Demography (2018) 55:295–318
https://doi.org/10.1007/s13524-017-0635-6

Comparing Observed and Unobserved Components of Childhood: Evidence From Finnish Register Data on Midlife Mortality From Siblings and Their Parents

Hannes Kröger[1,2] · **Rasmus Hoffmann**[1] · **Lasse Tarkiainen**[3] · **Pekka Martikainen**[3,4,5]

Published online: 18 December 2017

Abstract In this study, we argue that the long arm of childhood that determines adult mortality should be thought of as comprising an observed part and its unobserved counterpart, reflecting the observed socioeconomic position of individuals and their parents and unobserved factors shared within a family. Our estimates of the observed and unobserved parts of the long arm of childhood are based on family-level variance in a survival analytic regression model, using siblings nested within families as the units of analysis. The study uses a sample of Finnish siblings born between 1936 and 1950 obtained from Finnish census data. Individuals are followed from ages 35 to 72. To explain familial influence on mortality, we use demographic background factors, the socioeconomic position of the parents, and the individuals' own socioeconomic position at age 35 as predictors of all-cause and cause-specific mortality. The observed part—demographic and socioeconomic factors, including region; number of siblings; native language; parents' education and occupation; and individuals' income, occupation, tenancy status, and education—accounts for between 10 % and 25 % of the total familial influence on mortality. The larger part of the influence of the family on

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s13524-017-0635-6) contains supplementary material, which is available to authorized users.

✉ Hannes Kröger
hkroeger@diw.de

[1] European University Institute, Florence, Italy

[2] Socio-economic Panel Study (SOEP), German Institute for Economic Research (DIW), Berlin, Germany

[3] University of Helsinki, Helsinki, Finland

[4] Max Planck Institute for Demographic Research, Rostock, Germany

[5] Centre for Health Equity Studies (CHESS), Stockholm University and Karolinska Institutet, Stockholm, Sweden

mortality is not explained by observed individual and parental socioeconomic position or demographic background and thus remains an unobserved component of the arm of childhood. This component highlights the need to investigate the influence of childhood circumstances on adult mortality in a comprehensive framework, including demographic, social, behavioral, and genetic information from the family of origin.

Keywords Mortality · Long arm of childhood · Siblings · Register data · Finland

Introduction

The influence of the family of origin on adult mortality has been established in many studies (Galobardes et al. 2008; Turrell et al. 2007). The common approach in estimating the social influence of the family is to take observed socioeconomic characteristics—such as parental education, occupation, or income—to predict the child's mortality. Within a life course approach, the effects of childhood on adult health outcomes and mortality are sometimes referred to as the "long arm of childhood" (Hayward and Gorman 2004). Socioeconomic position (SEP) in adulthood is, in this perspective, seen as an important mediator of childhood SEP as well as an independent predictor of mortality. Research in this tradition shows that people from disadvantaged social backgrounds in childhood have higher mortality and lower life expectancy, and that a considerable proportion of the effects of these early-life conditions is mediated by achieved social status (Pakpahan et al. 2017; Palloni 2006; Pudrovska and Anikputa 2014).

Midlife mortality is of relevance when assessing the importance of childhood because it is the first major period in which many individuals are no longer under the direct influence of their family of origin. Mortality differences in this age range, due to practical constraints defined in this study as deaths occurring between ages 35 and 72, are of particular interest from a health equity perspective and also as a focal point for social policy. Further, midlife is the period in the life course in which health is most stratified by social characteristics (House et al. 1994).

Theoretical models of previous studies have focused on narrow ranges of observable childhood characteristics, putting the spotlight on different features of childhood, depending on discipline and research question. However, health is increasingly viewed as influenced by complex interactions of individuals' social and biological conditions through the life course (Ben-Shlomo and Kuh 2002; Blane et al. 2013; Galea et al. 2010; Shanahan and Hofer 2005). In focusing only on observed characteristics of the family of origin, previous studies have often deliberately chosen more parsimonious models to reduce analytic complexity. To the best of our knowledge, we present here the first systematic attempt to give an estimate of the total familial influence on midlife mortality; decompose it into unobserved and observed family factors, and direct and indirect pathways; and relate it back to the total childhood influence on adult mortality. We use Finnish register data and a family approach to estimate how much of the total childhood influence on adult mortality can be explained with a parsimonious set of childhood socioeconomic and demographic variables as well as how much is mediated through adulthood characteristics in a pathway model.

Observed and Unobserved Parts of the Long Arm of Childhood

We propose that within the framework of childhood influences on adult mortality, the total effect of the long arm of childhood (CH_t) should be thought of as comprising two components. The first is the *observed component* (Δ_{obs}), widely investigated in previous studies; it is estimated by the joint influence of *observed measures* for SEP and family characteristics. Its counterpart is the component *of the arm* that is *not observed* (Δ_{uno}), reflecting the influence of unobserved childhood characteristics. This addition is crucial because the influence of childhood and the family of origin can extend far beyond the socioeconomic and demographic factors that are typically observed and used in studies implementing parsimonious models of the long arm of childhood.

We therefore define the total childhood influence as the sum of the observed and unobserved parts:

$$\begin{aligned} &CH_t = \Delta_{obs} CH_t + \Delta_{uno} CH_t \\ &\Delta_{obs} + \Delta_{uno} = 1 \\ &0 \le \{\Delta_{obs}; \Delta_{uno}\} \le 1 \end{aligned} \tag{1}$$

To obtain a better understanding of what is observed and unobserved in the study of childhood influences on midlife mortality, we draw on two complementary theoretical frameworks. The first divides childhood influences into four dimensions; in the second, we divide the influence of childhood into direct and indirect effects according to two life course perspectives. We apply the division into observed and unobserved factors to both approaches by employing family-based design. As shown later in the article, obtaining a direct estimate of Δ_{obs} and Δ_{uno} is impossible. The family approach is therefore necessary in order to enable indirect inferences about the two parts of the long arm of childhood.

The Four Dimensions of the Long Arm

First, adapting the approach of Pescosolido et al. (2008), we divide the influence of the long arm of childhood on health and mortality in factors attributable to fundamental cause theory ($\Delta_{FC_{ch}}$) (Link and Phelan 1995); stress process theory ($\Delta_{ST_{ch}}$) (Pearlin 1989; Szanton et al. 2005); social safety net theory ($\Delta_{SN_{ch}}$), including social support and coping (Pescosolido and Levy 2002; Turner et al. 2014); genetic influences (Δ_{GE}); and the interaction of all four dimensions ($g(FC_{ch},ST_{ch},SN_{ch},GE)$).

Consequently, the total influence of childhood on mortality can be defined as the additive components plus an unknown function of the interactions of the four dimensions:

$$\begin{aligned} &CH_t = \Delta_{FC_{ch}} CH_t + \Delta_{ST_{ch}} CH_t + \Delta_{SN_{ch}} CH_t + \Delta_{GE} CH_t + \Delta_{g(FC_{ch},ST_{ch},SN_{ch},GE)} CH_t \\ &\Delta_{FC_{ch}} + \Delta_{ST_{ch}} + \Delta_{SN_{ch}} + \Delta_{GE} + \Delta_{f(FC_{ch},ST_{ch},SN_{ch},GE)} = 1 \\ &0 \le \left\{\Delta_{FC_{ch}}; \Delta_{ST_{ch}}; \Delta_{SN_{ch}}; \Delta_{GE}; \Delta_{f(FC_{ch},ST_{ch},SN_{ch},GE)}\right\} \le 1. \end{aligned} \tag{2}$$

Although such a framework is necessarily a strong simplification, we can link most previous research to one or more of the four dimensions. Several studies have investigated parental education and occupation or the financial or material situation of the household (Agahi et al. 2014; Case and Paxson 2010; Elo et al. 2014; Hayward and

Gorman 2004; Link et al. 2017; Palloni 2006; Turner et al. 2016), which can clearly be thought of as representing factors attributed to fundamental cause theory (FC_{ch}).

Stressors (ST_{ch}) can be such measures as crowded housing (Falkstedt et al. 2011) or family experience of imprisonment or substance abuse (Kelly-Irving et al. 2013), as well as direct indicators of childhood health or illness that have been used previously (Case and Paxson 2010; Pakpahan et al. 2017; Palloni 2006). In particular, exposure to infectious diseases has been suggested as a specific mechanism for the effect of childhood health on adulthood health and mortality (Bengtsson and Lindström 2000, 2003; Dowd et al. 2009). These infections are expected to leave a direct (scarring) effect (Bengtsson and Broström 2009), with a negative impact in particular on cardiovascular-related mortality in later life, although some demographic evidence calls their importance into question (Gagnon and Mazan 2009). Another related pathway through which early-life stressors might negatively impact adult mortality is cognitive ability (Kuh et al. 2009).

Examples for the safety net (SN_{ch}) include family structure, such as the early death of parents (Campbell and Lee 2009), the influence of peer groups, or the strength of parent-child relationships (Andersson 2016).

