



Published in final edited form as:

J Gerontol A Biol Sci Med Sci. 2006 December ; 61(12): 1253–1261.

Longevity and Correlated Frailty in Multigenerational Families

Gilda Garibotti^{1,3}, Ken R. Smith^{1,2}, Richard A. Kerber⁴, and Kenneth M. Boucher⁴

¹Huntsman Cancer Institute, University of Utah, Salt Lake City.

²Human Development and Family Studies, University of Utah, Salt Lake City.

³Centro Regional Universitario Bariloche, Universidad Nacional del Comahue, Argentina.

⁴Department of Oncological Sciences, University of Utah, Salt Lake City.

Abstract

Multigenerational pedigrees provide an opportunity for assessing the effects of unobserved environmental and genetic effects on longevity (i.e., frailty). This article applies Cox proportional hazards models to data from three-generation pedigrees in the Utah Population Database using two different frailty specification schemes that account for common environments (shared frailty) and genetic effects (correlated frailty). In a model that includes measures of familial history of longevity and both frailty effects, we find that the variance component due to genetic factors is comparable to the one attributable to shared environments: Standard deviations of the correlated and the shared frailty distributions are 0.143 and 0.186, respectively. Through simulations, we also show a greater reduction in the bias of parameter estimates for fixed covariates through the use of the correlated frailty model.

ASSESSING the sources of variation in human life span is a fundamental objective in biodemography and gerontology. The familial component of longevity has been a topic of considerable interest over the past century (1–19). Family-based studies have, in general, provided support for a modest genetic influence on life span. There is a small, simple correlation between the age at death of parents and offspring and stronger correlations between the ages at death of siblings (1,5,11–13). Several investigators found that a heritable component is present mainly in later-life survival and that life span does not seem to be heritable if parents live shorter lives (6,20). Gender differences in the inheritance of longevity have also been reported (6,20,21). Heritability estimates for age at death vary from nearly 0 (13) to 0.33 (22). In an attempt to separate the impact of genetic factors and the effect of family environment, several investigators have generated heritability estimates based on twin data. Studies comparing monozygotic and dizygotic twins have supported the prediction that life span is more correlated in monozygotic than dizygotic twins (9,22–24). Sorensen and colleagues' (14,15) studies of longevity in adopted children showed that premature death in adults has a strong genetic component.

It is clear that genetic factors affect adult survival probabilities, although the relative importance of environment and genes is still not fully determined. The opportunities for researchers to conduct studies that address this question have grown due to the increasing availability of survival data for families and multigenerational pedigrees. Data from large pedigrees contain valuable information with which to assess the role of genetics and environment in understanding variability in longevity. Accessible and flexible statistical

tools needed to analyze these complex pedigrees still need to be fully developed and evaluated.

Historically, longevity data have been studied by calculating correlations and analysis of variance between uncensored life spans of siblings as well as between parents and offspring. When the data contain censored observations, the analysis of variance approach is not appropriate and instead survival methods need to be considered, such as the widely used Cox proportional hazards model. One of the basic assumptions in the Cox model is independence of survival times of individuals given the observed values of covariates. However, in studies involving multiple individuals from the same family, the independence assumption is not plausible unless all important familial factors were measured and controlled for in the model. It is now standard practice to adjust for these family-based sources of statistical dependencies in survival models through the introduction of a frailty component. In this usage, frailty refers to a susceptibility to death that is not captured by observed covariates. Typically, frailty includes factors that affect an individual's survival chances such as genes and unmeasured attributes of the environment, all of which may or may not be shared to some degree with others in a family or pedigree. The most common frailty specification is the "shared frailty" extension of the proportional hazards regression model. The way in which shared frailty has been used in these models has largely been to collapse all factors that are not measurable into a single random effect that is shared by individuals within the group (e.g., family). Guo (25) and Guo and Rodriguez (26) used survival techniques that incorporate a shared frailty component to study genetic and environmental influences on longevity among sets of siblings across numerous families. They interpreted the frailty component as the sum of unobserved shared factors that were likely to affect longevity, including genetic effects shared among siblings, shared gene-gene interactions, common influence of parental competence, and other shared household effects that were not captured by observed covariates in their study. Hougaard and colleagues (27) used several versions of the frailty model for bivariate survival to fit data for Danish monozygotic and dizygotic twins. In a shared frailty model, the frailties are unobserved random variables assumed to be independent and to follow a probability distribution, the shape of which is described by a few parameters. These models provide insight into the effects of familial risk on mortality, but have not been designed to incorporate complex genetic relationships found in family data of varying size and structure. In other words, the shared frailty model pools all effects, shared genetic and shared environmental, into a single random effect without being informed by the genetic relationships that link any two relatives.

