The influence of observed and unobserved family background on mortality – evidence from Finnish register data on siblings and their parents

Abstract

In this study, we address the question as to how much the socioeconomic position of siblings and their parents contributes to the explanation of differences in mortality risk between families. We provide three estimates of the family's overall influence on mortality that are based on the family-level variance in a survival analytic regression model, using siblings nested in 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 age 35 up to age 72. To explain familial influence on mortality, we use demographic background factors, the socioeconomic position of the parents, and the siblings' own socioeconomic position at age 35 as predictors of all-cause and cause-specific mortality. Results show that familial influence is higher for CVD, accidents and alcohol and lung cancer-related deaths than for all-cause mortality, and than for mortality related to other forms of cancer than lung cancer. Jointly, demographic and socioeconomic factors, including region, number of siblings, native language, education and occupation of parents, income, occupation, tenancy status, and education of the siblings explain between 10 and 25% of the total familial influence on mortality. The siblings' socioeconomic variables make the largest contribution in explaining familial influence for all causes of mortality. However, a large portion of the influence of the family on mortality is not explained by individual and parental socioeconomic position, highlighting the need to investigate familial influence on mortality in a comprehensive framework including demographic, social, behavioral, and genetic information.

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 to estimating the social influence of the family is to take observed socioeconomic characteristics like the parents' education, occupation, or income to predict the mortality of their children. 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 within this tradition shows that people from disadvantaged social backgrounds in childhood have higher mortality and lower life expectancy, and that much of these early life conditions are mediated by achieved social status (Palloni, 2006; Palloni et al., 2009; Pudrovska and Anikputa, 2014).

This approach, however, mostly disregards all unobserved familial influences that are not captured by the observed measures of SEP. The influence of the family can extend far beyond these observable socioeconomic factors. Transmission of behavior, values, preferences, and genetic endowments, as well as family-related neighborhood and school characteristics can all be part of the "long arm of the family" on children's health, health behavior, and finally on their mortality risk later in life. Furthermore, as we understand familial influence explicitly as more than just parental influence, the siblings' similar social environment is also an important aspect of the influence of family. As depicted in figure 1, the net effect of the combination of these factors can be regarded as an inherited *frailty* (Vaupel, 1988) that is shared among the offspring of parents. To assess the influence of this shared 'frailty for mortality', we propose to estimate the variance of the frailty parameter between families, based on a multi-level survival model that uses siblings nested in families. This shared frailty of siblings can be regarded as the sum of all influences of the family on the siblings' mortality hazard. This approach of estimating total familial influence through sibling similarity is widespread in the study of the transmission of SEP (Björklund and Jäntti, 2012; Duncan et al., 2001; Solon et al., 1991) and has also found use in research on health inequalities (Johnson et al., 2012; Merlo, 2011). We will use this sibling approach to estimate the total influence of the family on mortality, and to see how much of this influence is attributable to the SEP of the parents and their children. This approach stands in contrast to other recent studies that simply want to control for shared characteristics in siblings, but do not take this shared frailty as an *explanandum* (e.g. Elo et al., 2014; Næss et al., 2012; Tarkiainen et al., 2015). To the best of our knowledge, we

present the first systematic attempt to explain the influence of unobserved family factors shared by siblings on mortality by structural demographic and socioeconomic determinants.

An important aspect when analyzing the influence of family on mortality is genetic endowment, or the heritability of longevity. Despite the fact that the influence of genes might seem obvious, studies that have actually quantified the degree of heritability of longevity (based on twins studies) come to the conclusion that between 15%-30% of variation may be due to heritability, and another 25% to factors that are fixed by the age of 30 (Beekman et al., 2013; McGue et al., 1993; Vaupel, 1998; Vaupel et al., 1998). The contribution of heritability tends to increase the older the individuals under observation are (Brooks-Wilson, 2013), and also might be more important if people become ill (Hoffmann, 2011). Furthermore, in genome-wide association studies, only two gene variants have consistently been identified as predictors of longevity (APOE and FOXO3A), and in their case mostly for old age mortality and the mortality of centenarians (Gentilini et al., 2013; Murabito et al., 2012). Consequently, we cannot assume that a majority of (midlife) mortality is determined by genetically heritable factors.

In addition to the genetic factors, siblings also share their (social) environments in childhood. For example, there may be behavioral similarities between siblings acquired throughout childhood and adolescence. These can include dietary factors (Baker et al., 2000; Kaati et al., 2007) or risk-taking behavior, but also substance abuse or violence (Bantle and Haisken-DeNew, 2002; Wickrama et al., 1999), or metabolic risk factors like obesity, diabetes, high LDL cholesterol, or hypertension (Khoury et al., 1983; Kloch-Badelek et al., 2014). This intergenerational transmission of health behavior can be direct, e.g. through malnutrition during pregnancy or early childhood (Barker et al., 1993), or via achieved adult characteristics such as education, knowledge, and lifestyle (Lawlor et al., 2006; Mirowsky and Ross, 1998). Siblings also partially share their childhood social environments. Previous evidence has shown that, for example, parental low education or crowded housing have effects on mortality in middle age (Elo et al., 2014).

Our concept of familial influence is a compound measure of all these aspects, including both shared genetic endowment and the social environment. For the purpose at hand it is neither possible nor necessary to disentangle each of the factors that contribute to similarity. Rather, we want to compare the degree to which familial influence on siblings' adult mortality is related to social characteristics of parents and siblings. We will further stratify the analyses by broad groups of causes of death. Following the argument that similarity in mortality hazard is partially determined by biological heritability, but also partially by social characteristics acquired through the family, we see two competing hypotheses

regarding differences between causes of death: On the one hand, causes of death that have a stronger behavioral component could be the ones indicative of the greatest similarity between siblings. Consequently, lung cancer, cardio-vascular diseases, and accidental and violent causes, as well as alcohol-related causes, are expected to show a higher level of sibling similarity than other forms of cancer and all-cause mortality. On the other hand, causes of death that arguably have a higher genetic determination, such as cancers other than lung cancer, could show a higher level of determination by family of origin if genetic factors dominate the shared frailty characteristics (Mackenbach et al., 2015; Plug et al., 2012).

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 to 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. This way of identifying siblings is more in line with a social notion of siblings, meaning being raised by at least one common parent in the same family, instead of a biological definition of siblings. The vast majority (> 93%) of siblings in our sample can be linked to both the mother and father.

The identification of siblings in the early census of 1950 avoids common problems of estimating longevity based on later reports of the survival of siblings and other relatives, which is a method sometimes used in survey research (Gakidou and King, 2006). This design also means that our results refer to midlife and early old age mortality (deaths in the age range 35-72). All surviving individuals are censored at the end of year 2007 when they are aged between 57 and 72. As there is no mortality information before 1970, the analyses exclude all deaths in early life until age 35, and refer only to those who survived to this age. This restriction reduces the age range upon which we can draw inference, but avoids the problem of variation in left truncation that can create biased inference on the estimated parameters (Berg and Drepper, 2015; Hoffmann, 2008). Not only those who died, but also those who emigrated before 1970 are absent in the analysis. Therefore, 15,065 of those siblings included in the 1950 census sample are not contributing to the mortality analysis. This is largely due to extensive emigration to Sweden in the 1960s. Prior studies on the same data set have shown that this leads to a minor overrepresentation of women, individuals born before 1945, those from low SES backgrounds,

and mother-only families in the sample (Elo et al., 2014). This bias is so small that it is unlikely to impact on our results.

The sample results in 94,042 individuals nested in 32,544 families, making up 2,598,805 person-years of analysis time. We divide mortality into all-cause mortality, and mortality due to a) lung cancer b) other forms of cancer c) cardiovascular diseases d) alcohol-related deaths, and e) accidents and violent deaths. Other groups of mortality do not provide sufficient number of deaths in the data set to analyze them separately. Alcohol-related causes include, among other things, 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 middle age (male) mortality in Finland (Elo et al., 2014; Herttua et al., 2008; Tarkiainen et al., 2016). Accidents and violence include, among other causes, suicides, traffic accidents, poisoning (excluding alcohol poisoning), and homicide. The coding of causes of death in the Finnish death register, especially in broader categories, has been shown to be reliable (Lahti and Penttilä, 2001).

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

The second group contains information on parental SEP from the 1950 census, and includes the highest level of education of the parents (no schooling, primary, or past primary education), as well as the occupational class of the father, categorized in analogy to the Erikson-Goldthorpe-Portocarero scheme (EGP). If paternal information was not available, the occupational status of the mother was used. Further, housing conditions – as persons per heated room – are used as an indicator of the parents' socioeconomic resources. The third group of variables measures the adult SEP of the siblings. We use the highest educational degree of each sibling. The degrees are categorized into 1) basic, 2) lower secondary level, 3) upper secondary level, 4) lowest level tertiary, and 5) lower-degree level tertiary and higher-degree tertiary. Occupational status is measured based on occupational coding comparable to the EGP class scheme. The categories used are 1) 'Business owners and the self-employed, 2) upper white-collar workers, 3) lower white-collar workers, and 4) blue-collar workers. Tenancy status distinguishes between siblings who are renting and those who own their home, or shares of it. After-tax income is categorized into deciles representing the relative income position in the year of the census

closest to the year when the child turned 35. The variables are measured using the census in 1970, 1975, 1980 and 1985, depending on the birth cohort.

Statistical Approach

We use three estimates to quantify the influence of shared family characteristics on mortality. These are sibling similarity, the median hazard ratio, and the family-age ratio. Each of these quantities is based in a different form of the variance of the shared frailty parameter from a multilevel survival model (Θ). The frailty parameter is shared between siblings, making families the higher level (level 2) units. It should be noted that frailty is used here in the statistical sense of survival analysis, which takes variation between different levels into account (Hougaard, 1995; Vaupel et al., 1979; Wienke, 2010). It is not a measurement of frailty as a clinical indicator for health, as often used in ageing research (Aalen et al., 2015; Gobbens et al., 2010; Romero-Ortuno and Kenny, 2012).

First, we will describe the estimation strategy for sibling similarity, and then follow with an explanation of the median hazard ratio and the family-age ratio.

Although we will interpret the coefficients of our models in terms of hazard ratios, it helps to look at the model by first applying the accelerated failure time (AFT) metric (Lambert et al., 2004), to understand how sibling similarity is estimated.