The last part of the framework is genetic endowment (*GE*). Studies that have quantified the degree of heritability of longevity (based on twins studies) have concluded that between 15 % to 30 % of the variation in longevity may be due to genetic heritability, with another 25 % being the result of environmental factors that are fixed by age 30 (Beekman et al. 2013; McGue et al. 1993; Vaupel et al. 1998). The framework further acknowledges that genetic endowment always interacts with early-life social environment (*FC*; *ST*; *SN*), beginning in the *in utero* stage and extending postnatally. We can thus expect a complex interaction of genes and social environment to determine longevity. Such gene-environment interactions have been shown in health-related outcomes, such as smoking (Boardman 2009), physical activity (Aaltonen et al. 2016), and obesity (Boardman et al. 2014; Bouchard 2008; Qi and Cho 2008). These results speak against interpreting genetic family influences on mortality as a result of the purely mechanistic heredity of genes and in favor of finding possible evidence and explanations for gene-environment interactions and the related processes (Freese 2008; Freese and Shostak 2009). However, as defined earlier, we can also find environment-environment interactions—for example, if the influence of parental SEP is moderated by parent-child relationships (Andersson 2016).

The observed part of the long arm of childhood is therefore the explanatory power of the observed variables (*OV*) used to represent the four dimensions (*D*) making up the long arm of childhood (CH_t).

$$\Delta_{obs} = \sum_{D \in \{FC;ST,SN,GENE\}} \Delta_{OV_{D_{ch}}} + \Delta_{g\left(OV_{D_{ch}}\right)}. \tag{3}$$

The unobserved part is, by definition, a residual category and should be seen as a benchmark of the explanatory power that different approaches to the study of childhood circumstances and their relation to adult mortality provide. This component is therefore conditional on data, research focus, and the state of research in the field in general.

A Life Course Perspective on the Long Arm of Childhood

The second perspective important in the study of the long arm of childhood is the life course approach in the study of health, disease, and mortality. This perspective stresses the concepts of critical period, accumulation, and pathways through the life course (Ben-Shlomo and Kuh 2002). Childhood is a critical period when influences on the child can have scarring effects, leading to an underdevelopment of organs and the metabolic system that manifests itself only in an increased risk of (for example) cardiovascular disease (CVD) in midlife and consequently in a higher risk of premature mortality. The ideas of cumulative (dis)advantage (Dannefer 2003; DiPrete and Eirich 2006) and the pathway model can be seen as analogs to sociological models that link individuals' family of origin to their own socioeconomic status (SES) (Blau and Duncan 1967). Both accumulation and the pathway model suggest that early childhood disadvantages are translated into midlife (socioeconomic) disadvantages and might therefore have an increasing impact on mortality risk throughout the life course. To distinguish between the idea of critical period and the pathway model, we can divide childhood impact on adult mortality into a direct effect (Δ_{direct}), an indirect ($\Delta_{indirect}$) effect, and the effects of possible interactions between childhood and adulthood status ($g(CH, ADULT)$), signifying diverging development trajectories.

$$\begin{aligned} &CH_t = \Delta_{direct} CH_t + \Delta_{indirect} CH_t + \Delta_{g(CH,\,ADULT)} \\ &\Delta_{direct} + \Delta_{indirect} = 1 \\ &0 \leq \{\Delta_{direct}; \Delta_{indirect}\} \leq 1. \end{aligned} \tag{4}$$

Previous studies have investigated interaction patterns ($g(CH, ADULT)$) but have found them to have little or no effect on adult health or mortality compared with the critical period or pathway model (Hayward and Gorman 2004; Kröger et al. 2016; Mishra et al. 2009, 2013). To reduce the complexity of our study, we therefore disregard such a pattern of interaction for the remainder of the study and assume that $\Delta_{g(CH,\,ADULT)} = 0$. Many studies have estimated the extent to which childhood influences are mediated through adulthood characteristics ($\Delta_{indirect}$), often taking adulthood SEP, health behavior, or health status as ways that childhood influences mortality or adult health (Hayward and Gorman 2004; Link et al. 2017; Pakpahan et al. 2017).

A Family Perspective on the Long Arm of Childhood

To obtain an estimate of the observed and unobserved part of the long arm of childhood, we need to superimpose another approach on the domain-specific approach to childhood influences on adult mortality. In our study, we take a family perspective on the childhood effects, based on the assumption that family and family-related characteristics are the most important compound factors for determining mortality in adulthood. We therefore further define the total childhood influence as the sum of the shared family component (Δ_{fam}) plus the individual influences (Δ_i) specific to the individual and not shared within the family.

$$\begin{aligned} &CH_t = \Delta_i CH_t + \Delta_{fam} CH_t \\ &\Delta_i + \Delta_{fam} = 1 \\ &0 \leq \{\Delta_i; \Delta_{fam}\} \leq 1. \end{aligned} \tag{5}$$

The family part of the childhood influence (Δ_{fam}), which will be central to our analyses, can be divided into an observed and an unobserved part in a similar manner as the total childhood influence on mortality.

$$\begin{aligned} &\Delta_{fam} = \Delta_{fam,obs}\Delta_{fam} = \Delta_{fam,uno}\Delta_{fam} \\ &\Delta_{fam,obs} + \Delta_{fam,uno} = 1 \\ &0 \le \left\{\Delta_{fam,obs}; \Delta_{fam,uno}\right\} \le 1. \end{aligned} \tag{6}$$

As defined earlier, the observed part is the explanatory power that we obtain from the observed variables (OV) representing the four domains that make up CH_t. However, because we are calculating the observed part on the *family level,* only differences *between* families are taken into account.

$$\Delta_{fam,obs} = \sum_{D \in \{FC;ST,SN,GENE\}} \Delta_{fam,OV_{D_{ch}}} + \Delta_{fam,g\left(OV_{D_{ch}}\right)}.$$

We calculate the direct and indirect (mediated by adulthood characteristics) effects on the family level similar to the decomposition into indirect and direct effects on the observed variable ($OV_{D_{ch}}$) level.

$$\begin{aligned} &\Delta_{fam} = \Delta_{fam,direct}\Delta_{fam} = \Delta_{fam,\,indirect}\Delta_{fam} \\ &\Delta_{fam,direct} + \Delta_{fam,indirect} = 1 \\ &0 \le \left\{\Delta_{fam,\,direct}; \Delta_{fam,\,indirect}\right\} \le 1. \end{aligned} \tag{7}$$

Before we turn to the derivation of estimates for Δ_{fam} and its components, we discuss the kind of conclusions that we can draw from such estimates on their own, and then how this lets us draw conclusions about Δ_{obs}, Δ_{uno}, Δ_{direct}, and $\Delta_{indirect}$—the elements that make up the whole long arm of childhood that affects adult mortality (CH_t).

What We Can Learn From the Family Part of the Long Arm of Childhood

One of the key features of the family approach to the long arm of childhood is the possibility of linking insight from Δ_{fam} back to CH_t and its components (the original theoretical interest of our study). The kind of conclusions that we can draw depends on the assumption we make regarding the relationship between the components of Δ_{fam} and those of Δ_i. Because we cannot estimate Δ_i or its components directly, we cannot verify any of the assumptions, but we think it is helpful to divide them into three comprehensive scenarios listed in Table 1.

With the strong Assumption 1 stating that the proportion observed is equal for the individual and family components, our estimate of the family components is a direct estimate of the observed and unobserved components of the total childhood influence. The same argument holds if the direct effect is equal for the family and individual components.

If we assume that the observed component of the family part is *larger* than the observed component of the individual part (Assumption 2), we obtain an upper-bound estimate for the observed effect and a lower-bound estimate for the unobserved effect. Thus, the observed part of the total influence of the long arm of childhood cannot be larger than the observed family part and will be smaller to a certain degree. As a result,

Table 1 Relationship between family component and total childhood influence on adult mortality depending on different assumptions

	Assumptions	Possible Conclusions	
1	$\Delta_{fam,obs} = \Delta_{i,obs}$ $\Delta_{fam,indirect} = \Delta_{i,indirect}$	$\Delta_{obs} = \Delta_{fam,obs}$ $\Delta_{uno} = \Delta_{fam,uno}$ $\Delta_{indirect} = \Delta_{fam,indirect}$ $\Delta_{direct} = \Delta_{fam,direct}$	Direct estimates of Δ_{obs}, Δ_{uno}, and Δ_{direct} and $\Delta_{indirect}$
2	$\Delta_{fam,obs} > \Delta_{i,obs}$ $\Delta_{fam,indirect} > \Delta_{i,indirect}$	$\Delta_{obs} < \Delta_{fam,obs}$ $\Delta_{uno} > \Delta_{fam,uno}$ $\Delta_{indirect} < \Delta_{fam,indirect}$ $\Delta_{direct} > \Delta_{fam,direct}$	Upper-bound estimate for the observed Δ_{obs} and indirect effect $\Delta_{indirect}$ and lower-bound estimate for the unobserved and direct effect
3	$\Delta_{fam,obs} < \Delta_{i,obs}$ $\Delta_{fam,indirect} < \Delta_{i,indirect}$	$\Delta_{obs} > \Delta_{fam,obs}$ $\Delta_{uno} < \Delta_{fam,uno}$ $\Delta_{indirect} > \Delta_{fam,indirect}$ $\Delta_{direct} < \Delta_{fam,direct}$	Lower-bound estimate for the observed Δ_{obs} and indirect effect Δ_{direct} and upper-bound estimate for the unobserved and direct effect

for the total influence, the unobserved factors are even more important than for the family level.