In order to describe more complicated dependencies, Yashin and Iachine (28) and Yashin and colleagues (29) introduced a model of bivariate survival that allows for the incorporation of correlations between individual frailties. These models are known as "correlated frailty" models, which have been used with twin data to assess genetic and environmental factors influencing mortality. Ripatti and Palmgren (30) and Therneau and colleagues (31) extended correlated frailty models to survival data on n individuals, rather than on two individuals (as is the case in twin studies). Unlike the shared frailty model, in which frailties are assumed to be independent, in the correlated frailty model proposed by Ripatti and Palmgren (30) and Therneau and colleagues (31), one individual's frailty is associated (but not necessarily perfectly) with the frailty of another individual who is from the same family of origin (related genetically) or family of procreation (related by marriage). In this case, frailties are assumed to be random variables drawn from a multivariate normal distribution with an arbitrary covariance structure.

Pankratz and colleagues (32) recently applied the correlated frailty model to investigate the aggregation of breast cancer within families. They considered a correlated frailty model with

one random effect per individual, with the random effects correlated to reflect the degree of genetic relationship between individuals. They interpreted the variance of this model as a polygenic (involving two or more genes) variance component and concluded that there is significant heritability of age-at-onset of breast cancer. This model assumed that family risk depends only on genetic relationships. Pankratz and colleagues (32) also explored the possibility of a shared family environment role in breast cancer by fitting a model with both polygenic and shared family random effects. They provided evidence for polygenic variance components and suggestive evidence for shared family environmental variance components.

The purpose of this article is to report how the correlated frailty model constitutes a valuable data analysis tool to incorporate both unobserved genetic and environmental influences in human longevity that allows for the inclusion of observable covariates. The utility of a correlated frailty model that incorporates both genetic and environmental sources of frailty is described here and applied to data from an ongoing family-based study of longevity called the Fertility, Longevity, and Aging (FLAG) study. We also examine, via simulations, how ignoring the dependencies among observations affects the estimation of the regression parameters and their standard errors. The problem of model misspecification related to ignoring the true dependencies in the data has been studied by Wei and colleagues (33) in the context of recurrent event data, and by Guo (25) in the case of shared frailty models. Our results point to the importance of using correlated frailty models to analyze complex pedigree data that allow for both genetic and environmental contributions to frailty.

Methods

Shared Frailty Model

The shared frailty model extends the proportional hazards model by introducing an unobserved frailty term. Suppose that there are n individuals, $i = 1, \dots, n$, who are each members of one of q families, $j = 1, \dots, q$. The conditional hazards function for individual i is:

$$\lambda_i(t|b) = \lambda_0(t) \exp(X_i\beta + Z_i b), \quad (1)$$

where λ_0 is an unspecified baseline hazard function, $b = (b_1, \dots, b_q)$ is a vector of random effects that represent family-specific frailties, assumed independent, X_i is a vector of measured covariates, β is a vector of unknown regression coefficients, and Z_i is a vector with $Z_{ij} = 1$ if individual i is a member of group j , and 0 otherwise. The distribution of b_j is specified a priori. Distributions of b_j commonly considered are the log gamma and the normal distribution for their computational advantages.

The shared frailty model is attractive because it explicitly acknowledges the potential role of unobserved factors affecting an individual's risk of mortality. It does, however, assume that unobservable characteristics are perfectly shared with others in the family and that unobserved factors that are not shared are not considered.