The formal regression equation for the frailty model on the AFT metric that we will use is defined as:

$$\log(y_{fst}) = a + b_w * t_{fs} + c * male_{fs} + b_m * t_{fs} * male_{fs} + \boldsymbol{X_{fs}\gamma} + \log(Z_f) + \log(e_{fst}) \quad (1)$$

$$\log(Z_f|X_{fs}) \sim N(0, \sigma_Z); \text{VAR}(\log(Z_f)) = \Theta$$

$$\log(e_{fs}) \sim G\left(0, \frac{\pi}{\sqrt{6}}\right);$$

The index f stands for family, s for sibling, and t for the age in years above 35. . Z_f is the frailty shared by siblings in a family. It is assumed that the random components $\log(Z_f)\, and\, \log(e_{fs})$ are uncorrelated and $\log(Z_f)$ is assumed to follow a normal distribution with a mean of zero which is a formulation more closely related to the methodological tradition of multilevel modeling than of shared frailty models (Pankratz et al. 2005). In our sensitivity analyses we also specify gamma and inverse Gaussian distributions for $\log(Z_f)$ - which is more common in the literature on shared frailty – and discuss the (small) differences in comparison to the normality assumption The shape parameter b,

representing the increase in mortality hazard per year, is allowed to differ between men ($b_w + b_m$) and women (b_w). **X** is a vector of observed predictors and $\boldsymbol{\gamma}$ the respective vector of coefficients. $\mathrm{Log}(y_{fs})$ is the log. survival time of each sibling nested in a family. This is the accelerated failure time metric which we use for illustrational purposes to demonstrate the estimation of the sibling similarity. The error term e_{fs}follows the standard exponential distribution with a mean and variance of 1. The logarithm of a standard exponentially distributed variable is the standard extreme value distribution or standard Gumbel (G) distribution, which has a variance of $\frac{\pi^2}{6}$ (Allison 2010, p. 78). The family specific error-term is estimated as the variance of the intercept α, which is the log-baseline survival time. The metric of the family error component is therefore on the log-survival time scale and to put it into relation to this log-survival time scale it has to be compared to the variance of the standard Gumbel distribution.

Following the method suggested by Goldstein et al. (2002), we calculate sibling similarity as an approximation of the intra-class correlation (ICC) in a linear multilevel model from the estimated variance on the family level and the assumed variance of the individual error component:

$$\hat{\rho}_f = \frac{\widehat{\Theta}}{\widehat{\Theta}+\sigma_e^2} = \frac{\widehat{\Theta}}{\widehat{\Theta}+\frac{\pi^2}{6}} \quad (2)$$

A confidence interval can be derived for this statistic based on a standard error, which is derived using the delta method (Stata's nlcom command, see also (1992)).

The equivalence between our estimate of sibling similarity and the calculation of an ICC for a continuous outcome is that the estimate is bounded by 0 and 1, and that higher values indicate higher similarity between siblings relative to the differences between families. However, as the variance on the sibling level is fixed, not freely estimated, we must be cautious in the interpretation of the numerical value of the sibling similarity. For low and average amounts of variation between families, the value can be interpreted as an approximation of the variance in survival time (or the hazard of mortality) on the family level compared to the total variation, which consists of family level plus individual (sibling) level. This assumes a latent variable approach in which either we think of the individual level as survival times that vary between individuals and have a known distribution, or we treat the hazard as an individual characteristic that is unobserved, but which varies between individuals. We do not observe this individual hazard, but only its realization in death or censoring at a certain age.

This sibling similarity can be interpreted as the total effect of all influences that siblings share and that make them more alike than any two individuals randomly chosen from the population. Importantly, it also gives an overall estimate of the influence of the family in all its aspects.

Note that the estimation of sibling similarity does not require the AFT metric, as proportional hazard and AFT are equivalent models on a different metric. Therefore, sibling similarity is not affected by our choice of metric. In the proportional hazards metric, the model is defined as:

$$h_{fs}(t) = Z_f * \exp\left(-\left(a + b_w * t_{fs} + c * male_{fs} + b_m * t_{fs} * male_{fs} + \boldsymbol{X_{fs}\gamma}\right)\right) h_0(t) \quad (3)$$

Our second measure of familial influence on siblings' mortality hazard is the median hazard ratio (MHR), a complementary approach to an ICC-based estimate of level-2 influence in non-linear models (Merlo et al., 2006). In contrast to the ICC, it is not a measure of similarity within groups, but a measure of *dissimilarity* between groups. It can be interpreted as the average (median) difference in mortality hazard between families, or the average increase in mortality that would occur if a random individual from a random family were put in another family. It is estimated based on the variance term on the family level:

$$MHR = \exp\left(\sqrt{2 * \theta} * 0.6745\right) \quad (4)$$

One advantage of the MHR is the equivalence of its scale to the parameters of the predictors in the model. Consequently, direct comparisons are possible. A fictional example would be a statement like: The average difference between families in mortality hazard (MHR 1.6) is about the same size as the difference between high and low educated individuals (HR 1.65).

In addition to the MHR and sibling similarity, a third statistic that can signify the importance of familial influence is calculated by relating the estimate of the familial variance to the shape parameter of the Gompertz model. The shape parameter tells us how much the hazard increases per year of age. Consequently, we can divide the estimate of the square root of the familial variance by the shape parameter and make this statement: Being in a family with one SD more of frailty influences the mortality hazard in the same way as X years of ageing. We call this the family-age ratio (FAR):

$$\widehat{FAR} = \frac{\sqrt{\hat{\theta}}}{\hat{b}} \quad (5)$$

As for sibling similarity and MHR, a confidence interval can be derived for this ratio based on a standard error which is derived using the delta method. While sibling similarity (as an intra-class correlation of variance partition component) and the MHR have been used in previous research, we propose the FAR as a new approach of estimating the relevance of family for mortality.

Our baseline model includes only the variables birth cohort and gender of the siblings. After estimating the baseline model, demographic characteristics of the individuals and families are introduced to the model (demography model). The third model includes parental SEP variables (parental SEP model). The last model includes achieved socioeconomic characteristics of the siblings at age 35 (siblings' SEP model). This model provides information on the contribution of similarity not due to common parental SEP, but due to similarity between siblings in their individual SEP.

We then compare the three measures of familial influence on mortality from the null model to the subsequent three models. Our approach has two main advantages. First, it allows us to estimate the total influence of the family on siblings' adult mortality. Second, we get an estimate of how much of this total familial influence can be attributed to SEP of the parents and to siblings' SEP in adulthood, and how much of the familial influence is left unexplained and can attributed to the unobserved family background.

Using this framework, the study presents an additional opportunity to use household level data for life course analysis of socioeconomic and demographic determinants of mortality. 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

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

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 estimate of sibling similarity of 0.18. We can say that almost a fifth of the variation in survival time is estimated as resulting from differences between families, while 82% is the result of differences between siblings. Compared to the average increase in hazard per year, being in a family that is one standard deviation below the average in survival time is approximately the same as 8.81 years of ageing (FAR averaged across gender). The family's impact on the hazard, expressed as the

FAR, is illustrated in figure 2. It plots the predictions of the hazard of mortality from the shared frailty model for an average family and a family that is one standard deviation above (high risk family) and below the average family (low risk family), respectively. The family-age ratio is visually represented as the thick dash-dot lines. They show that high risk men reach a hazard of 1% at age 51.88, while men from an average family first reach this hazard at age 61.64. For women, a hazard of 0.5% is reached at age 55.90 in the high risk families, but only at age 64.71 in the low risk group. The median hazard ratio in the baseline model is 1.77. This means that, on average, the difference in mortality risk is 77% higher in the higher risk family than in the lower risk family, taking a pair of families drawn randomly from the population.

The demography model adds variables on differences in parents' age at the birth of their child between regions in Finland, the number of siblings in the family, and an indicator for individuals with Swedish as their mother tongue. Children whose mother tongue is Swedish have a substantially reduced mortality risk of 0.60. There are also notable mortality differences between the regions in Finland, with Eastern Finland and the province of Uusimaa having slightly increased mortality compared to Western Finland (HR respectively: 1.13 and 1.15). There is no strong association between parents' age at birth (maternal or paternal) and mortality. The differences in mortality between number of siblings in the family is also small and non-significant. Overall, sibling similarity is not influenced notably (0.17), and neither is the family-age-ratio 8.66 or the MHR (1.75), meaning that similarity in mortality risk between siblings cannot be traced back to similarity of siblings in language, region, or parental age at birth.

The parental SEP model includes education and occupation, measured in 1950, when the siblings were aged between 0 and 15. Lower parental education ("less than primary school or no information" compared to "past primary school") is associated with higher mortality (HR 1.16). We can further see that parental occupational position is also associated with midlife mortality. Compared to professionals (higher white-collar), the HR for blue-collar and farm workers is 1.16; other differences are smaller and not statistically significant. The number of people living in the household per heated room is an indicator of lack of resources in childhood, but is only slightly and non-significantly associated with mortality. Regarding sibling similarity and family-age ratio on mortality, we can see that the inclusion of parental SEP variables changes little. Sibling similarity is still 0.17, the family–age ratio decreases slightly to 8.47, and MHR is reduced to 1.73. Substantively, these changes are negligible and we can conclude that parental SEP has some association with siblings' mortality, but cannot make a relevant contribution to the explanation of sibling similarity and family impact on mortality.

The siblings' SEP model adds variables on the SEP of the children at age 35. This includes education, income, occupational position and home ownership, and their employment status. All of the dimensions of siblings' SEP exert an influence on mortality separately. For example, individuals in the lowest income decile have a mortality risk 1.96 times greater than those in the highest decile. Compared to those with higher tertiary education, those with only basic or unknown education have a mortality risk which is 1.67 times higher. Siblings who rent have a significantly increased mortality risk compared to those who own (shares) of a house at the age of 35 (HR 1.31). Lastly, compared to upper white-collar workers, blue-collar workers have a 1.23 times higher risk of mortality.

The socioeconomic stratification variables of the individuals at age 35 explain a larger portion of the sibling similarity, in addition to the small explanatory contribution of the demography and the parental SEP model. The last model reduces the conditional sibling similarity to 0.14. The family-age ratio is reduced to 7.54 years of ageing. The effect of the median hazard ratio is 1.64. We can now see that this average difference between families is larger than any difference between occupational groups (max HR 1.23) and between home owners and non-owners. It is about the same size as the difference between the lowest educational group and the highest. It is somewhat smaller than the difference between the highest and lowest income decile, and also somewhat smaller than the difference between unemployed and employed individuals. For all-cause mortality, we can conclude that, first, the average difference in mortality risk between families is almost as large as the strongest differences we find between social groups and, second, that only an individual's own SEP contributes a relevant effect to the explanation of familial influences on all-cause mortality. In total, only about 20% of familial influence could be explained jointly by demographic, parental, and siblings' socioeconomic factors.

