569.
- Lindeboom M, van Doorslaer E. Cut-point shift and index shift in self-reported health. *Journal of Health Economics*. 2004; 23:1083–1099. [PubMed: 15556237]
- Lynch S. Cohort and Life-Course Patterns in the Relationship between Education and Health: A Hierarchical Approach. *Demography*. 2003; 40(2):309–331. [PubMed: 12846134]
- McMunn A, Nazroo J, Breeze E. Inequalities in health at older ages: a longitudinal investigation of the onset of illness and survival effects in England. *Age and ageing*. 2009; 38:181–187. [PubMed: 19029098]
- Mirowsky J, Ross CE. Education and self-related health: cumulative advantage and its rising importance. *Research on Aging*. 2008; 30(1):93–122.
- Noymer A. Mortality Selection and Sample Selection: A Comment on Beckett. *Journal of Health and Social Behavior*. 2001; 42(3):326–327. [PubMed: 11668777]
- Organisation for Economic Co-operation and Development. *Help Wanted?: Providing and Paying for Long-Term Care*. OECD Publishing; Paris: 2011.

- Petrie D, Allanson P, Gerdtham U-G. Accounting for the dead in the longitudinal analysis of income-related health inequalities. *Journal of Health Economics*. 2011; 30:1113–1123. [PubMed: 21820193]
- Prus S. Age, SES, and health: a population level analysis of health inequalities over the lifecourse. *Sociology of health and illness*. 2007; 29(2):275–296. [PubMed: 17381817]
- Sacker S, Clarke P, Wiggins RD, Bartley M. Social Dynamics of Health Inequalities: A Growth Curve Analysis of Aging and Self Assessed Health in the British Household Panel Survey 1991-2001. *Journal of Epidemiology and Community Health*. 2005; 59(6):495–501. [PubMed: 15911646]
- Siermann, C.; van Teeffelen, P.; Urlings, L. Equivalentiefactoren 1995-2000. Centraal Bureau voor de Statistiek; Voorburg/Heerlen: 2004.
- Smith J. Unraveling the SES-health connection. *Population and Development Review*. 2004; 30(Supp 4):108–132.
- Smith J. The impact of Socioeconomic status on health over the life course. *Journal of Human Resources*. 2007; 42(4):739–764.
- Statistics Netherlands. [accessed on November 21 2011] Statline. 2011. <http://statline.cbs.nl/statweb/>
- Tchetgen Tchetgen EJ, Glymour MM, Shpitser I, Weuve J. To weight or not to weight? On the relation between inverse-probability weighting and principal stratification for truncation by death. *Epidemiology*. 2012a; 23(1):132–137.
- Tchetgen Tchetgen EJ, Glymour MM, Weuve J, Robins J. Specifying the correlation structure in inverse-probability-weighting estimation for repeated measures. *Epidemiology*. 2012b; 23(4):644–646. [PubMed: 22659551]
- van Kippersluis H, Van Ourti T, O'Donnell O, van Doorslaer E. Health and income across the life cycle and generations in Europe. *Journal of Health Economics*. 2009; 28(4):818–830. [PubMed: 19651452]
- van Kippersluis H, O'Donnell O, van Doorslaer E, Van Ourti T. Socioeconomic differences in health over the life cycle in an Egalitarian country. *Social Science & Medicine*. 2010; 70(3):428–438. [PubMed: 19897296]
- Weuve J, Tchetgen Tchetgen EJ, Glymour MM, Beck TL, Aggarwal NT, Wilson RS, Evans DA, Mendes de Leona CF. Accounting for bias due to selective attrition: the example of smoking and cognitive decline. *Epidemiology*. 2012; 23(1):119–128. [PubMed: 21989136]
- Wooldridge J. Inverse probability weighted M-estimators for sample selection, attrition, and stratification. *Portuguese Economic Journal*. 2002; 1(2):117–139.

- We provide evidence on the evolution of income related health inequalities over the life cycle
- Selective dropout over the life cycle can lead to an underestimation of health inequalities
- We distinguish between selective dropout due to mortality and institutionalization
- 8 Years of registry data for over 2.5 million individuals are used to estimate selective dropout
- We use aggregated group inverse probability weights to correct for selective dropout

Figure 1.
Baseline life cycle patterns of poor health for men and women of the highest and lowest quartiles in the Netherlands (1998-2005)

Figure 2.