Correlated Frailty Model

The correlated frailty model differs from the shared frailty model in that it allows for individual-level frailties that can be correlated across individuals within the family. We describe the correlated frailty model that incorporates unmeasured random effects into the analysis of censored survival data applied to a family-based sample of size n . Let X_i and Z_i be vectors of known covariates, and $b = (b_1, \dots, b_q)$ a vector of frailties. Given b , the event times are assumed independent with the conditional hazard function for individual i following the proportional hazards specification

$$\lambda_i(t | b) = \lambda_0(t) \exp(X_i\beta + Z_i b), \quad (2)$$

where λ_0 is an unspecified baseline hazard function and β is a vector of unknown regression coefficients. It is assumed that the distribution of the vector of frailties belongs to a known family of distributions with mean 0 and covariance matrix $D(\sigma)$, with σ denoting a vector of unknown parameters.

We assume that b follows a multivariate normal distribution with mean 0 and covariance matrix $D(\sigma)$. This assumption makes it computationally convenient to impose the desired covariance structure for the frailty distribution. Ripatti and Palmgren (30) give an approximate likelihood for this model and propose a computational approach for calculating the approximate maximum likelihood estimates of β and σ . Therneau and colleagues (31) implemented the approximate maximum likelihood estimation approach as R software, in the Kinship package (this software can be downloaded from <http://cran.r-project.org/>). This model includes as a special case the shared frailty model with normal and log normal distributions.

An important property of the correlated frailty model is that it allows for the simultaneous incorporation of unobserved genetic and environmental frailty parameters in the analysis of life-span data. As a result, genetic and environmental components of frailty can be jointly estimated from the data.

We assume that the conditional hazard of death for individuals i , $\lambda_i(t | b)$, follows the model given in Equation 2. To evaluate the separate influences of unobserved genetic and environmental contributions to life span, we decompose the covariance matrix $D(\sigma)$ into two components:

$$D(\sigma) = \sigma_f^2 \Sigma_f + \sigma_p^2 \Sigma_p, \quad (3)$$

where σ_f^2 and σ_p^2 represent the shared environmental influences and the shared polygenic effect, respectively, and $\sigma = (\sigma_f, \sigma_p)$. The matrix Σ_f is a fixed matrix that incorporates the degree of shared environment among individuals, whereas Σ_p captures the shared polygenic factors between genetically related family members.

In our analyses, the elements of Σ_p represent the degree of genetic resemblance that two individuals would be expected to have by chance. The Σ_p matrix is a function of $K = (K_{ij})$, a matrix of kinship coefficients. The kinship matrix has elements K_{ij} , the kinship coefficients, that measure the probability that a gene selected randomly from individual i and another selected from individual j will be identical by descent (ibd) at a given locus. Thus, if we assume no consanguinity in previous generations, the kinship coefficient is 0.25 for first-degree relatives (parent/child and sibling pairs), 0.125 for second-degree relatives (aunts/uncles paired with nieces/nephews), and 0.0625 for third-degree relatives (first cousins), etc. Kinship coefficients represent half the expected proportion of the genome that is shared by pairs of relatives. To generate Σ_p , the kinship coefficient for the diagonal element is first set to 0.5 and $\Sigma_p = 2K$. Hence, Σ_p has values of 1 on the diagonal, 0.5 for parent/child and sibling pairs, 0.25 for grandparent/grandchild, uncle/niece, and so forth.

The matrix $\Sigma_f = (a_{ij})$ with $a_{ij} = 1$ if individuals i and j belong to the same family and 0 otherwise. By "family," we mean the group of all individuals that are related genetically or by marriage. Clearly, the definition of family in this respect can be made more flexible for a given application. Note that, in the case where two individuals are from different families,

the elements of their Σ_p (shared polygenic effects) and Σ_f (shared environmental effects) matrices will both be 0. Our choice of Σ_f in this analysis restricts all individuals in the family to have the same family-specific effect. Other correlation structures that incorporate alternative environmental relationships between individuals in the data set could be accommodated with this model.