Cause-specific sibling similarity

In this section, we look at differences in the magnitude of sibling similarity, and the fraction of similarity explained by the demography, parental, and siblings' SEP models for different causes of death. Table 3 lists the relative frequency of causes of death in the sample. Figure 3 shows sibling similarity by cause of death, figure 4 shows the FAR by cause of death, and figure 5 shows the MHR by cause of death.

The highest similarity is estimated for alcohol related deaths (0.36), but similarity in CVD (0.33), and accidental and violent (0.25) deaths are also markedly higher than for all-cause mortality. Lung cancer (0.29) also shows higher sibling similarity than all-cause mortality. Other types of cancer show a total familial influence comparable to that of all-cause mortality (0.19). The MHR follows the same pattern as sibling similarity; while the average difference in mortality risk between two families is about 70% for

all-cause and cancer mortality, it is about 100% in accidents and violence, 120%-130% in lung cancer and CVD, and about 150% in alcohol related mortality. It is interesting that the FAR comparison does not follow the same patterm. Lung cancer estimates are about the same size as for all-cause mortality (approximately 9 years). This means that, if we take a greater increase in mortality risk with increasing age into account, the relative influence of family is the same in all-cause as in lung cancer-specific mortality. Therefore, how we measure familial influence makes a difference for the analysis of familial influence with regard to different causes of death. Furthermore, a reliable estimate for accidents and violence could not be calculated, because the shape parameter (log. increase in hazard with age) approached zero, and the values became unreasonably high, which is a common drawback of a ratio. Therefore, the result established here – that the causes of death that are more strongly linked to behavior show a higher total familial influence than all-cause mortality and deaths related to other forms of cancer – holds only for an absolute measure of familial influence not relative to the increase in hazard with age.

Similar to the result for all-cause mortality discussed above, parental and sibling SEP can only explain a small fraction of familial influence on mortality. The largest part is explained by siblings' SEP for mortality due to lung cancer. As smoking shows a strong social gradient, individual SEP can explain an estimated 27% of the differences in lung cancer mortality between families. The cumulative explanatory power for other causes of death lie in the range of 10% (alcohol related) to 15.41% (accidents and violence), and are smaller than the familial influence explaining all-cause mortality. Despite the fact that we can find clear and strong social gradients in all causes of death, we can attribute mortality differences between families only to a maximum of about a quarter of our measures of social stratification. All three measures show that the differences in the level of familial influence between causes of death are much higher than the share of familial influence which can be explained by SEP (the differences between models within each cause of death).

Sensitivity Analysis

In our analysis, the estimation of the variance term for the shared frailty parameters is the key element. In the results reported above, we assumed a Gaussian distribution of the frailty parameters. However, the calculation of sibling similarity, the MHR, and the FAR might be sensitive to this assumption. Therefore, we reran the analyses for all-cause mortality using gamma and inverse Gaussian distribution, to see whether the results would change substantially due to the specification of the distribution of the

frailty parameter. The gamma model for the frailty distribution has been used by numerous researchers (e.g. Manton and Stallard, 1981; Vaupel and Yashin, 1985). The other common distribution, inverse Gaussian, was introduced as a frailty model by Hougaard (1984). Figures 6-8 in the appendix show the measures of familial influence for all-cause mortality according to the distributional assumption for shared frailty parameters. We can see that a Gaussian distribution results in slightly lower overall estimates of familial influence, but the change in familial influence for the different models is proportional across specifications. We therefore make a slightly conservative estimate regarding the size of influence of shared family characteristics; however, our conclusions regarding the explanation of familial influence are not affected by the choice of distribution for the shared frailty parameters.

A second aspect that might influence the conclusions from our analyses is gender specificity in familial influence on mortality hazard. In our analysis, we combined brothers and sisters. However, gender-specific parenting, as well as gender differences in mortality (Hamil-Luker and O'rand, 2007), could lead to different results if we estimated familial influence separately for brothers and sisters. To test the gender specificity of our results, we repeated our analyses for men and women separately, dropping all families with less than two children of the same gender, meaning with only one son or daughter respectively. We also repeated the analyses for cause-specific mortality if the brother and sister subsample still contained more than 500 deaths to ensure enough between and within-family variation. The results of brother and sister similarity are reported in figures 9-14 in the appendix. We can see that the differences in estimates of familial influence on the pooled sample are minor, and that the explanatory contribution of the four models are very similar in size. The strongest deviation from the main results that we observed is a somewhat stronger difference between CVD and alcohol-related familial influence for brothers, compared to the full sample. We therefore conclude that familial influence on mortality hazard is not gender specific for our sample.

Discussion

Based on Finnish register data for cohorts born between 1936 and 1950 of individuals with at least one sibling, our study indicates three major findings.

First, we show that midlife mortality exhibits clear social gradients with respect to income, education, occupation, and measures of wealth. These results are in line with previous studies that took a

complementary approach to analyses based on siblings, which found that the educational gradient in adult cardio-vascular diseases (Madsen et al., 2014) and cause specific mortality (Elo et al., 2014; Næss et al., 2012) were only partially explained by shared sibling characteristics;the income gradient was only modestly affected at best in a study also based on Finnish register data (Tarkiainen et al., 2015).

Second, we find substantial unobserved familial influence in all-cause and cause-specific mortality, measured as sibling similarity, median hazard ratio (MHR), or family-age-ratio (FAR). Individuals from a low risk family (1 standard deviation below the mean hazard) have a risk of mortality that is similar to an individual from an average family who is between 9 and 15 years younger, depending on the specific cause of death. Expressed as MHR, we can say that – on average – the mortality risk more than doubles for CVD, alcohol, and lung cancer-related deaths, and is about 77% to 150% (depending on cause of death) higher for all-cause mortality, if an individual were to change to a random higher risk family. From these estimates, we can see that – for mortality –family is of considerable importance, and consequently siblings are much more alike (sibling similarities ranging between 0.18 and 0.35) than individuals chosen at random. We can say that between 18 percent and 15 percent of the variation in log survival times is between families, the rest of the variation is between siblings.

Third, we showed that only about 20% (up to 27% for lung cancer) of familial influence can be explained by the joint effect of demography, parental SEP, and siblings' SEP on mortality. This does not mean that the differences between socioeconomic groups are of minor importance. On the contrary, the models show that there are significant differences between them. Rather, it tells us that other characteristics of the family of origin that we have been unable to observe directly are extremely powerful in determining midlife mortality. Depending on cause of death, these unobserved factors contribute between 4 and 5 times more to the differences in midlife mortality between families than observed factors.

From the final model we are able to see that this average mortality difference between families is almost as large as – and in some cases even larger than – the strongest differences between social groups. This argument may also be expressed conversely: Despite the fact that observed sociodemographic characteristics explain only between 10-27% of sibling similarities, the strongest differences between social groups are larger than the average difference between families (MHR) with regard to mortality risk. This interpretation highlights the fact the social gradient in mortality is indeed sizable, even when compared to familial differences which include unobserved (genetic) characteristics. Familial influence on lung cancer should be mentioned specifically, because for this cause of death, the overall sibling similarity is larger than for all-cause mortality, and the fraction that can be attributed to

the social stratification of parents and their children is also larger. This indicates that determinants of lung cancer mortality, mostly smoking (Fenelon and Preston, 2012), are especially subject to social influences, a result that has been found in other studies as well (e.g. Geyer, 2008; Kulik et al., 2013; Mackenbach et al., 2004). Shared family background therefore plays an important role in determining mortality beyond the already important differences based on individual social characteristics.

Considering the transferability of our results to other contexts, we can make the following observations. There is convincing evidence from many developed countries that adult SEP has strong predictive power for mortality. Therefore, the part of our results showing associations with observed parental and sibling characteristics is replicable across countries, time, and cohorts, even if the specific strength of the associations varies. It is also reasonable to expect sizable differences in mortality between families (sibling similarity) in other contexts as well. While each of the two factors – association of observed and unobserved shared sibling characteristics – will vary across context, it is reasonable to predict that the order of magnitude in the relation between overall familial influence (as measured by sibling similarity) and overall SEP (as measured by parental SEP and siblings' SEP) will not be very different. This means that also, in other contexts, we would expect the non-observed factors to play a more important role than the observed sociodemographic characteristics of parental and individual SEP.

One possible way of systematically extending the study of sibling similarity in mortality is to take a multigenerational perspective and include grandparents and cousins into the analyses (Mare, 2011). The limiting factor for such approaches is of course data availability, even when using longitudinally linked census data. However, sometimes data is available and a study based on settlers of Cape Colony shows that, even in the third generation, mortality is linked within families (Piraino et al., 2014).

Limitations

The benefits of using register data also come with certain disadvantages. For example, we do not have information on the income of the households when the siblings where young, although our results show that there are substantial differences in mortality risk between income groups in adulthood. This might underestimate some of the effects of parental SEP, especially because parental education is also only measured in three broad categories, and therefore a poorer proxy of income than for the younger generation. In addition, information regarding parenting styles, supervision, and encouragement in the family home, as well as direct information on shared genetic traits. is entirely lacking. Thus, we cannot empirically separate genetic factors from other factors acting in childhood for reasons of data

limitations, as has been done in studies comparing monozygotic with dizygotic twins (Tan et al., 2013). We can, however, interpret the effect of unobserved family factors, which is shown as sibling similarity after controlling for observable demographic, parental, and individual variables; the joint effect of genetic and unobserved social family characteristics. Even on a very fundamental conceptual level, these two elements may not be two separate factors, because of the complex interplay between (epi)genetic factors and environment (Capri et al., 2014). On a more practical level, our study was unable to analytically separate these two groups of factors, so we only made a modest attempt to analyze the differences between causes of death, and their dependence on genes versus behavior. Our finding that cancer - other than lung cancer - exhibits less sibling similarity than the other causes of death led to the conclusion that the behavioral aspects in our unobserved family traits might be more important than the genetic aspects.

Further, our analyses are limited to midlife mortality. Early-life mortality and old-age mortality might show different patterns regarding total familial influence and sibling similarity. It is hard to predict their magnitude relative to midlife mortality. 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, and 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). It would therefore be interesting for future research to compare total familial influence on mortality in different stages of the life course and for different cohorts.

Another limitation derives from our inability to determine the exact degree of relatedness of all siblings in the register data. While for each individual in the data at least a common mother or a common father is identified, it is not always clear whether the siblings share both parents. It is therefore not possible to differentiate clearly between full, half, and step siblings. . This misclassification is likely to lead to under-estimation of shared frailty. Additionally, orphans and institutionalized children are also not included in the analyses, although, in the cohorts under investigation, they make up about 4.6% of the population.