Assumption 3 posits that the observed component of the family part is *smaller* than the observed component of the individual part. In this case, we get a lower-bound estimate for the observed effect and an upper-bound estimate for the unobserved effect.

All three assumptions apply equally to the relationship between the direct and indirect part of the long arm of childhood, with indirect being the equivalent to observed and direct being the equivalent to unobserved (see Table 1).

In sum, our intent is to estimate Δ_{fam}—and its components, defined in Eqs. (6) and (7)—to assess how much of the long arm of childhood can be observed and to determine how much of the effects of the long arm of childhood are direct effects and thus not mediated by adulthood characteristics.

When we assess what the family part (Δ_{fam}) of the long arm of childhood can teach us, it is also important to note what we are *excluding* with the choice of our focus. We are disregarding all influences from FC_{ch}, ST_{ch}, SN_{ch}, and GE on adult mortality that are not shared in the family but might be different for different members of the family. For example, parental investment of resources from FC_{ch} can vary between children in one family (Becker and Tomes 1976). First, illnesses might strike one sibling but not the other, leading to differential stress exposure (ST_{ch}). Second, family relations might be different, with parents having closer or weaker ties and support (e.g., related to birth order) for particular children (SN_{ch}). Last, genetic endowment varies by definition between siblings and between families. Together, these elements constitute important influences on mortality acquired in childhood that are individual-specific (Δ_i) and not family-specific, but these effects are disregarded when focusing on the family component.

Causes of Death

We stratify our analyses by groups of causes of deaths. Previous research has established that childhood circumstances are related in different ways to different

causes of death (Galobardes et al. 2004). We investigate whether this also holds true for the family component of the total influence of childhood. CVD (and related mortality) has often been argued to build up over the life course starting in childhood with both scarring (critical period) effects (Bengtsson and Lindström 2000, 2003) and the accumulation of risk factors (Davey Smith et al. 1997); this is also true for lung cancer, with smoking as a naturally cumulating behavioral risk factor (Lynch et al. 1997). In the age range under investigation (35–72), major groups of causes of deaths in Finland (in addition to CVD and cancer) are an almost equal proportion of combined deaths related to accidents, violence, or alcohol (see Table S2 in Online Resource 1). These causes warrant special attention because their development through the life course and the link to childhood might be more indirect and thus mediated by adulthood social risks and health behaviors.

Data and Methods

We use a 10 % sample from the Finnish 1950 census for our analyses. Statistics Finland linked the individuals to the death register between 1970 and 2007 using personal identification codes. Siblings are identified as persons aged 0–14 at the time of the 1950 census (birth cohorts from 1936–1950) and having the status of child in the same family. This excludes all siblings living in different households, orphans, and institutionalized children, and treats adopted children as full siblings. Identifying siblings this way is in line with the wider social notion of siblings—that is, being raised by at least one common parent in the same family—instead of a biological definition of siblings (although in the majority of cases, these definitions converge).

All surviving individuals are censored at the end of year 2007. Because there is no mortality information before 1970, the analyses exclude all deaths in early life (here, before age 35). This restriction reduces the age range from which we can draw inferences but avoids the problem of variation in left truncation that can create biased inference of the estimated parameters (Hoffmann 2008; van den Berg and Drepper 2015). This design also means that our results refer to midlife and early old-age mortality (deaths in the age range 35–72). Also excluded are individuals who emigrated before 1970. As a result, 15,065 of those individuals included in the 1950 census sample make no contribution to the mortality analysis, largely because of extensive emigration to Sweden in the 1960s. A prior study on the same data set have shown a minor overrepresentation of women, those born before 1945, individuals from low-SES backgrounds, and mother-only families in the sample (Elo et al. 2014). This bias is so small, however, that it is unlikely to affect our results. The sample includes 94,042 individuals nested in 32,544 families, yielding 2,598,805 person-years of analysis time.

We divide mortality into all-cause mortality and mortality due to (1) cancers of the lung, larynx, trachea, and bronchus (truncated as “lung cancer”); (2) other forms of cancer; (3) CDV; (4) alcohol-related deaths; and (5) accidents and violence-related deaths. Other cause-of-death categories do not include sufficient numbers of deaths in the data set to analyze them separately. Among the alcohol-related causes are alcoholic liver disease, accidental alcohol poisoning, alcoholic diseases of the pancreas, alcoholic cardiomyopathy, alcohol dependence syndrome, and other mental and behavioral disorders resulting from alcohol use. They are important causes of midlife (male)

mortality in Finland (Elo et al. 2014; Herttua et al. 2008; Tarkiainen et al. 2016). Accidents and violence include suicides, traffic accidents, poisoning (excluding alcohol poisoning), and homicide. The coding of causes of death in the Finnish death register, especially in broader categories such as these, has been shown to be reliable (Lahti and Penttilä 2001).

We arrange the factors explaining mortality differences between families into three categories. The first category contains demographic factors shown to be associated with mortality in Finland. This category includes native language (Swedish or Finnish), parental age at conception (Gavrilov and Gavrilova 2001; Hubbard et al. 2009; Myrskylä et al. 2014), number of siblings (Hart and Davey Smith 2003), and region of residence (Blomgren et al. 2004; Saarela and Finnäs 2009).

The second category contains information on parental SEP from the 1950 census and includes the highest educational level attained by both parents (no schooling, primary, or past primary education) as well as the occupational class of the father, categorized according to the Erikson-Goldthorpe-Portocarero scheme (EGP). If paternal information was not available, the occupational status of the mother was used. Further, housing conditions—measured as persons per heated room—are used as an indicator of the socioeconomic resources of the parents.

The third category of variables measures the individual's SEP at age 35. We use the highest educational degree of each sibling. The degrees are categorized into basic (ISCED classification 2011 code: 2), lower track of upper secondary (ISCED: 3), higher track of upper secondary (ISCED: 3–4), lowest and lower-level tertiary (ISCED: 5–6), and highest-degree tertiary (ISCED: 7–8). Again, occupational status is measured based on occupational coding comparable with the EGP class scheme. The categories used are employers and self-employed, upper white-collar workers, lower white-collar workers, and blue-collar workers. Tenancy status distinguishes between individuals who rent and those who own or partially own their home. Personal income before taxes is categorized into deciles for those who earn taxable income, plus a category for those who do not earn taxable income. This variable represents the relative income position in the year of the census closest to the year when the individual turned 35 and not necessarily the relative position within the sample.

Table S2 in Online Resource 1 shows the summary statistics for all variables used in the sample.

Because our sample excludes all only children, Table S4 in Online Resource 1 shows the differences between the sample of individuals from families with at least two siblings and the only-children who were thus excluded (21,902 individuals). With respect to the relevant characteristics, the samples are fairly similar. Singletons tend to have a higher probability of having Swedish as their native language and are slightly better educated than those who have siblings.

Statistical Approach

Identifying the Familial Influence on Mortality

As an identification strategy for Δ_{fam}, we estimate the variance of the shared frailty parameter based on a multilevel survival model that uses siblings nested within families. This approach of estimating total familial influence is widespread in the study

of SES transmission (Björklund and Jäntti 2012; Duncan et al. 2001; Solon et al. 1991) and has also been used in research on health inequalities (Johnson et al. 2012; Merlo 2011).

Next, we conduct a step-by-step introduction of sets of observed factors representing childhood and early adulthood conditions. Adding the observed demographic and socioeconomic characteristics of the parents to the basic model, we show how much of this total familial influence on mortality can be attributed to these observed characteristics (observed part of the family part of the long arm of childhood, $\Delta_{fam,obs}$) and how much of the familial influence is left unexplained (unobserved component of the family part of the long arm, $\Delta_{fam,no}$). The same approach applies to the introduction of the later-life SEP of a family's children into the model, which identifies the direct ($\Delta_{fam,direct}$) and indirect ($\Delta_{fam,indirect}$) family pathways.

Quantifying the Familial Influence on Mortality

We use the median hazard ratio (MHR) to quantify the total familial influence on mortality. The MHR is a relative measure of dissimilarity in mortality risk between families and is reported in the hazard ratio metric. In Online Resource 1, we report similar analyses for two other ways of estimating the total family influence—namely, equivalent years of aging and sibling similarity.

The MHR is based on the variance of the shared frailty parameter derived from a multilevel survival model (θ). The frailty parameter is shared between siblings, making families the higher-level (Level 2) units. *Frailty* is used here in the statistical sense of survival analysis, which takes variation between different levels into account (Hougaard 1995; Vaupel 1988; Vaupel et al. 1979; Wienke 2010); it is not a clinical indicator for health, as it is often used in aging research (Aalen et al. 2015; Gobbens et al. 2010).

We estimate a parametric survival model with an exponential distribution of the underlying hazard, the explicit introduction of analysis time (t) as a covariate, and a shared frailty parameter for siblings within the same family. Using exponential distribution and analysis time as covariate is equivalent to specifying a Gompertz distribution for the hazard:

$$h_t = \exp(a + b \times t).$$

The advantage of this approach is that it allows us to include both men and women in the model but still estimate the shape parameter of the Gompertz model separately for men and women, as is appropriate due to the much higher mortality risk of men in midlife.