Data: Exceptional Longevity in Families in the FLAG Study

Cox proportional hazards models with correlated frailty are applied to data that were generated as part of the ongoing FLAG project. The larger study selects individuals from long-lived (LL) pedigrees who have survived to extreme ages to identify genetic and environmental factors that have allowed them to experience exceptional survival and to determine whether they have fewer disabilities and lower rates of age-related diseases over the course of their lives. The overall objective of the FLAG study is to measure aging-related demographic, epidemiologic, social, cognitive, physiological, and molecular traits in these exceptional individuals to identify a “delayed aging phenotype” and to isolate genetic markers associated with slower rates of aging.

The Utah Population Database (UPDB) is used to identify the families that participate in the study. The UPDB contains over 8 million records, including the genealogies of the founders of Utah and their descendants. In the 1970s, approximately 170,000 Utah nuclear families were identified on “Family Group Sheets” from the archives at the Utah Family History Library, each with at least one member having had a vital event (birth, marriage, death) on the Mormon Pioneer Trail or in Utah. These families have been linked across generations; in some instances, the records span seven generations. The UPDB is an active genealogy; new families and their members are continually being added as the UPDB is linked to other sources of data, including birth and death certificates. Additional information on these individuals comes from sources such as driver's license and the Utah Cancer Registry. The UPDB now holds data from migrants to Utah and their descendants that include more than 1.8 million individuals born from the early 1800s to the mid-1900s and that are linked into multigeneration pedigrees.

We used familial excess longevity (FEL) as a genealogically based method for identifying LL pedigrees. FEL is a summary measure of excess longevity (EL) among all blood relations for a given individual (34). Calculating FEL first requires an estimate of the difference between an individual's attained age and the age to which that individual was expected to live according to an accelerated failure time model. This model uses three covariates that are available in the UPDB and that are associated with longevity: gender, birth year, and affiliation with the Church of Jesus Christ of Latter-day Saints (or LDS Church or Mormons). The variable describing whether a person in the UPDB is a member of the LDS Church is included because it is known that active members of the LDS Church have significantly longer lives (35).

The expected age at death is:

$$\hat{y} = \exp [\beta_0 + \beta_1 (\text{Gender}) + \beta_2 (\text{Birth Year}) + \beta_3 (\text{Religious Affiliation})] \quad (4)$$

with β_m regression coefficients, $m = 0, \dots, 3$. The values of the birth year variable are limited to the birth years represented in the sample analyzed. Its inclusion reflects an effort to control for secular increases in age at death. EL is $l = y - \hat{y}$, where y is the observed age at death or the age at the time last confirmed alive, in years. The FEL for an individual is calculated as the weighted average of individual EL of all blood relatives aged 65 years old or older. For a given individual, “blood relatives” refers to all persons aged 65 years old or

older in the database with whom the individual has a genetic relationship, regardless of their birth year and degree of genetic relationship. The number of relatives used to generate any one individual's FEL ranged from 62 to 105,893 with a median of 1987. The weights are the kinship coefficients. The FEL for individual i is

$$FEL_i = \frac{\sum_{j \in J} K_{ij} l_j}{\sum_{j \in J} K_{ij}} \quad (5)$$

where J is the set of all blood relatives of individual i .

In this study, we tested whether there is evidence that genetic and environmental factors are important determinants of age at death based on 146 kindreds of founders in the UPDB with very high FEL (the LL families) and 179 matched control kindreds. Based on the entire UPDB, LL families were identified as the top families (identified by their founder) (listed in terms of increasing p value) with a statistically significant ($p < .01$) excess of descendants surviving past the 97th percentile of EL or were members of the top founders most often selected by a Monte Carlo selection procedure. For the Monte Carlo procedure, variation was introduced by sampling from sibships in the data with replacement. During each iteration of the Monte Carlo procedure, the most likely LL founders were selected based on the EL of the sampled sibships. A key feature of this procedure was that each sibship was allowed to contribute to at most one of the selected founders in each iteration. Both methods were used in an effort to identify LL families that did not rely on a single selection method. The matched kindreds had founders with normative FEL values with comparable numbers of descendants as those among the exceptional FEL founders. Analyses were conducted excluding the founders themselves to eliminate the bias associated by including the exceptional founders (as they are, by definition, LL). Although all founders examined here generally have many descendants spanning up to 10 subsequent generations in the UPDB, only their children and grandchildren were used in the analysis. This was done to allow for a wider range of relationships than is done with twin studies (siblings, parents and offspring, uncle/aunt and niece/nephews, and first cousins) but also to restrict our assessment of environmental effects to circumstances where relatives are likely to share common environmental exposures. This restriction also allowed for less right-censoring because these children and grandchildren were born earlier in history and hence are more likely to be older (and, quite often, deceased). Our study focused on survival past age 50; therefore, only individuals 50 years old or older from the selected families are incorporated in the analysis. We required that individuals had to have survived to age 50 to focus our analysis on the age range during which mortality begins to rise but also to maximize our sample size as much as possible. In all, 7273 individuals 50 years old or older were used as the basis for the analysis, comprising approximately 23,300 sib pairs, 5600 parent/offspring pairs, 22,900 uncle/aunt–niece/nephew pairs, and 53,000 first-cousin pairs.