Conclusion

The mortality hazard of Finnish cohorts born between 1936 and 1950 shows considerable variation between families, which is measured as a shared frailty of siblings and represents the total familial influence on mortality. The degree of familial influence varies between causes of death, with alcohol-related causes showing most similarity between siblings, and all-cause mortality, as well as cancer

(except lung cancer), the lowest total familial influence. All types of mortality show strong social gradients, mostly with respect to siblings' SEP, but parental social background also plays a stratifying role. In combination with demographic characteristics these social characteristics account for about a fifth of the variation of all-cause mortality between families, and up to 27% of lung cancer mortality differences between families. Because a large proportion of sibling similarity is left unexplained, other family-related factors that are shared by siblings are immensely important in determining their mortality risk in mid and early old-age..

References

Aalen, O.O., Valberg, M., Grotmol, T., and Tretli, S. (2015). Understanding variation in disease risk: the elusive concept of frailty. Int. J. Epidemiol. *44*, 1408–1421.

Allison, P.D. (2010). Survival Analysis Using SAS: A Practical Guide, Second Edition (SAS Institute).

Baker, C.W., Whisman, M.A., and Brownell, K.D. (2000). Studying intergenerational transmission of eating attitudes and behaviors: Methodological and conceptual questions. Health Psychol. *19*, 376–381.

Bantle, C., and Haisken-DeNew, J.P. (2002). Smoke Signals: The Intergenerational Transmission of Smoking Behavior. SSRN ELibrary.

Barker, D.J., Gluckman, P.D., Godfrey, K.M., Harding, J.E., Owens, J.A., and Robinson, J.S. (1993). Fetal nutrition and cardiovascular disease in adult life. Lancet Lond. Engl. *341*, 938–941.

Beekman, M., Blanché, H., Perola, M., Hervonen, A., Bezrukov, V., Sikora, E., Flachsbart, F., Christiansen, L., De Craen, A.J.M., Kirkwood, T.B.L., et al. (2013). Genome-wide linkage analysis for human longevity: Genetics of Healthy Aging Study. Aging Cell *12*, 184–193.

Berg, G.J. van den, and Drepper, B. (2015). Inference for Shared-Frailty Survival Models with Left-Truncated Data. Econom. Rev. *in print*, 1–24.

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

Blomgren, J., Martikainen, P., Mäkelä, P., and Valkonen, T. (2004). The effects of regional characteristics on alcohol-related mortality—a register-based multilevel analysis of 1.1 million men. Soc. Sci. Med. *58*, 2523–2535.

Brooks-Wilson, A.R. (2013). Genetics of healthy aging and longevity. Hum. Genet. *132*, 1323–1338.

Capri, M., Santoro, A., Garagnani, P., Giulia Bacalini, M., Pirazzini, C., Olivieri, F., Procopio, A., Salvioli, S., and Franceschi, C. (2014). Genes of Human Longevity: An Endless Quest? Curr. Vasc. Pharmacol. *12*, 707–717.

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

Duncan, G.J., Boisjoly, J., and 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., Martikainen, P., and Myrskylä, M. (2014). Socioeconomic status across the life course and all-cause and cause-specific mortality in Finland. Soc. Sci. Med. *119*, 198–206.

Fenelon, A., and Preston, S.H. (2012). Estimating Smoking-Attributable Mortality in the United States. Demography *49*, 797–818.

Gakidou, E., and King, G. (2006). Death by survey: Estimating adult mortality without selection bias from sibling survival data. Demography *43*, 569–585.

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

Gavrilov, L.A., and Gavrilova, N.S. (2001). Human Longevity and Parental Age at Conception. In Sex and Longevity: Sexuality, Gender, Reproduction, Parenthood, J.-M. Robine, T.B.L. Kirkwood, and M. Allard, eds. (Springer Berlin Heidelberg), pp. 7–31.

Gentilini, D., Mari, D., Castaldi, D., Remondini, D., Ogliari, G., Ostan, R., Bucci, L., Sirchia, S.M., Tabano, S., Cavagnini, F., et al. (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., and Schols, J.M.G.A. (2010). Towards an integral conceptual model of frailty. J. Nutr. Health Aging *14*, 175–181.

Goldstein, H., Browne, W., and Rasbash, J. (2002). Partitioning Variation in Multilevel Models. Underst. Stat. *1*, 223–231.

Hamil-Luker, J., and O'rand, A.M. (2007). Gender differences in the link between childhood socioeconomic conditions and heart attack risk in adulthood. Demography *44*, 137–158.

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

Hayward, M.D., and 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., and 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. Am. J. Epidemiol. *168*, 1110–1118.

Hoffmann, R. (2008). Socioeconomic differences in old age mortality (Springer Science & Business Media).

Hoffmann, R. (2011). Illness, not age, is the leveler of social mortality differences in old age. J. Gerontol. B. Psychol. Sci. Soc. Sci. *66*, 374–379.

Hougaard, P. (1984). Life table methods for heterogeneous populations: Distributions describing the heterogeneity. Biometrika *71*, 75–83.

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

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

Jann, B. (2007). Making regression tables simplified. Stata J. *7*, 227.

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

Kaati, G., Bygren, L.O., Pembrey, M., and Sjöström, M. (2007). Transgenerational response to nutrition, early life circumstances and longevity. Eur. J. Hum. Genet. EJHG *15*, 784–790.

Khoury, P., Morrison, J.A., Laskarzewski, P.M., and Glueck, C.J. (1983). Parent-offspring and sibling body mass index associations during and after sharing of common household environments: The princeton school district family study. Metabolism *32*, 82–89.

Kloch-Badelek, M., Knez, J., Tikhonoff, V., Thijs, L., Sakiewicz, W., Ryabikov, A., Stolarz-Skrzypek, K., Jin, Y., Malyutina, S., Casiglia, E., et al. (2014). Heritability and other determinants of left ventricular diastolic function in the family-based population study: J. Hypertens. *32*, 1854–1861.

Kulik, M.C., Hoffmann, R., Judge, K., Looman, C., Menvielle, G., Kulhánová, I., Toch, M., Ostergren, O., Martikainen, P., Borrell, C., et al. (2013). Smoking and the potential for reduction of inequalities in mortality in Europe. Eur. J. Epidemiol. *28*, 959–971.

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

Lambert, P., Collett, D., Kimber, A., and Johnson, R. (2004). Parametric accelerated failure time models with random effects and an application to kidney transplant survival. Stat. Med. *23*, 3177–3192.

Lawlor, D.A., Sterne, J.A.C., Tynelius, P., Smith, G.D., and Rasmussen, F. (2006). Association of Childhood Socioeconomic Position with Cause-specific Mortality in a Prospective Record Linkage Study of 1,839,384 Individuals. Am. J. Epidemiol. *164*, 907–915.

Mackenbach, J.P., Huisman, M., Andersen, O., Bopp, M., Borgan, J.-K., Borrell, C., Costa, G., Deboosere, P., Donkin, A., Gadeyne, S., et al. (2004). Inequalities in lung cancer mortality by the educational level in 10 European populations. Eur. J. Cancer *40*, 126–135.

Mackenbach, J.P., Kulhánová, I., Bopp, M., Deboosere, P., Eikemo, T.A., Hoffmann, R., Kulik, M.C., Leinsalu, M., Martikainen, P., Menvielle, G., et al. (2015). Variations in the relation between education and cause-specific mortality in 19 European populations: a test of the “fundamental causes” theory of social inequalities in health. Soc. Sci. Med. 1982 *127*, 51–62.

Madsen, M., Andersen, P.K., Gerster, M., Andersen, A.-M.N., Christensen, K., and Osler, M. (2014). Are the educational differences in incidence of cardiovascular disease explained by underlying familial factors? A twin study. Soc. Sci. Med. 1982 *118*, 182–190.

Manton, K.G., and Stallard, E. (1981). Methods for evaluating the heterogeneity of aging processes in human populations using vital statistics data: explaining the black/white mortality crossover by a model of mortality selection. Hum. Biol. *53*, 47–67.

Mare, R.D. (2011). A Multigenerational View of Inequality. Demography *48*, 1–23.

McGue, M., Vaupel, J.W., Holm, N., and Harvald, B. (1993). Longevity Is Moderately Heritable in a Sample of Danish Twins Born 1870–1880. J. Gerontol. *48*, B237–B244.

Merlo, J. (2011). Contextual Influences on the Individual Life Course: Building a Research Framework for Social Epidemiology. Psychosoc. Interv. *20*, 109–118.

Merlo, J., Chaix, B., Ohlsson, H., Beckman, A., Johnell, K., Hjerpe, P., Råstam, L., and 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. J. Epidemiol. Community Health *60*, 290–297.

Mirowsky, J., and Ross, C.E. (1998). Education, Personal Control, Lifestyle and Health A Human Capital Hypothesis. Res. Aging *20*, 415–449.

Murabito, J.M., Yuan, R., and Lunetta, K.L. (2012). The Search for Longevity and Healthy Aging Genes: Insights From Epidemiological Studies and Samples of Long-Lived Individuals. J. Gerontol. A. Biol. Sci. Med. Sci. *67A*, 470–479.

Myrskylä, M., Elo, I.T., Kohler, I.V., and Martikainen, P. (2014). The association between advanced maternal and paternal ages and increased adult mortality is explained by early parental loss. Soc. Sci. Med. 1982 *119*, 215–223.

Næss, Ø., Hoff, D.A., Lawlor, D., and Mortensen, L.H. (2012). Education and adult cause-specific mortality—examining the impact of family factors shared by 871 367 Norwegian siblings. Int. J. Epidemiol. dys143.

Oehlert, G.W. (1992). A Note on the Delta Method. Am. Stat. *46*, 27–29.

O'Rand, A., and Henretta, J.C. (1999). Age And Inequality: Diverse Pathways Through Later Life (Boulder, Colo: Westview Press).

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

Palloni, A., Milesi, C., White, R.G., and Turner, A. (2009). Early childhood health, reproduction of economic inequalities and the persistence of health and mortality differentials. Soc. Sci. Med. *68*, 1574–1582.

Pankratz, V.S., de Andrade, M., and Therneau, T.M. (2005). Random-effects Cox proportional hazards model: General variance components methods for time-to-event data. Genet. Epidemiol. *28*, 97–109.

Piraino, P., Muller, S., Cilliers, J., and Fourie, J. (2014). The transmission of longevity across generations: The case of the settler Cape Colony. Res. Soc. Stratif. Mobil. *35*, 105–119.