In the proportional hazards metric, the model is defined as follows:

$$h_{fs}(t) = Z_f \times \exp\left(-\left(a + b_w \times t_{fs} + c \times male_{fs} + b_m \times t_{fs} \times male_{fs} + \mathbf{X}_{fs}\boldsymbol{\gamma}\right)\right)h_0(t).$$

Because the distributional assumption of this model is equivalent to the Gompertz distribution, its use for midlife mortality seems appropriate. See Online Resource 1 for a nonparametric test that supports this assumption.

Our central measure of familial influence on mortality is the MHR (Merlo et al. 2006), a measure of *dissimilarity between* groups. MHR is the average increase in mortality that would occur if a random individual from a random family were to be put in another higher-risk family. The MHR can be estimated based on the variance term on the family level (Merlo et al. 2006):

$$MHR = \exp\left(\sqrt{2 \times \theta} \times 0.6745\right).$$

Our baseline model includes only two variables: the birth cohort and gender of the individual. After estimating the baseline model, we introduce the individual- and family-level demographic characteristics in the second model (demography model). The third model includes parental SEP variables (parental SEP model). The last model includes the individuals' own achieved socioeconomic characteristics at age 35 (own SEP model). This last model provides information on the contribution to the total familial influence resulting not from common parental SEP but instead from similarities between siblings in their individual, adult SEP.

We then compare the MHR from the null model with MHRs from the subsequent three models. Comparing the difference in MHR after introducing demographic and socioeconomic characteristics of the parents gives us an estimate of the observed part of the family part of the long arm of childhood ($\Delta_{fam,obs}$). The change after introducing adulthood characteristics gives an estimate of the indirect effect on the family level ($\Delta_{fam,indirect}$).

All data preparation and all analyses are performed using Stata version 14.1 with the *mestreg* command and additional user written commands (Jann 2007).

Results

The baseline model contains gender, cohort, a gender-specific shape factor, and a random intercept term (shared frailty) for each family (group of siblings). Table 2 contains the estimates of individual and family-level characteristics on all-cause mortality. The variance estimate for frailty is 0.36 on the hazard scale, which translates into an MHR of 1.77. Thus, on average, between a pair of families randomly drawn from the population, the difference in mortality risk is 77 % higher in the higher-risk family than in the lower-risk family.

In the demography model, we add variables for the age of parents at the individual's birth, differences between regions in Finland, the number of siblings in the family, and an indicator for individuals with Swedish as their mother tongue. The only major difference in mortality risk is between children whose mother tongue is Swedish compared with the Finnish-speaking majority (HR = 0.61). Overall, the MHR (1.75) is not influenced notably, meaning that familial influence on mortality risk cannot be traced back to similarity of siblings with regard to language, regional parity, or parental age at birth.

The parental SEP model includes the education and occupation of the parents. A lower parental education level (less than primary school or no information, compared with past primary school) is associated with higher mortality (HR = 1.16). Further, parental occupational status is also associated with midlife mortality. Compared with

Table 2 Influences of observed and unobserved family characteristics on all-cause mortality: Hazard ratios (HR) and standard errors (SE)

	Baseline		Demography		Parental SEP		Own SEP	
	HR	SE	HR	SE	HR	SE	HR	SE
Age (Gompertz shape parameter b_w)	1.07***	0.00	1.07***	0.00	1.07***	0.00	1.07***	0.00
Male	2.94***	0.16	2.94***	0.16	2.94***	0.16	3.15***	0.17
Male × Age (Gompertz shape parameter b_m)	0.99**	0.00	0.99**	0.00	0.99**	0.00	1.00	0.00
Demography								
Native language (ref. = Finnish)								
Swedish			0.61***	0.03	0.62***	0.03	0.72***	0.04
Mother's age at birth (ref. = 14–24)								
25–35			1.00	0.03	1.00	0.03	1.01	0.03
35+			1.07	0.04	1.06	0.04	1.06	0.04
No valid info			1.23*	0.10	1.19*	0.10	1.21*	0.10
Father's age at birth (ref. = 14–24)								
25–35			0.98	0.04	1.00	0.04	0.98	0.04
35+			0.96	0.04	0.98	0.04	0.94	0.04
No valid info			1.07	0.06	1.04	0.06	1.00	0.05
Region (ref. = Western Finland)								
Eastern Finland			1.13***	0.03	1.11***	0.03	1.10***	0.03
Lapland			1.03	0.05	0.99	0.05	1.06	0.05
Uusimaa			1.15***	0.04	1.14***	0.04	1.14***	0.04
Number of siblings (ref. = 2)								
3			1.02	0.03	1.02	0.03	1.01	0.03
4			1.05	0.03	1.04	0.03	1.01	0.03
5+			1.04	0.03	0.99	0.03	0.97	0.03
Parental SEP								
Education (ref. = more than primary)								
Did not go to school/unknown					1.16**	0.06	1.00	0.05
Primary school					1.06	0.05	0.97	0.04
Occupational status (ref. = professionals)								
Workers and agriculture workers					1.16***	0.04	1.04	0.04
Farmers					1.01	0.04	0.89**	0.04
Farmer (10+ ha)					0.89*	0.05	0.83***	0.04
Employer/self-employed					1.04	0.05	0.96	0.05
Other/unknown					1.33**	0.12	1.14	0.10
Persons per heated room (ref. = less than 1)								
1–2 persons					1.01	0.05	0.95	0.05
2–3 persons					1.02	0.06	0.93	0.05
More than 3 persons					1.10	0.06	0.95	0.06
Unknown					0.99	0.11	0.89	0.10
Own SEP								
Education (ref. = highest tertiary)								
Basic or unknown							1.57***	0.09
Upper secondary (lower track)							1.33***	0.08
Upper secondary (higher track)							1.26***	0.08
Lower-degree tertiary							1.07	0.08
Income (ref. = 10th decile)								
1st decile							1.99***	0.10
2nd decile							1.72***	0.09
3rd decile							1.53***	0.08
4th decile							1.43***	0.08
5th decile							1.31***	0.07
6th decile							1.23***	0.06
7th decile							1.07	0.05

Table 2 (continued)

	Baseline		Demography		Parental SEP		Own SEP	
	HR	SE	HR	SE	HR	SE	HR	SE
8th decile							1.01	0.05
9th decile							1.00	0.05
No income							1.72***	0.12
Homeownership (ref. = homeowner)								
No owner							1.31***	0.03
Unknown							0.10***	0.01
Occupational status (ref. = higher white collar)								
Self-employed							0.94	0.05
Lower white collar							1.14**	0.05
Blue collar							1.23***	0.06
Other/unknown							1.12	0.07
Employment status (ref. = employed)								
Unemployed							1.88***	0.12
Homemakers							0.85**	0.05
Other/unknown							1.92***	0.08
Birth cohort (ref. = 1936)								
1937	0.98	0.07	0.99	0.05	0.99	0.05	0.98	0.05
1938	0.97	0.07	1.02	0.05	1.02	0.05	0.95	0.05
1939	0.88	0.06	0.95	0.05	0.96	0.05	0.88**	0.04
1940	0.86*	0.06	0.92	0.05	0.92	0.05	0.84***	0.04
1941	0.89	0.06	0.96	0.05	0.96	0.05	0.89*	0.04
1942	0.86*	0.07	0.91	0.05	0.92	0.05	0.86**	0.05
1943	0.84*	0.06	0.87**	0.05	0.88*	0.05	0.80***	0.04
1944	0.88	0.07	0.89*	0.05	0.90*	0.05	0.81***	0.04
1945	0.94	0.07	0.92	0.05	0.92	0.05	0.83***	0.04
1946	0.84*	0.06	0.88*	0.05	0.89*	0.05	0.79***	0.04
1947	0.77***	0.06	0.90	0.05	0.91	0.05	0.78***	0.04
1948	0.88	0.07	0.94	0.05	0.94	0.05	0.84**	0.05
1949	0.90	0.07	0.94	0.05	0.93	0.05	0.81***	0.05
1950	0.82*	0.07	0.90	0.06	0.90	0.06	0.77***	0.05
Family-Level Variance (θ)	0.36***	0.03	0.35***	0.03	0.33***	0.03	0.27***	0.03
MHR	1.77***	0.04	1.75***	0.04	1.73***	0.04	1.64***	0.04
Total Person-Years at Risk	2,598,805							
Individuals	94,042							
Families	32,544							
Deaths	10,948							

*$p < .05$; **$p < .01$; ***$p < .001$

professionals (higher white-collar workers), the HR for blue-collar and farm workers is 1.16; other differences are smaller and not statistically significant. Our measure of total familial influence (MHR) is minimally reduced to 1.73 after the inclusion of parental SEP variables. Substantively, these changes are very small. Taken together, the observed part ($\Delta_{fam,obs}$) is just 5.2 % of the total familial influence. We thus conclude that parental SEP has some association with mortality but does not contribute substantially to the explanation of total familial influence on midlife mortality.

The individual SEP model adds education, income, homeownership, occupational position, and employment status at age 35. All the dimensions of individuals' SEP exert an influence on mortality separately. For example, the mortality risk for individuals in

the lowest income decile is 1.99 times higher than the risk for those in the highest decile. Compared with those with higher tertiary education, individuals with only basic or unknown education have a mortality risk that is 1.57 times higher. Individuals who rent have significantly increased mortality risk compared with those who own or partially own a house at the age of 35 (HR = 1.31). Last, compared with upper white-collar workers, blue-collar workers' mortality risk is 1.23 times higher.