Results

Descriptive statistics for individuals in LL and control kindreds are shown in Table 1. There were 1416 women and 1550 men among the individuals in the LL pedigrees and 2078 women and 2229 men in the control group. Given the range of birth years, there are right censored observations although most births occur in the 1800s. Accordingly, right censoring was limited (and similar) across the two groups with 9% for the LL group and 10% for the control group. The percentage of individuals who are identified as active members of the LDS Church in the LL group was 62.9% and in the control group 58.6%.

We fit a series of mixed effects proportional hazards models that allow us to assess the influence of genetic and environmental factors in longevity. We include four fixed covariates that influence longevity: birth year, gender, affiliation with the LDS Church, and FEL. A fifth covariate is also included that takes the value 1 if individuals belong to an LL kindred and is equal to 0 if they are members of a control kindred. The LL variable was derived from the survival experience of individuals in the sample and was used as the basis for sampling individuals from both longevous and normative families. The LL variable is therefore included as a control for the sample design in which we explicitly oversampled persons from LL pedigrees. Given this sampling strategy, we would expect the LL variable to be a significant predictor of longevity. What is not known is the extent to which it is associated with individual survival and the degree to which the association between LL and survival is affected by the model specification, especially how frailty is considered. FEL is included in the same model to assess how a family history of longevity, when measured as a continuous variable, affects survival after adjusting for the effects of oversampling LL pedigrees through the inclusion of the LL variable.

The first model is a proportional hazards model that ignores the potential statistical dependencies between individuals in the same family (relations attributable to common genes or environment) and corresponds to the standard proportional hazards model. This means that we treat all individuals as being statistically independent after controlling for the covariates. The second model assumes that all members of a family share the same unmeasured risk and represents the shared frailty model. We interpret frailty as the combined shared environmental and genetic influences that are not measured by FEL and the LL dummy variable. The third model, representing the correlated frailty model, also allows for familial risk, but this risk depends only on genetic relationships and does not explicitly consider any shared environmental effects. In the decomposition of the covariance matrix given in Equation 3, this model corresponds to $\sigma_f^2=0$. The final and most general model incorporates both shared and correlated frailties. This model provides separate estimates for the role that genetic and shared environments play in the mortality hazard rate.

In Table 2, we present the hazard rate ratios ($\exp(\beta)$) and their 95% confidence intervals (CI). The four models use the same data and have the same covariates; the difference between them is the manner in which frailty is modeled. All the parameter estimates except σ_f^2 and σ_p^2 represent relative risks. With respect to the observed covariates, all show, in a qualitative sense, the same effects although some differences exist. These models show that male mortality hazard rates are 23%–27% higher than those of females and that as birth year increases by 1 year individuals enjoy a reduction in their hazard rate by 1%. Individuals affiliated with the LDS Church have hazard rates that are not significantly different than those of persons not affiliated with the LDS Church. The null effect of LDS affiliation is unexpected but is partially explained by the fact that FEL is included in the model and it controls, indirectly, for the effects of LDS status. For this cohort, men benefit much more from their affiliation with the LDS Church than do women. The results shown here pool men and women in the same sample leading to a weaker association with mortality. An interaction of gender and LDS affiliation was not estimated for reasons of parsimony. As expected, both FEL and the indicator variable for being in an LL kindred indicate that individuals with a family history of exceptional longevity have significantly lower rates of mortality. Individuals sampled from LL kindreds have a mortality risk that is 14%–19% lower than that found in control kindreds; for each year of increase in FEL there is a 5%–6% decrease in the mortality hazard rate.