Plug, I., Hoffmann, R., Artnik, B., Bopp, M., Borrell, C., Costa, G., Deboosere, P., Esnaola, S., Kalediene, R., Leinsalu, M., et al. (2012). Socioeconomic inequalities in mortality from conditions amenable to medical interventions: do they reflect inequalities in access or quality of health care? BMC Public Health *12*, 346.

Pudrovska, T., and Anikputa, B. (2014). Early-Life Socioeconomic Status and Mortality in Later Life: An Integration of Four Life-Course Mechanisms. J. Gerontol. B. Psychol. Sci. Soc. Sci. gbt122.

Romero-Ortuno, R., and Kenny, R.A. (2012). The frailty index in Europeans: association with age and mortality. Age Ageing afs051.

Saarela, J., and Finnäs, F. (2009). Geographic ancestry and cause-specific mortality in a national population. Popul. Res. Policy Rev. *28*, 169–194.

Solon, G., Corcoran, M., Gordon, R., and Laren, D. (1991). A Longitudinal Analysis of Sibling Correlations in Economic Status. J. Hum. Resour. *26*, 509–534.

Tan, Q., Christiansen, L., Thomassen, M., Kruse, T.A., and Christensen, K. (2013). Twins for epigenetic studies of human aging and development. Ageing Res. Rev. *12*, 182–187.

Tarkiainen, L., Martikainen, P., Laaksonen, M., and Aaltonen, M. (2015). Childhood family background and mortality differences by income in adulthood: fixed-effects analysis of Finnish siblings. Eur. J. Public Health *25*, 305–310.

Tarkiainen, L., Martikainen, P., and 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.

Turrell, G., Lynch, J.W., Leite, C., Raghunathan, T., and 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. J. Epidemiol. Community Health *61*, 723–730.

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

Vaupel, J.W. (1998). Demographic analysis of aging and longevity. Am. Econ. Rev. *88*, 242–247.

Vaupel, J.W., and Yashin, A.I. (1985). The Deviant Dynamics of Death in Heterogeneous Populations. Sociol. Methodol. *15*, 179.

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

Vaupel, J.W., Carey, J.R., Christensen, K., Johnson, T.E., Yashin, A.I., Holm, N.V., Iachine, I.A., Kannisto, V., Khazaeli, A.A., Liedo, P., et al. (1998). Biodemographic trajectories of longevity. Science *280*, 855–860.

Wickrama, K.A.S., Conger, R.D., Wallace, L.E., and Elder, G.H. (1999). The Intergenerational Transmission of Health-Risk Behaviors: Adolescent Lifestyles and Gender Moderating Effects. J. Health Soc. Behav. *40*, 258–272.

Wienke, A. (2010). Frailty Models in Survival Analysis (Chapman and Hall/CRC).

Figure 1 - Conceptual relationship between family influence, shared frailty and mortality

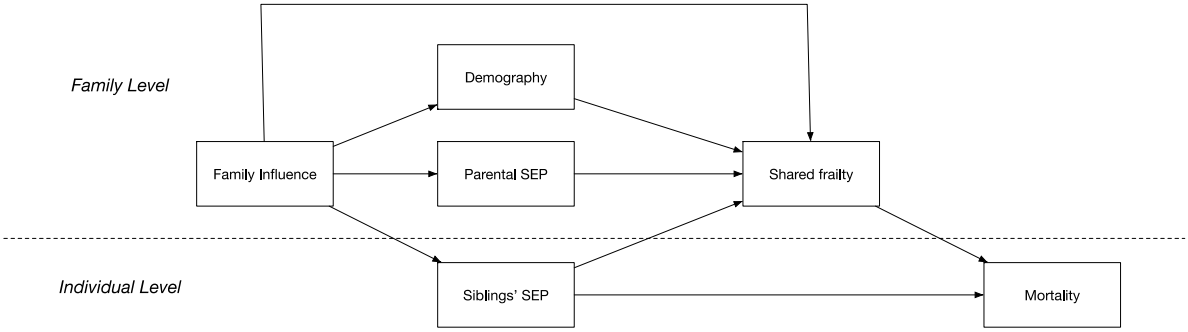


Figure 2 – Family-age ratio for men and women in the baseline model

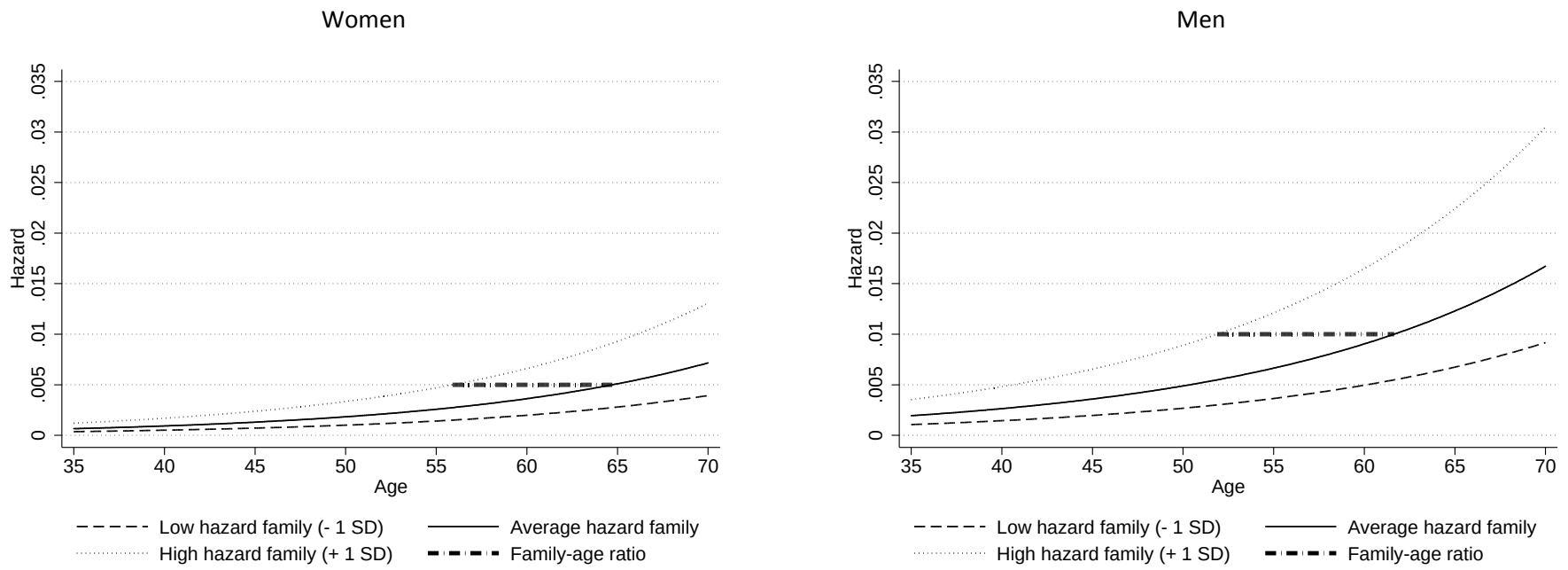


Figure 3 – Differences in sibling similarity between models and by cause of death

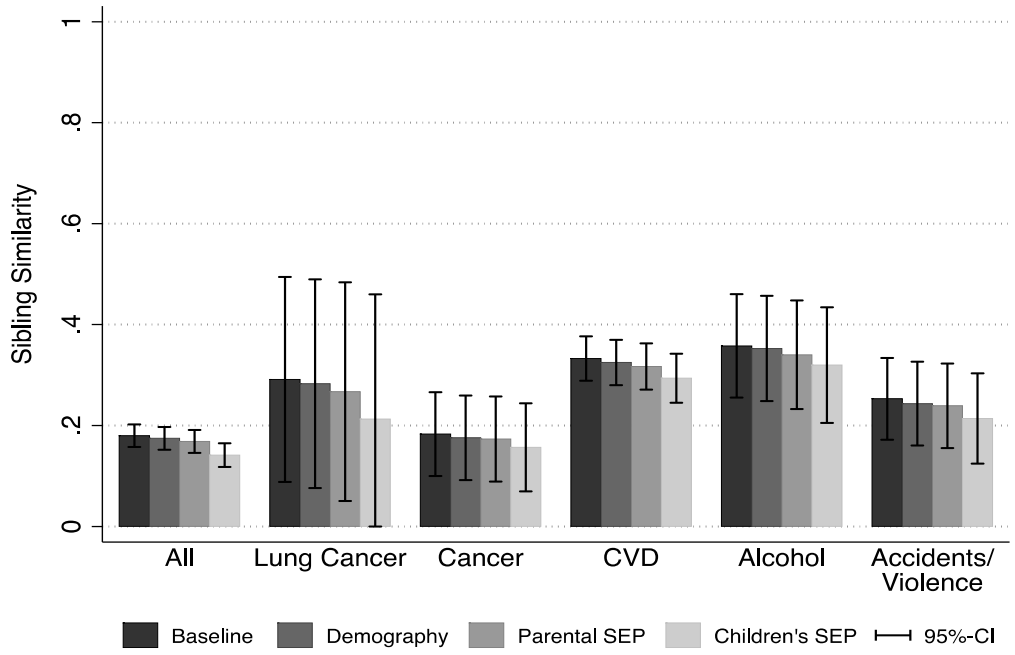
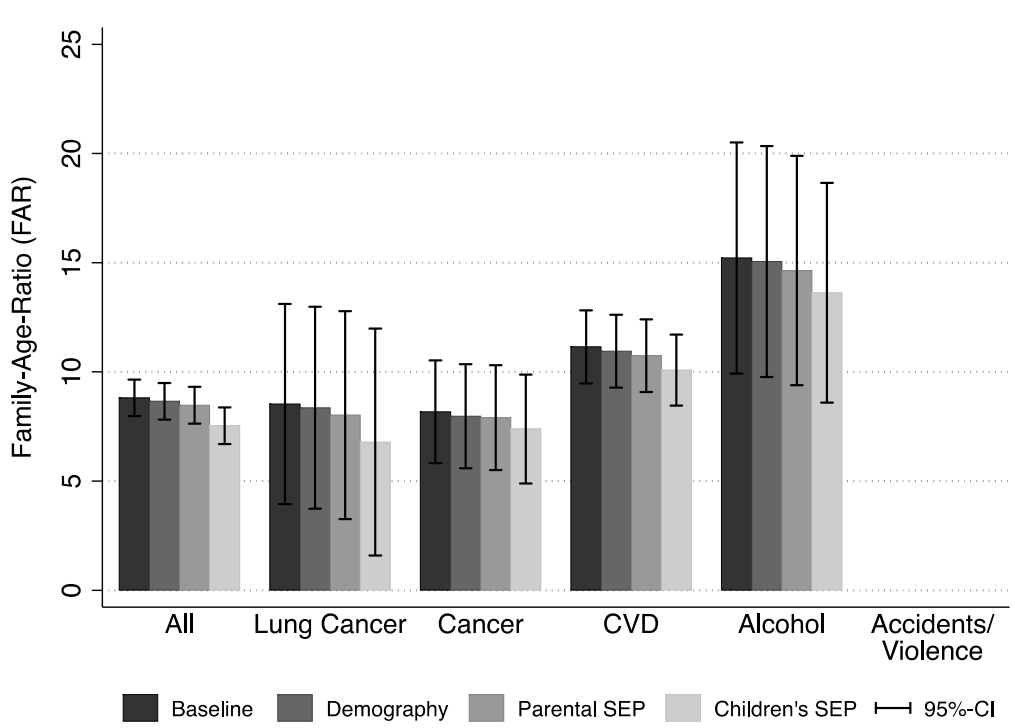


Figure 4 – Differences in family-age ratio between models and by cause of death



Note: Estimates for accidents and violence are missing because they were unreliably high, because the estimate for the shape parameter (the nominator) approached zero.