Average IPWs due to selective mortality between 1998 and 1999 for men and women by income and hospitalization

Figure 3.

Average IPWs due to selective institutionalization between 1998 and 1999 for men and women by income and hospitalization

Figure 4.
Mortality-corrected patterns of poor health for men and women over the life cycle in the Netherlands (1998-2005)

Figure 5.
Institutionalization-corrected patterns of poor health for men and women over the life cycle in the Netherlands (1998-2005)

Table 1

Summary statistics of the individual data underlying quartile 1 and 4 in the SLC sample

Income quartile		1998	1999	2000	2001	2002	2003	2004	2005
1	% poor health	44%	46%	45%	42%	44%	44%	44%	49%
	mean income	€10,845	€9,926	€10,372	€13,487	€12,935	€13,097	€12,821	€13,079
	median income	€11,558	€10,761	€11,268	€14,145	€13,619	€13,806	€13,506	€13,682
	iqr income	3,678	3,831	3,775	3,915	4,098	4,043	3,856	3,886
Males		N	4,203	1,912	1,038	1,186	1,006	1,003	434
4	% poor health	13%	14%	13%	16%	16%	17%	18%	18%
	mean income	€45,729	€43,977	€47,481	€50,540	€48,718	€49,355	€47,337	€51,866
	median income	€38,302	€37,242	€38,691	€44,594	€43,685	€44,213	€43,426	€43,769
	iqr income	13,442	12,223	12,969	13,173	13,142	13,189	11,715	15,574
		N	6,033	2,784	1,320	1,464	1,334	1,362	598
1	% poor health	45%	49%	46%	45%	46%	47%	48%	52%
	mean income	€10,262	€9,794	€10,239	€12,972	€12,459	€12,762	€12,610	€12,460
	median income	€10,668	€10,141	€10,445	€13,228	€12,691	€13,078	€13,035	€12,610
	iqr income	3,818	3,707	3,753	4,400	4,291	4,095	3,999	3,909
Females		N	6,832	3,096	1,503	1,689	1,587	1,544	665
4	% poor health	17%	20%	19%	21%	19%	20%	22%	19%
	mean income	€45,545	€43,832	€45,824	€51,487	€50,522	€49,564	€48,380	€51,094
	median income	€38,260	€37,247	€39,036	€45,483	€44,524	€44,485	€43,793	€44,846
	iqr income	13,373	12,422	13,385	14,756	14,411	13,155	13,253	15,709
		N	5,005	2,224	1,919	1,409	1,259	1,185	498

Note: iqr: inter quartile range

Table 2

Summary statistics of the RIS sample

		1998	1999	2000	2001	2002	2003	2004	2005
Males	Shared of total sample dropped out due to: ^a								
	Death within next year	1.18%	1.31%	1.40%	1.53%	1.61%	1.67%	1.75%	
	Moving to an institution	0.08%	0.09%	0.09%	0.10%	0.11%	0.17%	0.19%	
	Average age	52.7	53.5	54.3	55.1	55.8	56.6	57.4	
	Share in income group:								
	€0-€12,500	24%	23%	20%	16%	14%	14%	15%	20%
	€12,500-€17,500	31%	31%	30%	28%	27%	28%	28%	32%
	€17,500-€30,000	31%	32%	33%	34%	35%	35%	34%	31%
	€30,000 and higher	14%	15%	16%	22%	24%	23%	23%	17%
	Share hospitalized	9%	9%	9%	9%	10%	11%	12%	
Females	Shared of total sample dropped out due to: ^a								
	Death within next year	0.72%	0.83%	0.94%	1.07%	1.19%	1.26%	1.36%	
	Moving to an institution	0.18%	0.18%	0.20%	0.22%	0.25%	0.32%	0.37%	
	Average age	53.1	53.9	54.8	55.7	56.5	57.3	58.2	
	Share in income group:								
	€0-€12,500	28%	26%	24%	19%	17%	17%	17%	24%
	€12,500-€17,500	30%	30%	29%	28%	28%	28%	29%	31%
	€17,500-€30,000	29%	30%	31%	32%	33%	33%	33%	29%
	€30,000 and higher	12%	13%	15%	20%	22%	22%	21%	16%
	Share hospitalized	10%	10%	10%	10%	11%	12%	13%	

^aNote: The total sample at the baseline year 1998 consisted of 1,249,737 men and 1,271,385 women