The standard deviation (*SD*) of the frailty random effect (σ_f) from the shared frailty model is 0.189 (95% CI, 0.150–0.229). This estimate indicates that typical values of the family-

specific risk of mortality are approximately 21% larger or smaller [$\exp(0.189) = 1.208$] than the overall mortality risk, *ceteris paribus*, suggesting an important shared familial influence in life span that is unaccounted for by the included covariates. This model assigns the same familial relative risk to all members of a family without taking explicit account of shared genetic effects between individuals in the same kindred. However, it is not possible in this model to rule out some common shared alleles among family members.

The correlated frailty (polygenic) model incorporates an individual-specific random effect that is allowed to be correlated with the frailties of other relatives, as indicated by the kinship matrix K . The *SD* of the frailty distribution for the polygenic frailty model (σ_p) is 0.338 (95% CI, 0.258–0.418). This result indicates that the individual relative risk of mortality is likely to be as much as ($\exp(0.338)$) 40% larger or smaller than the average risk. This provisional result is consistent with the idea that life span has a sizable heritable component.

The shared frailty model conflates genetic with shared environmental factors, and the correlated frailty model assumes that familial risk depends only on genetic variants that are shared to varying degrees among blood relatives. The model with both shared family and correlated random effects represents a more general approach that incorporates genetic and nongenetic traits separately. The *SD* of the correlated frailty distribution (σ_p) in this model is 0.143 (95% CI, 0.000–0.282), and the *SD* of the shared frailty distribution (σ_f) is 0.186 (95% CI, 0.147–0.230). Given the model and this sample of kindreds, both shared familial and polygenic factors are found to play a role in explaining the variation in longevity although the size of σ_f is larger. As two of the fixed covariates, FEL and LL, are likely correlated with unobserved genetic factors and environmental factors, it is difficult to interpret the difference in the magnitude of the estimates of σ_f and σ_p . In an attempt to determine whether the difference in the size of the estimates of σ_f and σ_p is an indication that shared environmental factors play a larger role than genetic traits on the mortality hazard rate, we refit the model with the same decomposition of the covariance matrix but with only birth year, gender, and affiliation with the LDS Church as fixed covariates. In this model, the estimated *SD* of the correlated frailty is 0.265, whereas the *SD* for shared frailty is now 0.294. In comparing the two sets of results, we make two observations. First, the small difference in the size of the two new *SD* values suggests that shared environmental factors play a similar albeit slightly larger role than genetic variants shared among sets of relatives. Here, the small difference between the two *SD* values (i.e., a relatively large *SD* for the polygenic correlated frailty effect in relation to what the literature suggests) may reflect the stronger role played by alleles associated with longevity arising from our selection of families characterized by EL. The second observation is that the exclusion of both the LL and FEL variables resulted in similar increases in the *SD* for both the shared and correlated frailty distributions. This finding supports the idea that the LL and FEL variables capture factors that have both environmental and genetic origins.

Effect of Ignoring Frailties on the Estimates of a Cox Proportional Hazards Model

Analysts frequently possess data based on individuals clustered into families and kindreds, and yet the analysis is conducted under the assumption of independence of survival times given the observed values of covariates. Using simulated data from correlated frailty models, we examined whether ignoring the associations in the data biases the relative risk estimates and, if so, to what degree. Additionally, we investigated whether the standard errors of the parameter estimates are inefficient under the assumption of independence.