Figure 5 – Differences in median hazard ratio between models and by cause of death

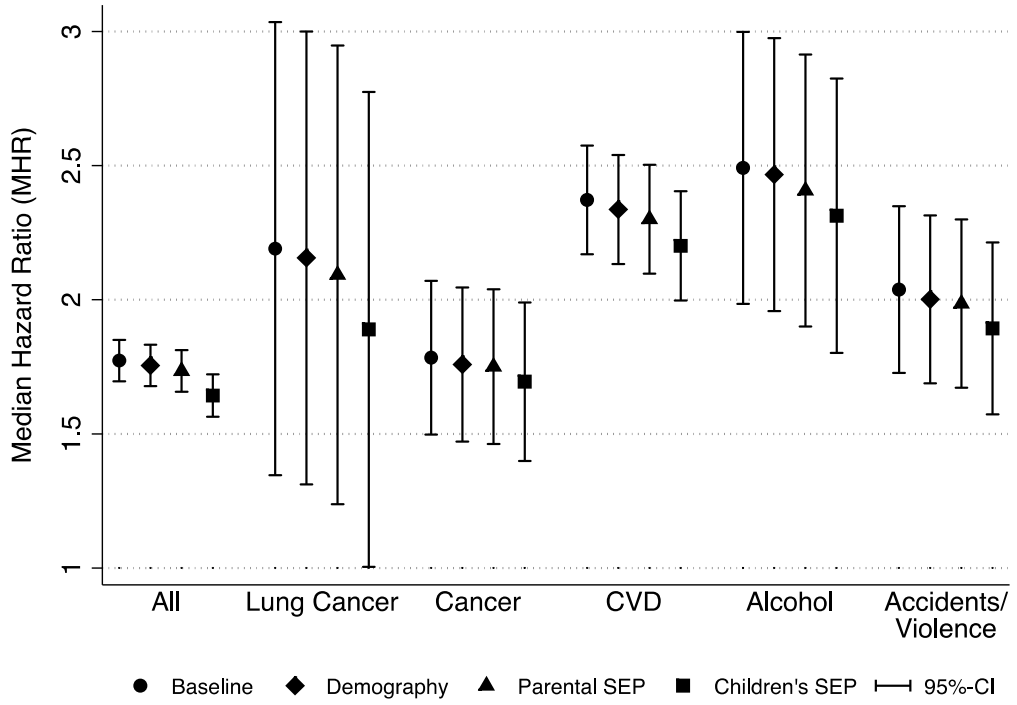


Table 1 – Summary statistics

	Mean	SD	Minimum	Maximum
Demography				
Birth cohort	*1943.73*	*4.19*	*1936.00*	*1950.00*
Age (years) at last observation	*61.62*	*5.69*	*35.00*	*71.00*
Female	*0.49*			
Native language:				
Finnish	0.93			
Swedish	0.07			
other	0.00			
Mother's age at birth:				
14-24	0.21			
25-35	0.58			
35+	0.19			
no valid info	0.01			
Father's age at birth:				
14-24	0.08			
25-35	0.52			
35+	0.33			
no valid info	0.06			
Region				
Western Finland	0.44			
Eastern Finland	0.38			
Lapland	0.06			
Uusimaa	0.12			
Number of siblings				
2	0.34			
3	0.27			
4	0.18			
5+	0.21			
Parental SEP				
Parental education:				
Did not go to school, unknown	0.17			
Primary school	0.73			
Past primary	0.10			
Father's occupational status				
Professional/administrative	0.14			
Workers & agriculture workers	0.41			
Farmers	0.27			
Farmers (10+ ha)	0.08			
Employer/self-employed	0.08			
Other, unknown	0.01			
Number of persons per heated room				
up to 1 person	0.05			
1-2 persons	0.37			
2-3 persons	0.28			
3 and more persons	0.29			
Unknown	0.01			
Siblings' SEP (at age 35)				
Siblings' Income				
1st decile	0.10			
2nd decile	0.08			
3rd decile	0.08			
4th decile	0.08			
5th decile	0.08			
6th decile	0.08			
7th decile	0.08			
8th decile	0.08			
9th decile	0.08			
10th decile	0.09			
Unknown	0.16			
Siblings' education				
Basic	0.53			
Lower secondary level	0.24			
Upper secondary level	0.11			
Lowest level tertiary	0.04			
Lower-degree level tertiary	0.03			
Higher-degree level tertiary	0.05			
Siblings' home ownership				
No owner	0.30			
Owns house/share	0.56			
Unknown	0.14			
Siblings' occupation				
Self-employed	0.11			
Upper white-collar	0.12			
Lower white-collar	0.26			
Blue-collar	0.31			
Other/unknown	0.20			
Siblings' employment status				
Employed	0.71			
Unemployed	0.02			
Homemakers	0.06			
Others/Unknown	0.21			

Table 2 – Influences of observed and unobserved family characteristics on all-cause mortality

	Baseline		Demography		Parental SEP		Siblings' SEP	
	HR	SE	HR	SE	HR	SE	HR	SE
Age (Gompertz shape parameter b)	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)	3.16***	(0.17)
Male # Age (Gompertz shape parameter b)	0.99**	(0.00)	0.99**	(0.00)	0.99**	(0.00)	1.00	(0.00)
Demography								
Native language (ref. Finnish):								
Swedish			*0.60****	*(0.03)*	*0.62****	*(0.03)*	*0.72****	*(0.04)*
other			0.70	(0.22)	0.67	(0.21)	0.78	(0.25)
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)	1.20*	(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)	1.00	(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)	0.96	(0.04)
Occupational status (ref. Professionals)								
Workers & 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)
3 and more persons					1.10	(0.06)	0.94	(0.06)
Unknown					0.99	(0.11)	0.89	(0.10)
Siblings' SEP								
Education (ref: Highest tertiary)								
Basic or unknown							1.67***	(0.12)
Lower secondary level							1.41***	(0.10)
Upper secondary level							1.34***	(0.09)
Lowest level tertiary							1.14	(0.09)
Lower-degree level tertiary							1.17	(0.11)
Income (ref. 10th decile)								
1st decile							1.96***	(0.10)
2nd decile							1.71***	(0.09)
3rd decile							1.52***	(0.08)
4th decile							1.40***	(0.07)
5th decile							1.30***	(0.07)
6th decile							1.22***	(0.06)
7th decile							1.06	(0.05)
8th decile							1.01	(0.05)
9th decile							0.98	(0.05)
Unknown							1.70***	(0.12)
Home ownership (ref: Home owner)								
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							1.13**	(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)
Others/Unknown							1.92***	(0.08)
Family-level variance (Θ)	0.36***	(0.03)	0.35***	(0.03)	0.33***	(0.03)	0.27***	(0.03)
Sibling Similarity	0.18***	(0.01)	0.17***	(0.01)	0.17***	(0.01)	0.14***	(0.01)
FAR	8.81***	(0.43)	8.66***	(0.43)	8.47***	(0.43)	7.54***	(0.43)
MHR	1.77***	(0.04)	1.75***	(0.04)	1.73***	(0.04)	1.64***	(0.04)
Total person years at risk				2598805				
Siblings				94042				
Families				32544				
Deaths				10948				

Table 3 – Relative frequency of causes of death in the sample of siblings

Cause of death	Freq.	Percent
Cancer	2462	22.49
Cardiovascular	3188	29.12
Alcohol	1085	9.91
Accidents and Violence	1952	17.83
Lung Cancer	613	5.60
Other	1321	12.07
	327	2.99
Total	10948	100

Appendix

Figure 6 – Sibling similarity for different models and distributional assumptions