The socioeconomic stratification variables of individuals at age 35 explain a larger proportion of the total family influence. The MHR is 1.64. The indirect part ($\Delta_{fam,indirect}$) thus makes up an additional 11.7 % of the total familial influence.

For all-cause mortality, we can conclude that (1) the average difference in mortality risk between families is almost as large as the strongest differences that we find between social groups, and (2) only the indirect pathway ($\Delta_{fam,indirect}$) through individuals' own SEP contributes a relevant portion to the explanation of familial influences on all-cause mortality. As we propose in the theoretical section, the *unobserved arm* is of greater magnitude than the *observed long arm of childhood*.

Cause-Specific Familial Influence

In this section, we examine differences in the magnitude of sibling similarity and the proportion of similarity explained by the demography, parental, and individual SEP models between causes of death. Table S3 in Online Resource 1 lists the relative frequency of causes of death in the sample. Figure 1 shows MHR by cause of death.

The highest MHR is found for alcohol-related deaths (2.49), but the median hazards in CVD (MHR = 2.37), lung cancer (MHR = 2.19), and accidental and violent (MHR = 2.04) deaths are also markedly higher than for all-cause mortality. Other types of cancer show a similar total familial influence to all-cause mortality (MHR = 1.78).

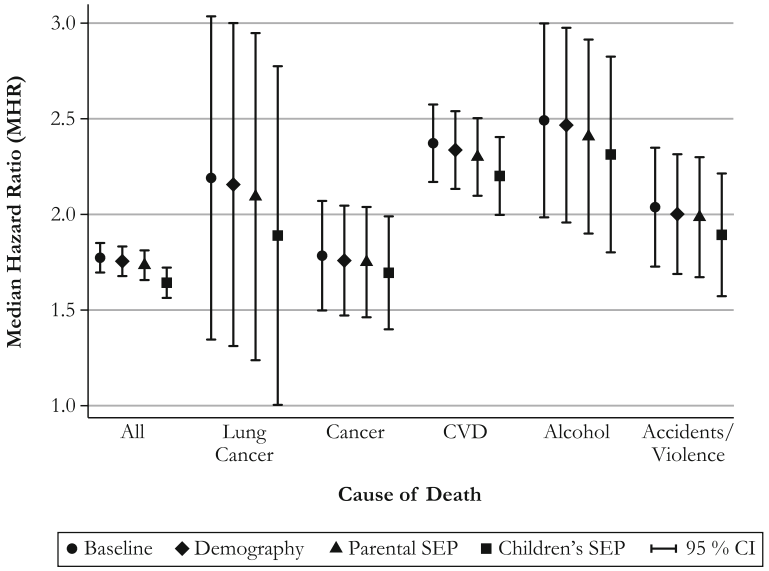

Fig. 1 Differences in median hazard ratio between models and by cause of death

Similar to the result for all-cause mortality, parental and individuals' own SEP explain only a small proportion of the familial influence on mortality. The largest fraction (21.5 %) is explained by the indirect effect ($\Delta_{fam,indirect}$) of individuals' SEP on mortality due to lung cancer, in addition to only 7 % of the observed part of total family influence ($\Delta_{fam,obs}$). The cumulative explanatory power ($\Delta_{fam,indirect} + \Delta_{fam,obs}$) for other causes of death lies between 10 % (alcohol-related) and 15.41 % (accidents and violence), which is smaller than the explicable familial influence on all-cause mortality. Despite finding clear and strong social gradients in all cause-of-death groups, we can attribute mortality differences between families to only a maximum of one-quarter of our measures of social stratification. The analyses show that the differences in the level of familial influence between causes of death are much higher than the share of familial influence that can be explained by SEP (the differences between models within each cause of death), indicating much more variation in the strength of the long arm of childhood across causes of death than between the observed and unobserved components.

We conducted several sensitivity analyses showing that our results are not sensitive to gender (analyses solely of brother-sister sibling pairings; see Figs. S8–S13 in Online Resource 1) or to alternative choices of distributional assumption about the shared frailty parameter (inverse Gaussian, Gamma distribution). The results are reported in Figs. S6 and S7 in Online Resource 1.

Discussion

We set out to uncover evidence that the long arm of childhood influences adult mortality in Finland. We showed that midlife mortality exhibits clear social gradients with respect to achieved income, education, occupation, and measures of wealth at age 35, but to a lesser degree with the socioeconomic characteristics of an individual's parents during childhood. Based on these analyses alone, we would have found little evidence for the long arm of childhood, with it acting mostly indirectly via individuals' own achieved SEP. The exception was Swedish as mother tongue, which showed a strong gradient favoring the Swedish minority, which could be only partially explained by adult SEP.

However, we proposed that in addition to the *observed long arm* of childhood (Δ_{obs}), an *unobserved counterpart* (Δ_{uno}) had even greater importance. We found substantial unobserved familial influences ($\Delta_{fam,uno}$) in all-cause and cause-specific mortality, measured as MHR and reflecting the family component of the total influence of childhood on adult mortality. On average, the mortality risk more than doubled for CVD–, alcohol-, and lung cancer–related deaths, and would be approximately 70 % higher for cancers other than lung cancer and all-cause mortality if an individual were to change to a randomly chosen higher-risk family. This finding confirms an unobserved counterpart to the arm of childhood that has substantial stratifying effects on midlife mortality. Further, we found that only approximately 20 % (up to 28 % for lung cancer) of the total familial influence on mortality could be explained by the joint effect of observed sociodemographic characteristics of parents ($\Delta_{fam,obs}$) and the indirect pathway through individuals' adult SEP ($\Delta_{fam,indirect}$), confirming our hypothesis that the unobserved counterpart of the long arm of childhood is in fact of greater importance

for midlife mortality than the visible arm. We also found that the larger part of the observed arm was the indirect pathway, which was mediated by adulthood SEP. The indirect pathway contributed much more to the explanation of the family component of the long arm than observed childhood characteristics. Nevertheless, the remaining direct pathway of family influence ($\Delta_{fam,direct}$) was more than twice the size of the mediated pathway, indicating potential for unrecognized scarring effects in childhood and thus making childhood an even more sensitive or critical period (or as yet unobserved factors of individuals in early adulthood).

The major part of the explained familial influence in all groups of causes was related to the indirect effect through individuals' own SEP, showing strong support for the pathway model. In comparison, observed parental factors were of much lesser importance. Several explanations are possible. First, the increase in economic and educational status from the parents' to the children's generation leads to a higher (observable) variation, especially in educational degrees. Because of the reduction in size of the lowest educational categories in the parent's generation, the variance in mortality risk between families can be explained to only a small degree by differences in education. Second, adulthood characteristics might be more important for health behaviors in the context of a rapidly changing economic, social, and technological situation in Finland after the 1950s. Finally, current living conditions—including economic and working conditions, but also level of education—might have a more direct relation to mortality, whereas childhood conditions and their latent effects do not (yet) show their influence in the age group under observation; this might be especially true for premature mortality.

We found considerable evidence to suggest that familial influence is strongest in accidents, violence, and alcohol-related deaths (as well as lung cancer), which reflects results from previous studies on childhood influences based on observed characteristics (Galobardes et al. 2004). The familial influence on lung cancer should be given special consideration because the overall familial influence on this cause of death was larger than for all-cause mortality, and the relative importance of the observed part of the arm—that is, the part that can be attributed to the observed sociodemographic characteristics of parents and their children—was considerably larger compared with other groups of cause of death. This finding indicates that determinants of lung cancer mortality, primarily smoking (Fenelon and Preston 2012), are especially subject to observable social influences—a result that has also been found in other studies (Geyer 2008; Kulik et al. 2013; Mackenbach et al. 2004). Note that although cancer other than lung cancer showed the smallest familial influence, we still found a considerable link between total family circumstances and these forms of cancer, suggesting that previous studies' finding of no evidence for links with observed SEP variables (Galobardes et al. 2004, 2008) might have taken too narrow or specific a view on childhood influences. In future studies, more in-depth analyses regarding familial influence on more specific groups of causes of death would be interesting, given that previous results indicate particular causes (such as stomach cancer or hemorrhagic stroke) that have especially strong links to observed childhood characteristics (Galobardes et al. 2004).

When we relate our estimates of $\Delta_{fam,uno}$ to Δ_{uno} and $\Delta_{fam,indirect}$ to $\Delta_{indirect}$, our conclusions depend on which of the three assumptions (1–3) we can defend. If we made the very unrealistic Assumption 1 that our observed variables would explain the same amount for the individual part of the long arm of childhood as for the family part,

we could generalize our statements to the total family influence on adult mortality. If our observed variables have more explanatory power on the family than the individual level (Assumption 2), our estimates for the unexplained direct family effects are conservative or lower-bound estimates. For the unexplained part of childhood effects, we think this is likely because almost all variables vary only between families, and to only a minor degree within families (with exceptions, such as age of parents at birth). Forming definitive conclusions regarding the indirect pathway is difficult, but we think that the results shown here are strong enough to conclude that for the total childhood influence, the set of observed variables captures at best only one-half of the long arm of childhood and probably less. At the very least, we can say that a substantial total childhood influence is left unexplained in a very parsimonious model, even if we cannot specify the exact proportion (Assumption 3).