We performed 1000 computer simulations of 100 families that represent the composition and structure of families in UPDB. The data comprise two generations within a family with four basic types of relationships: parents and offspring, siblings, aunts/uncles with nieces/

nephews, and first cousin dyads. Ages at death were simulated using a proportional hazards model. The baseline survival function was assumed to follow a Gompertz distribution, that is,

$$\lambda_0(t) = \alpha \exp(\lambda t), \quad (6)$$

with $\lambda = 0.1$ and $\alpha = 0.007$. The value of α represents the proportion of individuals in our sample that survived to age 50 years and died before reaching age 51 years (36,37). The value $\lambda = 0.1$ is consistent with an adult force of mortality that doubles every 7 years $\exp(0.1 \times 7) = 2$. The mortality hazard rate increased by a factor of $\exp(\beta)$ for a unit increase in a covariate X . The variable X followed a normal distribution with mean 2.0 and SD 1.9. Three different values for the regression parameter β were considered, $\beta = 2.0$, $\beta = 0.7$, and $\beta = 0.1$. These values represent hazard ratios of 7.4, 2, and 1.1, respectively. We also assumed that the risk of dying is influenced by an additive polygenic effect. Hence, the hazard rate is specified as:

$$\lambda_i(t | b_i) = \lambda_0(t) \exp(\beta X_i + b_i), \quad i = 1, \dots, n \quad (7)$$

where $b = (b_1, \dots, b_n)$ is a normally distributed random vector with mean 0 and covariance $\Sigma = \sigma_p^2 K$, with K the kinship matrix. This corresponds to $\sigma_f^2 = 0$ in Equation 3. We further assumed that the fixed covariates X_i , $i = 1, \dots, n$, are independent given b .

Two common but different situations are evaluated: first, the covariates X_i , $i = 1, \dots, n$, are generated to be independent of the polygenic frailty b ; second, X_i are specified so that they are correlated with the polygenic frailty b , with covariance between X_i and b_j equal to 0.2 for $i = j$, and 0 otherwise. The value 0.2 was chosen arbitrarily to represent a moderate degree of association between the known covariate and the unknown factor.

The effect on the parameter estimates of ignoring the genetic relationship in the data was assessed using two different models: a proportional hazards model and a shared frailty model with variance component σ_f^2 . The correlated frailty model is also estimated. In what follows, $\widehat{\beta}$, $\widehat{\sigma}_f^2$, and $\widehat{\sigma}_p^2$ will refer to the maximum likelihood estimators of β , σ_f^2 , and σ_p^2 , and $SD(\widehat{\beta})$ is the estimator of the standard deviation of $\widehat{\beta}$.

Table 3 shows results for the situation in which the simulated data sets were generated according to a model with X_i independent of the frailty b . Table 3 gives the empirical expected value of $\widehat{\beta}$, $SD(\widehat{\beta})$, $\widehat{\sigma}_f^2$, and $\widehat{\sigma}_p^2$; that is, the average of the maximum likelihood estimates from each of the 1000 simulated data sets. Within Table 3, the empirical SD values are given in parentheses.

Based on these simulations, we note a bias in the estimation in β . The estimate of β is seriously underestimated when fitting the proportional hazards model; this bias is reduced somewhat with the shared frailty model. Even when fitting the correlated frailty model, this bias persists, although we were able to recover an estimate of b that is much closer to the true value. Ripatti and Palmgren (30) also noticed this phenomenon and attributed it to the approximation of the likelihood used in the estimation. In addition, they observed an underestimation in the variance component of the correlated frailty model, a result that occurs in our simulations as well. The results show that, on average, the estimated SD of $\widehat{\beta}$ is smaller in the proportional hazards and the shared frailty models than in the correlated frailty model. This result is consistent with the fact that related observations contribute less

information in relation to a model that assumes independence; hence the standard errors of the fixed parameters are underestimated. Moreover, Ripatti and Palmgren (30) observed that the estimated SD for $\widehat{\beta}$, in the correlated frailty model, is underestimated because the variation in the estimation of the variance component is ignored (i.e., the estimator of the SD of $\widehat{\beta}$ is calculated as if σ were known).

The correlated frailty model assumes that the fixed covariates are independent of the frailty distribution; however, in many applied settings this will not be true. Table 4 includes the results of simulations in which the fixed covariate is correlated with frailty. The results are similar to those obtained when the fixed covariate and frailty are assumed to be independent. The new result is that, in some cases, the parameter β has a positive bias.