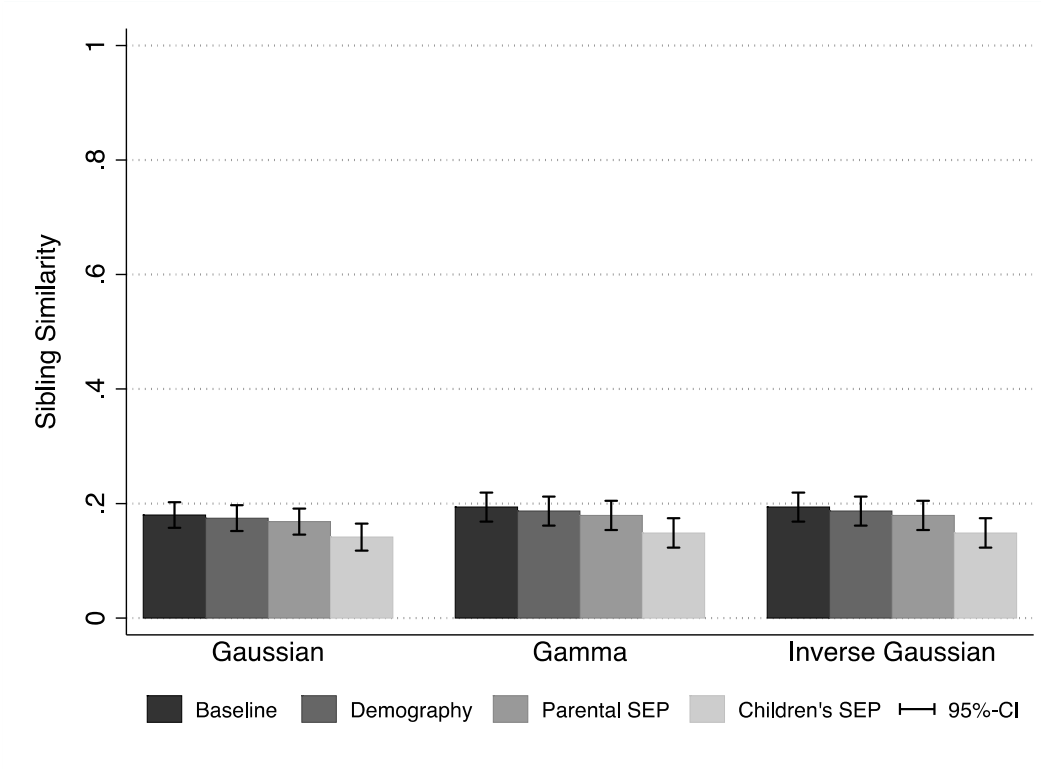


Figure 7 – Family-age ratio for different models and distributional assumptions

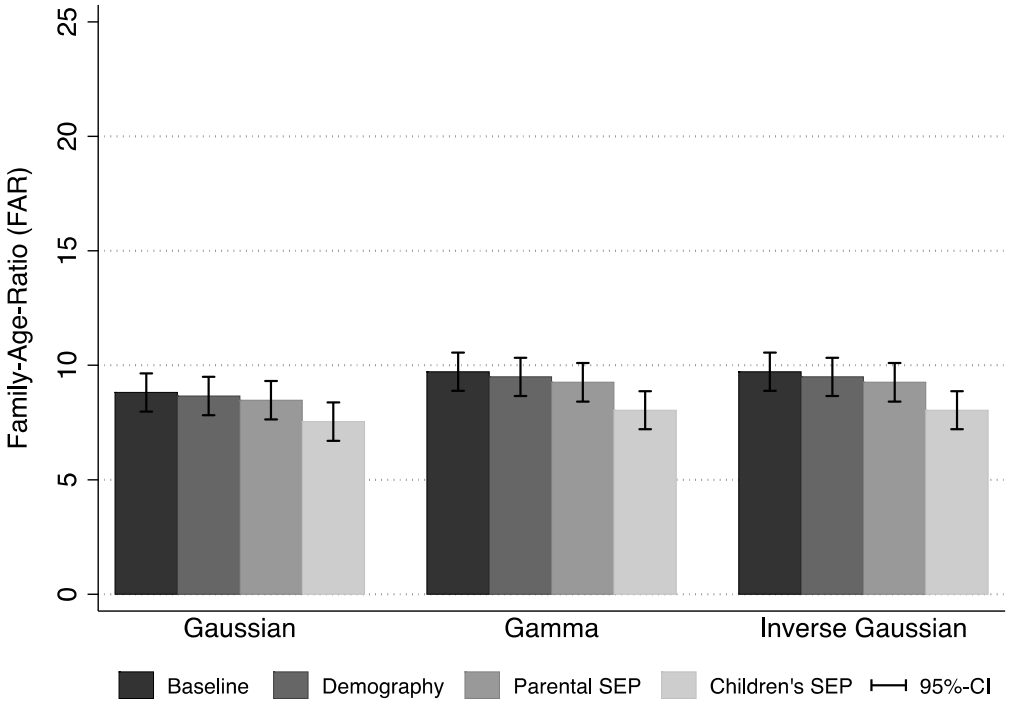


Figure 8 – Median hazard ratio for different models and distributional assumptions

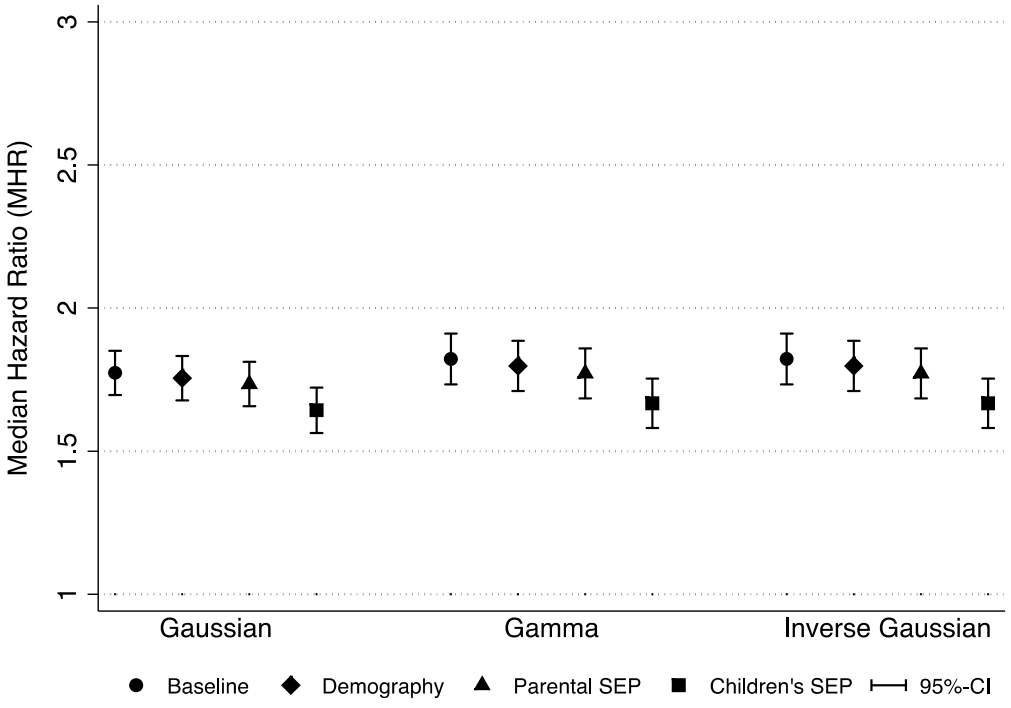
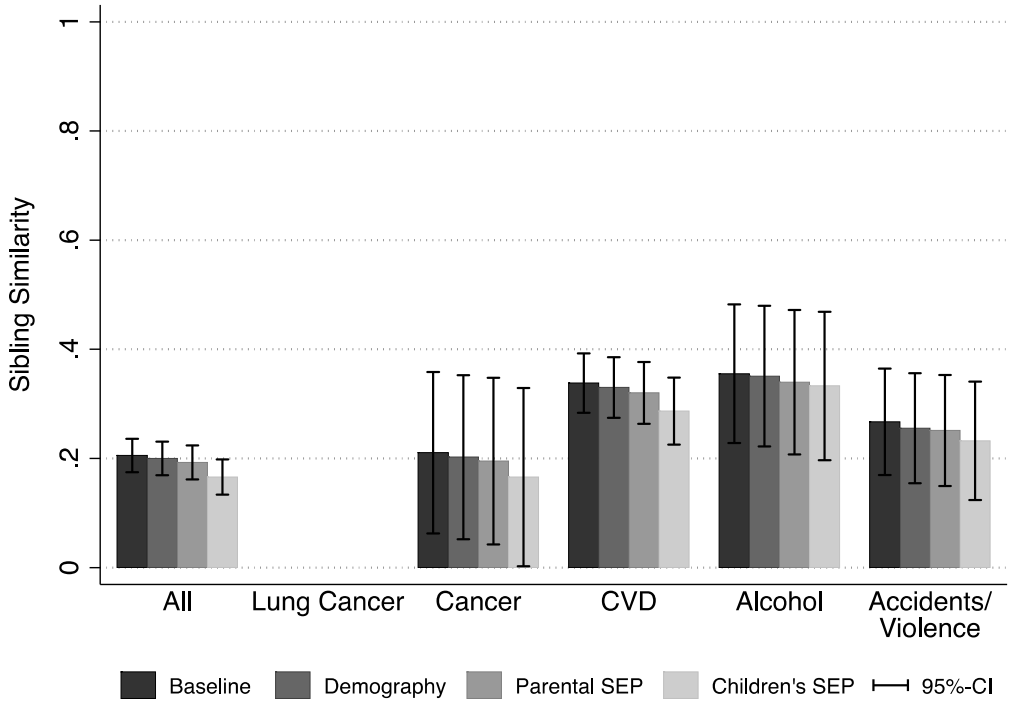
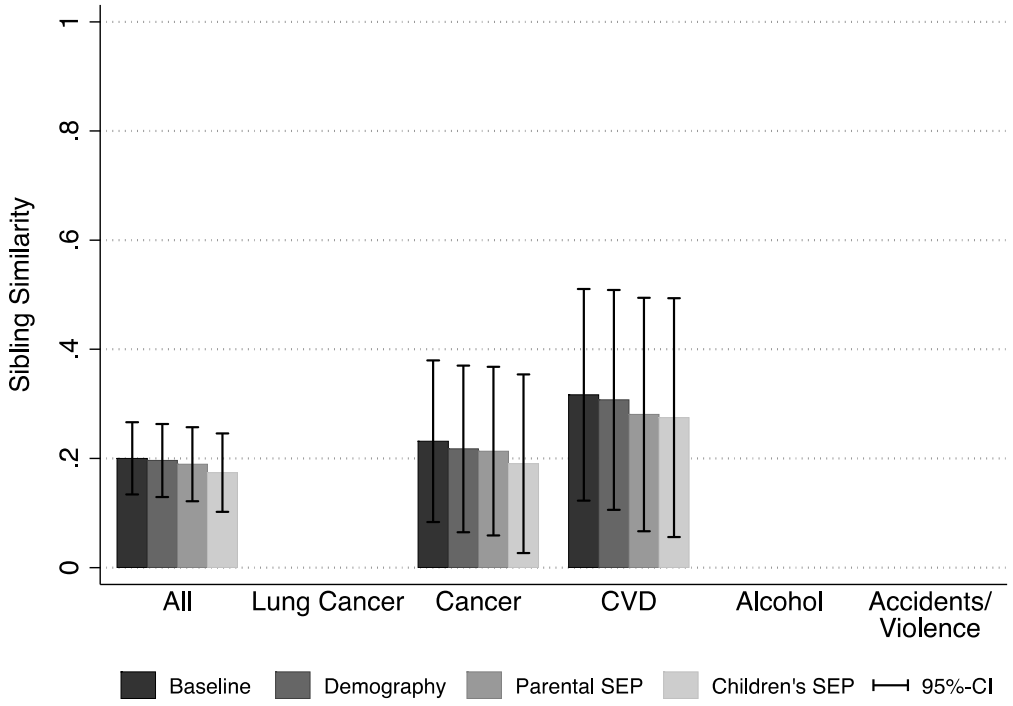