Overall, our results indicate that calls for more complex models of interaction between social and biological factors in the life course (Galea et al. 2010) are not merely aiming at minor improvements of existing parsimonious models but could potentially have strong additional predictive power when considering childhood circumstances and adult mortality.

We do not imply that the differences between socioeconomic groups are unimportant. On the contrary, our models shows significant differences between them. Nevertheless, other characteristics of the family of origin—ones that we have been unable to observe directly—are clearly extremely powerful in determining midlife mortality. Depending on cause of death, these unobserved factors contribute between three and five times more than observed factors to the differences in midlife mortality between families.

Comparing our study results with previous research shows similarities in the sense that most childhood socioeconomic influences can be explained by the pathway model. In the original study that gave the "long arm of childhood" research its name, Hayward and Gorman (2004) found that the effects of socioeconomic and demographic variables of the parents are mediated through individuals' own achieved SEP and health behaviors. Also, Case and Paxson (2010) found that the differences in adult health between different levels of childhood socioeconomic position are completely explained by attained adult social position. Several other studies investigating adult health instead of adult mortality have found similar results of (almost) complete mediation (Link et al. 2017; Pakpahan et al. 2017; Turner et al. 2016; Zajacova et al. 2015), although some evidence suggests that neither the effects of childhood socioeconomic conditions nor the effects of early-life health conditions on adult health can fully be accounted for by adulthood characteristics (O'Rand and Hamil-Luker 2005). A systematic review of the literature on childhood SEP and its association with adult mortality also corroborates the view that a large part of the effects of childhood SEP is mediated via adult characteristics (Galobardes et al. 2004, 2008). The key addition to previous studies is that we showed—while using a data set that yields similar results on observed socioeconomic variables as previous research—that the total childhood influence may exceed what can typically be observed several times over. We therefore argue that the general direction of our results would also hold in other contexts. Convincing evidence from many developed countries suggests that adult SEP has strong predictive power for mortality (Elo 2009; Mackenbach et al. 2008). Therefore, our results showing an association with observed parental characteristics should be replicable

across countries, time, and cohorts, even if the exact strength of the associations may vary. It is thus reasonable to expect sizable differences in mortality between families (sibling similarity) in other contexts as well.

Our results indicate that the factors that are not observed are important in determining adult mortality. In future studies, their relative contribution to the total childhood influence can be assessed in one of the following ways. Based on our results, we know that the combined effects of all factors not measured are approximately three to four times larger than observed adult SEP variables, which were the strongest predictors of mortality. Therefore, a rough guide for future studies would be to assess whether a new set of explanatory variables is as predictive or more predictive. For example, the estimates in our final model for all-cause mortality, with a family-level variance of 0.27, mean that a potential new predictor that is standardized to have a variance of 1 on the family level needs to show a HR of 1.68[1] to completely explain differences between families in mortality hazard. Although we should not take these estimates as exact guidelines because they are conditional on data and modeling differences, they can be seen as an order of magnitude estimate for the future introduction of new variables or for modeling complex interactions and assessing their contribution to the study of childhood influences on adult mortality.

Limitations

The advantages of using register data also come with certain disadvantages. First, when comparing the influence of the observed and unobserved arm of childhood, we run the systematic risk of underestimating the contribution of the visible arm due to poor measurement of our observed socioeconomic and demographic characteristics. For example, we do not have information on household income when the individuals were young, although our results show substantial differences in mortality risk between income groups in adult age. This might lead to an underestimation of some of the effects of parental SEP, especially because parental education is also measured in only three broad categories. Consequently, measurement error might be a driver of the low estimate of the influence of the visible arm of childhood and parental social characteristics on mortality—a finding observed in previous studies as well (Hayward and Gorman 2004; Kröger et al. 2017). However, we can also give a complementary explanation for the relatively minor contribution of parental SEP to the visible arm of childhood. The compression of parental characteristics into only three educational categories and much less variation in occupational class positions in the parental generation than in the children's generation are signs that stratification across these dimensions is much smaller than in the children's generation. This does not mean that there were no educational or occupational inequalities in the prewar generation in Finland, but rather that the advantaged groups (well-educated, upper white-collar workers) made up such a small part of the population that these dimensions can make only a modest contribution to the explanation of differences in mortality risk in midlife. In our assessment, both measurement error and lower stratification very likely play some role in the relatively minor contribution of parental characteristics to total familial influence. Another limitation of the measurement of parental characteristics stems from

[1] See Online Resource 1 for the calculation.

the fact that we have only one point of observation in childhood, at an age dictated by the timing of the census and not by theoretical choice. Although the parents' educational degrees can be assumed to be quite stable throughout childhood, economic conditions and occupations can change, and we cannot observe these changes or their implications for later-life mortality. However, socioeconomic position in adulthood is arguably well measured from the register data and still cannot account for the majority of variation in mortality risk between families.

Second, our analyses are limited to midlife and early old-age mortality. Early-life mortality and old-age mortality might show different patterns regarding total familial influence and sibling similarity. Predicting their magnitude relative to midlife mortality is difficult. On the one hand, genetic research shows that inheritance of mortality grows with age (Gentilini et al. 2013; Murabito et al. 2012). On the other hand, intracohort differentiation during the life course, as well as individual paths and influences from outside the family, might lead to higher heterogeneity between families and within families at older ages (Dannefer 1997; O'Rand and Henretta 1999). Therefore, an interesting undertaking for future research would be to compare total familial influence on mortality in different stages of the life course and for different cohorts.

Third, we are able to analyze causes of death by only very broad groups because of the limited number of deaths per family per cause. In terms of statistical models, we have to rely on the parametric assumptions of a Gompertz distribution of the hazard and normal distribution of the shared frailty parameter. The former yields a very similar prediction of the hazard as a nonparametric approach, and the latter is insensitive to specifying gamma or inverse Gaussian distributions for the frailty parameter.

A fourth limitation derives from our inability to determine the exact degree of relatedness of all individuals in the register data. Although we are at least able to identify a common mother or father for each individual in the sibling data, it is not always clear whether the individuals share both parents. We therefore cannot differentiate clearly between full, half-, and step-siblings. This misclassification is likely to lead to the underestimation of shared frailty. Additionally, orphans and institutionalized children are not included in the analyses; however, in the cohorts under investigation, they make up only 4.6 % of the population. Furthermore, using a sibling approach, by definition, excludes all only-children. Although we could not find substantive differences in terms of SEP between siblings and singletons, the long arm of childhood might manifest itself differently for only-children because of their only-child status per se.

Finally, the unobserved contribution of shared family effects is conditional on data and research question. We could not cover many other important factors that we included in our theoretical model. Other studies have shown that childhood health status is especially predictive of adult health, even beyond achieved adulthood characteristics (Case and Paxson 2010; Haas 2008; Link et al. 2017; O'Rand and Hamil-Luker 2005; Pakpahan et al. 2017; Zajacova et al. 2015), making childhood health status an important stressor that could explain the unobserved family component. We also had only very limited information on family structure and relationships inside the family, which might be important for the development of health and mortality risk in later life (Campbell and Lee 2009). In addition, we have no genetic information. If such information cannot be directly collected, one way to indirectly assess genetic endowment reducing the mortality risk is to calculate family excess longevity (if family members can be linked), which has been shown to explain a substantial part of the

correlations in mortality hazards between same-sex siblings (Smith et al. 2009). Finally, we did not consider complex interactions of any of the dimensions that influence mortality.

Conclusion

The midlife mortality hazard of Finnish cohorts born between 1936 and 1950 shows considerable variation between families, which to a significant extent is due to unobserved factors. Thus, to get a more comprehensive picture of the influence of childhood and family circumstances on mortality, the observed part of the long arm of childhood needs to be supplemented with the unobserved counterpart of the same arm. The degree of familial influence varies between causes of death, with alcohol-related causes showing the strongest influence from the family and all-cause mortality and cancer (except lung cancer) showing the lowest total familial influence. All types of mortality show strong social gradients, mostly with respect to the individuals' own SEP, but parental social background also plays a stratifying role. In combination with demographic characteristics, these observed social characteristics account for approximately one-fifth of the total variation of all-cause mortality between families and up to 28 % of lung cancer mortality differences between families. Because a large proportion of the total familial effect is left unexplained, other family-related factors that are shared within families are immensely important in determining the mortality risk in midlife and early old age, highlighting the potential for complex models of social biological interactions in a life course framework.

Acknowledgments We thank Diana Schacht and three anonymous reviewers for valuable feedback on earlier versions of this article. We especially thank Lea Kröger for feedback on and inspiration for the study. All remaining mistakes are, of course, solely the authors' responsibility.

References

Aalen, O. O., Valberg, M., Grotmol, T., & Tretli, S. (2015). Understanding variation in disease risk: The elusive concept of frailty. *International Journal of Epidemiology, 44,* 1408–1421.

Aaltonen, S., Kaprio, J., Vuoksimaa, E., Huppertz, C., Kujala, U. M., & Silventoinen, K. (2016). Genetic architecture of motives for leisure-time physical activity: A twin study. *Scandinavian Journal of Medicine & Science in Sports, 27,* 1431–1441.

Agahi, N., Shaw, B. A., & Fors, S. (2014). Social and economic conditions in childhood and the progression of functional health problems from midlife into old age. *Journal of Epidemiology and Community Health, 68,* 734–740.

Andersson, M. A. (2016). Chronic disease at midlife. *Journal of Health and Social Behavior, 57,* 373–389.

Becker, G. S., & Tomes, N. (1976). Child endowments and the quantity and quality of children. *Journal of Political Economy, 84*(4, Part 2), S143–S162.

Beekman, M., Blanché, H., Perola, M., Hervonen, A., Bezrukov, V., Sikora, E., . . . GEHA Consortium. (2013). Genome-wide linkage analysis for human longevity: Genetics of Healthy Aging Study. *Aging Cell, 12,* 184–193.

Bengtsson, T., & Lindström, M. (2000). Childhood misery and disease in later life: The effects on mortality in old age of hazards experienced in early life, southern Sweden, 1760–1894. *Population Studies, 54,* 263–277.

Bengtsson, T., & Lindström, M. (2003). Airborne infectious diseases during infancy and mortality in later life in southern Sweden, 1766–1894. *International Journal of Epidemiology, 32,* 286–294.

Bengtsson, T., & Broström, G. (2009). Do conditions in early life affect old-age mortality directly and indirectly? Evidence from 19th-century rural Sweden. *Social Science & Medicine, 68,* 1583–1590.

Ben-Shlomo, Y., & Kuh, D. (2002). A life course approach to chronic disease epidemiology: Conceptual models, empirical challenges and interdisciplinary perspectives. *International Journal of Epidemiology, 31,* 285–293.

Björklund, A., & Jäntti, M. (2012). How important is family background for labor-economic outcomes? *Labour Economics, 19,* 465–474.

Blane, D., Kelly-Irving, M., d'Errico, A., Bartley, M., & Montgomery, S. (2013). Social-biological transitions: How does the social become biological? *Longitudinal and Life Course Studies, 4,* 136–146.

Blau, P. M., & Duncan, O. D. (1967). *The American occupational structure*. New York, NY: Wiley.

Blomgren, J., Martikainen, P., Mäkelä, P., & Valkonen, T. (2004). The effects of regional characteristics on alcohol-related mortality—A register-based multilevel analysis of 1.1 million men. *Social Science & Medicine, 58,* 2523–2535.

Boardman, J. D. (2009). State-level moderation of genetic tendencies to smoke. *American Journal of Public Health, 99,* 480–486.

Boardman, J. D., Domingue, B. W., Blalock, C. L., Haberstick, B. C., Harris, K. M., & McQueen, M. B. (2014). Is the gene-environment interaction paradigm relevant to genome-wide studies? The case of education and body mass index. *Demography, 51,* 119–139.

Bouchard, C. (2008). Gene–environment interactions in the etiology of obesity: Defining the fundamentals. *Obesity, 16*(S3), S5–S10.

Campbell, C. D., & Lee, J. Z. (2009). Long-term mortality consequences of childhood family context in Liaoning, China, 1749–1909. *Social Science & Medicine, 68,* 1641–1648.

Case, A., & Paxson, C. (2010). Causes and consequences of early-life health. *Demography, 47,* S65–S85.

Dannefer, D. (1997). Aging as intracohort differentiation: Accentuation, the Matthew effect, and the life course. *Sociological Forum, 2,* 211–236.

Dannefer, D. (2003). Cumulative advantage/disadvantage and the life course: Cross-fertilizing age and social science theory. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 58,* S327–S337.

Davey Smith, G., Hart, C., Blane, D., Gillis, C., & Hawthorne, V. (1997). Lifetime socioeconomic position and mortality: Prospective observational study. *BMJ, 314,* 547. https://doi.org/10.1136/bmj.314.7080.547

DiPrete, T. A., & Eirich, G. M. (2006). Cumulative advantage as a mechanism for inequality: A review of theoretical and empirical developments. *Annual Review of Sociology, 32,* 271–297.

Dowd, J. B., Simanek, A. M., & Aiello, A. E. (2009). Socio-economic status, cortisol and allostatic load: A review of the literature. *International Journal of Epidemiology, 38,* 1297–1309.

Duncan, G. J., Boisjoly, J., & Harris, K. M. (2001). Sibling, peer, neighbor, and schoolmate correlations as indicators of the importance of context for adolescent development. *Demography, 38,* 437–447.

Elo, I. T. (2009). Social class differentials in health and mortality: Patterns and explanations in comparative perspective. *Annual Review of Sociology, 35,* 553–572.

Elo, I. T., Martikainen, P., & Myrskylä, M. (2014). Socioeconomic status across the life course and all-cause and cause-specific mortality in Finland. *Social Science & Medicine, 119,* 198–206.

Falkstedt, D., Lundberg, I., & Hemmingsson, T. (2011). Childhood socio-economic position and risk of coronary heart disease in middle age: A study of 49,321 male conscripts. *European Journal of Public Health, 21,* 713–718.

Fenelon, A., & Preston, S. H. (2012). Estimating smoking-attributable mortality in the United States. *Demography, 49,* 797–818.

Freese, J. (2008). Genetics and the social science explanation of individual outcomes. *American Journal of Sociology, 114*(Suppl.), S1–S35.

Freese, J., & Shostak, S. (2009). Genetics and social inquiry. *Annual Review of Sociology, 35,* 107–128.

Gagnon, A., & Mazan, R. (2009). Does exposure to infectious diseases in infancy affect old-age mortality? Evidence from a pre-industrial population. *Social Science & Medicine, 68,* 1609–1616.

Galea, S., Riddle, M., & Kaplan, G. A. (2010). Causal thinking and complex system approaches in epidemiology. *International Journal of Epidemiology, 39,* 97–106.

Galobardes, B., Lynch, J. W., & Davey Smith, G. (2004). Childhood socioeconomic circumstances and cause-specific mortality in adulthood: Systematic review and interpretation. *Epidemiologic Reviews, 26,* 7–21.

Galobardes, B., Lynch, J. W., & Davey Smith, G. (2008). Is the association between childhood socioeconomic circumstances and cause-specific mortality established? Update of a systematic review. *Journal of Epidemiology and Community Health, 62,* 387–390.

Gavrilov, L. A., & Gavrilova, N. S. (2001). Human longevity and parental age at conception. In J.-M. Robine, T. B. L. Kirkwood, & M. Allard (Eds.), *Sex and longevity: Sexuality, gender, reproduction, parenthood* (pp. 7–31). Berlin, Germany: Springer-Verlag.

Gentilini, D., Mari, D., Castaldi, D., Remondini, D., Ogliari, G., Ostan, R., . . . Vitale, G. (2013). Role of epigenetics in human aging and longevity: Genome-wide DNA methylation profile in centenarians and centenarians' offspring. *Age, 35,* 1961–1973.

Geyer, S. (2008). Social inequalities in the incidence and case fatality of cancers of the lung, the stomach, the bowels, and the breast. *Cancer Causes & Control, 19,* 965–974.

Gobbens, R. J. J., Luijkx, K. G., Wijnen-Sponselee, M. T., & Schols, J. M. G. A. (2010). Towards an integral conceptual model of frailty. *Journal of Nutrition, Health & Aging, 14,* 175–181.

Haas, S. (2008). Trajectories of functional health: The "long arm" of childhood health and socioeconomic factors. *Social Science & Medicine, 66,* 849–861.

Hart, C. L., & Davey Smith, G. (2003). Relation between number of siblings and adult mortality and stroke risk: 25 year follow up of men in the Collaborative study. *Journal of Epidemiology and Community Health, 57,* 385–391.

Hayward, M. D., & Gorman, B. K. (2004). The long arm of childhood: The influence of early-life social conditions on men's mortality. *Demography, 41,* 87–107.

Herttua, K., Mäkelä, P., & Martikainen, P. (2008). Changes in alcohol-related mortality and its socioeconomic differences after a large reduction in alcohol prices: A natural experiment based on register data. *American Journal of Epidemiology, 168,* 1110–1118.

Hoffmann, R. (2008). *Socioeconomic differences in old age mortality* (Vol. 25). New York, NY: Springer.

Hougaard, P. (1995). Frailty models for survival data. *Lifetime Data Analysis, 1,* 255–273.

House, J. S., Lepkowski, J. M., Kinney, A. M., Mero, R. P., Kessler, R. C., & Herzog, A. R. (1994). The social stratification of aging and health. *Journal of Health and Social Behavior, 35,* 213–234.

Hubbard, R. E., Andrew, M. K., & Rockwood, K. (2009). Effect of parental age at birth on the accumulation of deficits, frailty and survival in older adults. *Age and Ageing, 38,* 380–385.

Jann, B. (2007). Making regression tables simplified. *Stata Journal, 7,* 227–244.

Johnson, R. C., Schoeni, R. F., & Rogowski, J. A. (2012). Health disparities in mid-to-late life: The role of earlier life family and neighborhood socioeconomic conditions. *Social Science & Medicine, 74,* 625–636.