Figure 9 – Brother similarity in all-cause and cause-specific mortality



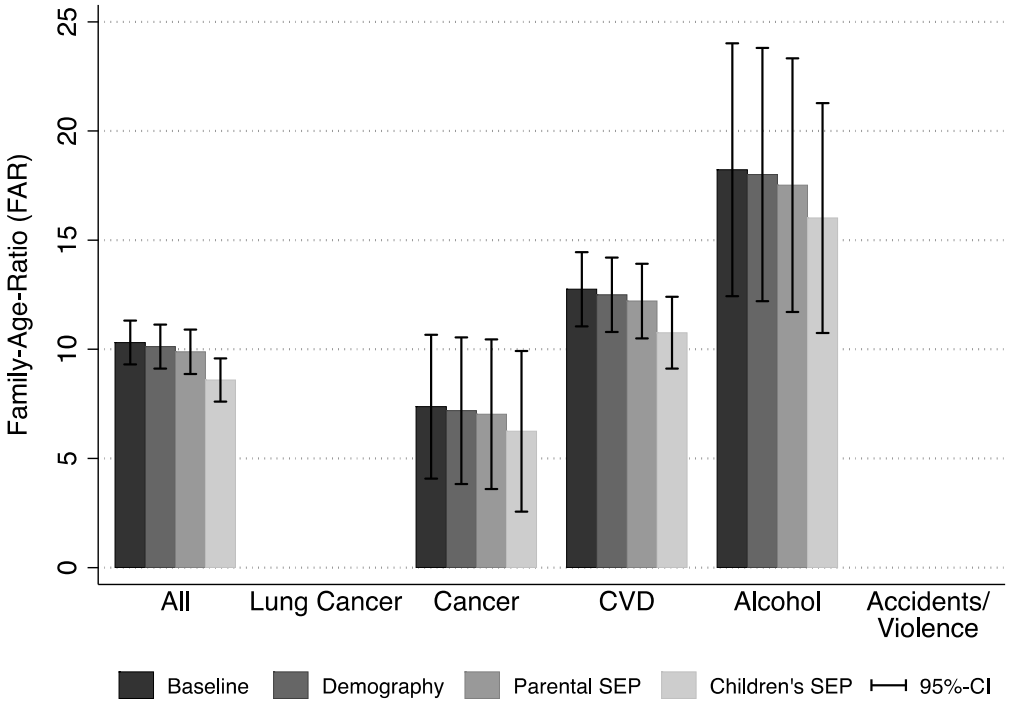
Note: There were not enough deaths related to lung cancer (<500) to get reliable estimates of the variation between families.

Figure 10 – Sister similarity in all-cause and cause-specific mortality



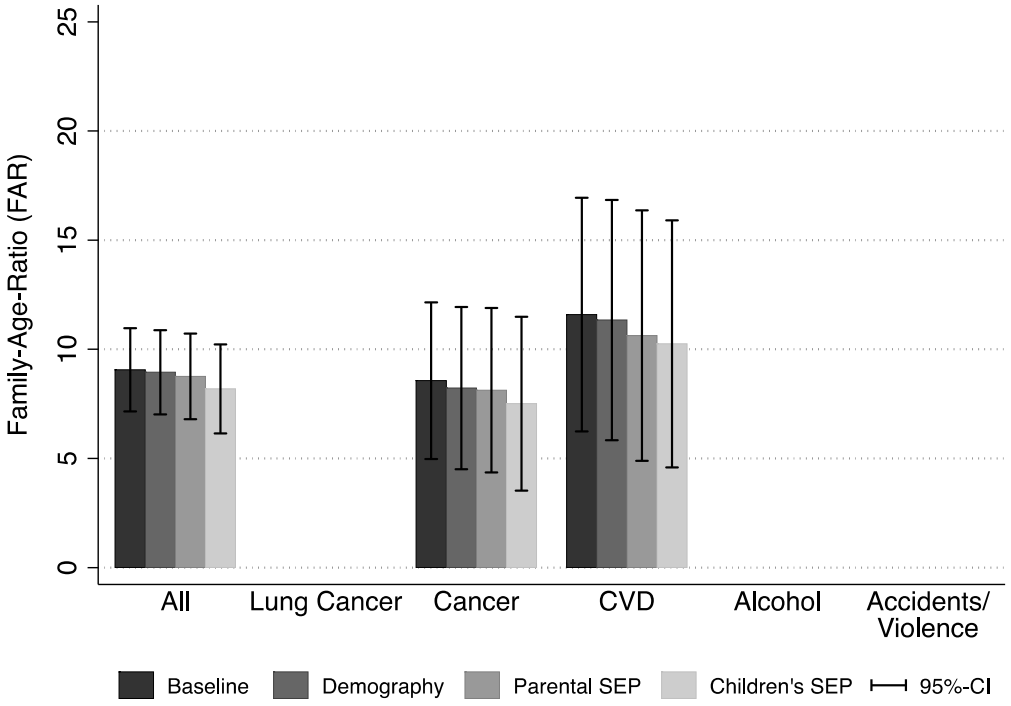
Note: There were not enough deaths related to lung cancer, alcohol, and accidents and violence (<500) to get reliable estimates of the variation between families.

Figure 11 – Family-ageratio in all-cause and cause-specific mortality - Brothers



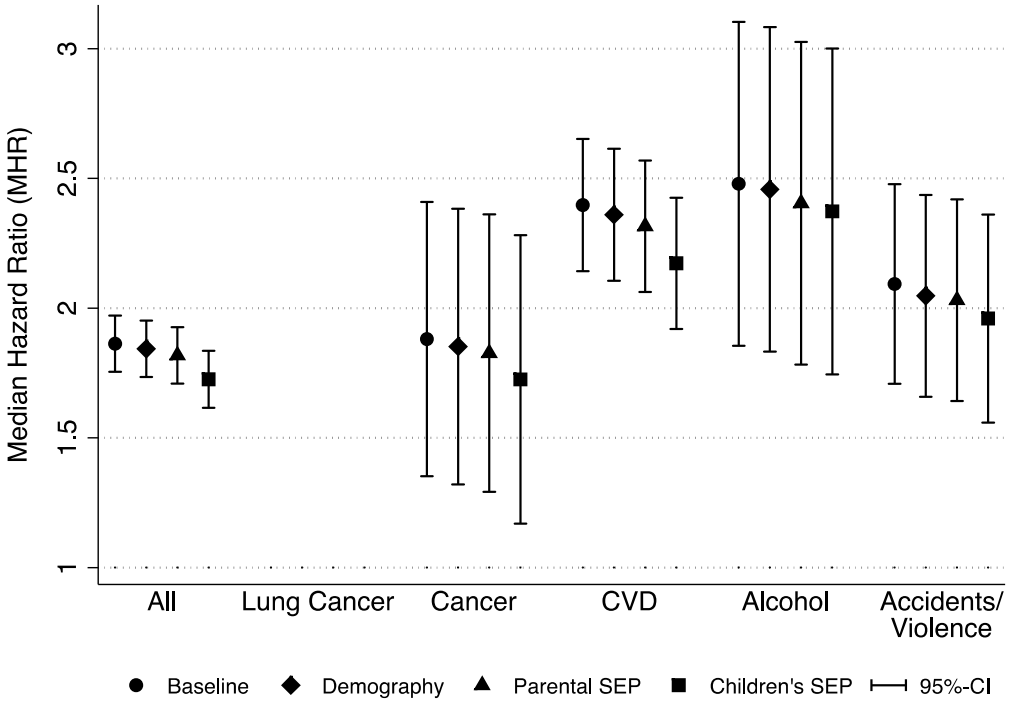
Note: There were not enough deaths related to lung cancer (<500) to get reliable estimates of the variation between families.

Figure 12 – Family-age ratio in all-cause and cause-specific mortality - Brothers



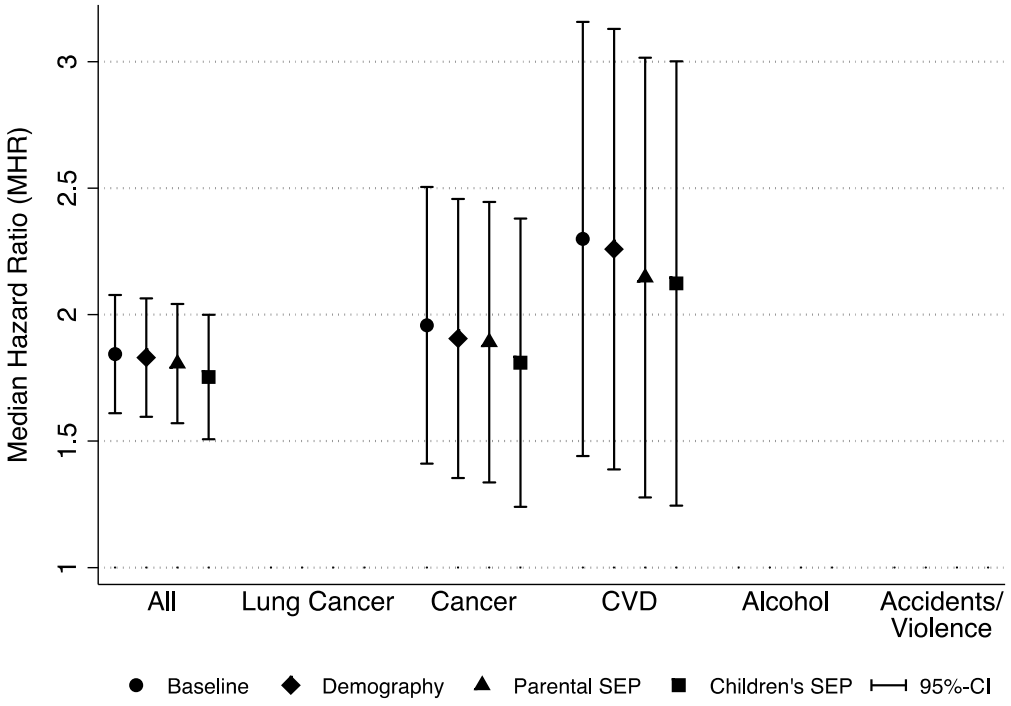
Note: There were not enough deaths related to lung cancer, alcohol, and accidents and violence (<500) to get reliable estimates of the variation between families.

Figure 13 – Median hazard ratio in all-cause and cause-specific mortality - Brothers



Note: There were not enough deaths related to lung cancer (<500) to get reliable estimates of the variation between families.

Figure 14 – Median hazard ratio in all-cause and cause-specific mortality - Brothers



Note: There were not enough deaths related to lung cancer, alcohol, and accidents and violence (<500) to get reliable estimates of the variation between families.