Kelly-Irving, M., Lepage, B., Dedieu, D., Lacey, R., Cable, N., Bartley, M., . . . Delpierre, C. (2013). Childhood adversity as a risk for cancer: Findings from the 1958 British birth cohort study. *BMC Public Health, 13,* 767. https://doi.org/10.1186/1471-2458-13-767

Kröger, H., Fritzell, J., & Hoffmann, R. (2016). The association of levels of and decline in grip strength in old age with trajectories of life course occupational position. *PLoS One, 11*(5), e0155954. https://doi.org/10.1371/journal.pone.0155954

Kröger, H., Kroh, M., Kroll, L. E., & Lampert, T. (2017). *Einkommensunterschiede in der Mortalität in Deutschland—Ein empirischer Erklärungsversuch* [Differences in income in mortality in Germany—An empirical explanation attempt]. *Zeitschrift für Soziologie, 46,* 124–146.

Kuh, D., Shah, I., Richards, M., Mishra, G., Wadsworth, M., & Hardy, R. (2009). Do childhood cognitive ability or smoking behaviour explain the influence of lifetime socio-economic conditions on premature adult mortality in a British post war birth cohort? *Social Science & Medicine, 68,* 1565–1573.

Kulik, M. C., Hoffmann, R., Judge, K., Looman, C., Menvielle, G., Kulhánová, I., . . . Mackenbach, J. P. (2013). Smoking and the potential for reduction of inequalities in mortality in Europe. *European Journal of Epidemiology, 28*, 959–971.

Lahti, R. A., & Penttilä, A. (2001). The validity of death certificates: Routine validation of death certification and its effects on mortality statistics. *Forensic Science International, 115,* 15–32.

Link, B. G., & Phelan, J. (1995). Social conditions as fundamental causes of disease. *Journal of Health and Social Behavior, 35,* 80–94.

Link, B. G., Susser, E. S., Factor-Litvak, P., March, D., Kezios, K. L., Lovasi, G. S., . . . Cohn, B. A. (2017). Disparities in self-rated health across generations and through the life course. *Social Science & Medicine, 174,* 17–25.

Lynch, W. J., Kaplan, A. G., & Salonen, T. J. (1997). Why do poor people behave poorly? Variation in adult health behaviours and psychosocial characteristics by stages of the socioeconomic lifecourse. *Social Science & Medicine, 44,* 809–819.

Mackenbach, J. P., Huisman, M., Andersen, O., Bopp, M., Borgan, J.-K., Borrell, C., . . . Kunst, A. E. (2004). Inequalities in lung cancer mortality by the educational level in 10 European populations. *European Journal of Cancer, 40,* 126–135.

Mackenbach, J. P., Stirbu, I., Roskam, A.-J. R., Schaap, M. M., Menvielle, G., Leinsalu, M., & Kunst, A. E. (2008). Socioeconomic inequalities in health in 22 European countries. *New England Journal of Medicine, 358,* 2468–2481.

McGue, M., Vaupel, J. W., Holm, N., & Harvald, B. (1993). Longevity is moderately heritable in a sample of Danish twins born 1870–1880. *Journal of Gerontology, 48,* B237–B244.

Merlo, J. (2011). Contextual influences on the individual life course: Building a research framework for social epidemiology. *Psychosocial Intervention, 20,* 109–118.

Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., . . . Larsen, K. (2006). A brief conceptual tutorial of multilevel analysis in social epidemiology: Using measures of clustering in multilevel logistic regression to investigate contextual phenomena. *Journal of Epidemiology and Community Health, 60,* 290–297.

Mishra, G., Nitsch, D., Black, S., De Stavola, B., Kuh, D., & Hardy, R. (2009). A structured approach to modelling the effects of binary exposure variables over the life course. *International Journal of Epidemiology, 38,* 528–537.

Mishra, G. D., Chiesa, F., Goodman, A., De Stavola, B., & Koupil, I. (2013). Socio-economic position over the life course and all-cause, and circulatory diseases mortality at age 50–87 years: Results from a Swedish birth cohort. *European Journal of Epidemiology, 28,* 139–147.

Murabito, J. M., Yuan, R., & Lunetta, K. L. (2012). The search for longevity and healthy aging genes: Insights from epidemiological studies and samples of long-lived individuals. *Journals of Gerontology, Series A: Biological Sciences and Medical Sciences, 67A,* 470–479.

Myrskylä, M., Elo, I. T., Kohler, I. V., & Martikainen, P. (2014). The association between advanced maternal and paternal ages and increased adult mortality is explained by early parental loss. *Social Science & Medicine, 119,* 215–223.

O'Rand, A. M., & Hamil-Luker, J. (2005). Processes of cumulative adversity: Childhood disadvantage and increased risk of heart attack across the life course. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 60*(Special Issue 2), 117–124.

O'Rand, A. M., & Henretta, J. C. (1999). *Age and inequality: Diverse pathways through later life*. Boulder, CO: Westview Press.

Pakpahan, E., Hoffmann, R., & Kröger, H. (2017). The long arm of childhood circumstances on health in old age: Evidence from SHARELIFE. *Advances in Life Course Research, 31,* 1–10.

Palloni, A. (2006). Reproducing inequalities: Luck, wallets, and the enduring effects of childhood health. *Demography, 43,* 587–615.

Pearlin, L. I. (1989). The sociological study of stress. *Journal of Health and Social Behavior, 30,* 241–256.

Pescosolido, B. A., & Levy, J. A. (2002). The role of social networks in health, illness, disease and healing: The accepting present, the forgotten past, and the dangerous potential for a complacent future. In J. A. Levy & B. A. Pescosolido (Eds.), *Advances in medical sociology: Social networks and health* (Vol. 8, pp. 3–25). Bingley, UK: Emerald Group.

Pescosolido, B. A., Perry, B. L., Long, J. S., Martin, J. K., Nurnberger, J. I., Jr., & Hesselbrock, V. (2008). Under the influence of genetics: How transdisciplinarity leads us to rethink social pathways to illness. *American Journal of Sociology, 114*(Suppl. 1), S171–S201.

Pudrovska, T., & Anikputa, B. (2014). Early-life socioeconomic status and mortality in later life: An integration of four life-course mechanisms. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 69,* 451–460.

Qi, L., & Cho, Y. A. (2008). Gene-environment interaction and obesity. *Nutrition Reviews, 66,* 684–694.

Saarela, J., & Finnäs, F. (2009). Geographic ancestry and cause-specific mortality in a national population. *Population Research and Policy Review, 28,* 169–194.

Shanahan, M. J., & Hofer, S. M. (2005). Social context in gene-environment interactions: Retrospect and prospect. *Journals of Gerontology, Series B: Psychological Sciences and Social Sciences, 60*(Special Issue 1), 65–76.

Smith, K. R., Mineau, G. P., Garibotti, G., & Kerber, R. (2009). Effects of childhood and middle-adulthood family conditions on later-life mortality: Evidence from the Utah Population Database, 1850–2002. *Social Science & Medicine, 68,* 1649–1658.

Solon, G., Corcoran, M., Gordon, R., & Laren, D. (1991). A longitudinal analysis of sibling correlations in economic status. *Journal of Human Resources, 26,* 509–534.

Szanton, S. L., Gill, J. M., & Allen, J. K. (2005). Allostatic load: A mechanism of socioeconomic health disparities? *Biological Research for Nursing, 7,* 7–15.

Tarkiainen, L., Martikainen, P., & Laaksonen, M. (2016). The contribution of education, social class and economic activity to the income-mortality association in alcohol-related and other mortality in Finland in 1988–2012. *Addiction, 111,* 456–464.

Turner, R. J., Thomas, C. S., & Brown, T. H. (2016). Childhood adversity and adult health: Evaluating intervening mechanisms. *Social Science & Medicine, 156,* 114–124.

Turner, R. J., Turner, J. B., & Hale, W. B. (2014). Social relationships and social support. In R. J. Johnson, R. J. Turner, & B. G. Link (Eds.), *Sociology of mental health selected topics from forty years 1970s–2010s* (pp. 1–20). New York, NY: Springer.

Turrell, G., Lynch, J. W., Leite, C., Raghunathan, T., & Kaplan, G. A. (2007). Socioeconomic disadvantage in childhood and across the life course and all-cause mortality and physical function in adulthood: Evidence from the Alameda County Study. *Journal of Epidemiology and Community Health, 61,* 723–730.

van den Berg, G. J., & Drepper, B. (2015). Inference for shared-frailty survival models with left-truncated data. *Econometric Reviews, 35,* 1075–1098.

Vaupel, J. W. (1988). Inherited frailty and longevity. *Demography, 25,* 277–287.

Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., Yashin, A. I., Holm, N. V, . . . Curtsinger, J. W. (1998). Biodemographic trajectories of longevity. Science*,* 280*,* 855–860.

Vaupel, J. W., Manton, K. G., & Stallard, E. (1979). The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography, 16,* 439–454.

Wienke, A. (2010). *Frailty models in survival analysis* (1st ed.). Boca Raton, FL: Taylor and Francis Group.

Zajacova, A., Walsemann, K. M., & Dowd, J. B. (2015). The long arm of adolescent health among men and women: Does attained status explain its association with mid-adulthood health? *Population Research and Policy Review, 34,* 19–48.