These results point to the importance of using correlated frailty models for the analysis of family-based data. To emphasize even more the undesirable consequences of ignoring the relationship among individuals and assuming independence, in our simulations we observed that none of the 95% CI values for β , obtained from the proportional hazards fit, contains the true value of β . This finding is due to the large bias and also possibly to the underestimation of the SD of $\widehat{\beta}$.

Discussion

A greater understanding of factors affecting human longevity requires analytic tools that permit the joint assessment of observable environmental and genetic covariates and an accounting of the role played by unobservable factors. The recent availability of genetic markers along with measures of social and environmental risk factors on large population-based cohort samples makes it possible to assess their distinct influence on mortality risks using standard survival analysis techniques. The simultaneous development of complex longitudinal databases of extended kindreds, such as the UPDB, means that we now have informative data on variables affecting longevity on extended families and large kindreds. The challenge is to identify methods that are optimally suited for the analysis of these data. This study has shown how the standard Cox proportional hazards model can be extended to incorporate unobservable factors through the introduction of random effects or frailty components by using different assumptions regarding the role of shared family and correlated genetic influences using family-based data. The performance of these models was assessed through data on persons from LL families and controls, identified from the UPDB, as well as through simulations.

The present study supports the long-established observation that longevity is subject to substantial familial aggregation. It also demonstrates that the family-specific risk of mortality, controlling for observed covariates, is typically 21% larger or smaller than the overall mortality risk. This level of variability can be attributed to environmental factors, and would likely decline as additional significant covariates were added to the model. We also show that the relative risk of mortality at the individual level that is attributable to polygenic factors is likely to be as much as 40% larger or smaller than the average risk. When we fit a model that allows for both the contributions of polygenic and shared familial traits simultaneously, we find that the variance component due to genetic factors is comparable to the one due to shared environment.

The application of the shared frailty and the correlated polygenic frailty models reported here relied on assumptions about the manner in which the environment and genetic factors are shared among relatives. Specifically, we assumed that all individuals in a family share the same environment throughout their lives and that those environmental factors that are

not shared or that change with time are nonexistent. This is a strong assumption, but alternative correlation structures that allow for more complex environmental conditions are possible with the methods described here. The method permits the analyst to modify the a_{ij} values to reflect assumed or measured levels of exposure-sharing between, for example, coresident relatives as opposed to kin who live in disparate geographic areas.

Our simulation analyses illustrate the consequences of assuming independence of survival times between individuals when family-based data are analyzed. Models that ignore the environmental or genetic association among related individuals result in downwardly biased relative risk estimates, a pattern that has been reported by others (38,39). We show that, in the presence of correlated frailty based on a general polygenic model, the correlated frailty specification nearly eliminates the bias in regression parameters. These simulations also indicate that the standard errors for the regression parameter estimates are also underestimated when the dependence among individuals is ignored. Our simulations show, for example, that none of the 95% CI values for the estimated regression parameters included the true value. Overall, downwardly biased regression coefficients coupled with narrow (and incorrect) CI values will lead to poor inferences about the significance of the covariates.

For analysts with survival data comprising statistically independent individuals, estimating the effects of the covariates on the hazard rate does not require any modifications to the methodology. If, however, the sample is based on a set of relatives (i.e., all individuals are connected to other individuals because they share a common ancestor), then it is important to consider modeling unobserved environmental and genetic factors within a general analytic approach.

Acknowledgments

This study was supported by National Institutes of Health Grant AG022095 (The Utah Study of Fertility, Longevity, and Aging).

We thank the Pedigree and Population Resource (funded by the Huntsman Cancer Foundation) for providing the data and valuable computing support. We also thank Dr. Terry Therneau at the Mayo Clinic College of Medicine for sharing his ideas and software used in this analysis. In addition, we acknowledge the contributions of Alison Fraser and Diana Lane Reed in providing database management and research assistance for this study.

REFERENCES

1. Abbott MH, Abbey H, Bolling DR, Murphy EA. The familial component in longevity—a study of offspring of nonagenarians: III. Intrafamilial studies. *Am J Med Genet.* 1978; 2:105–120. [PubMed: 263431]
2. Beeton M, Pearson K. Data for the problem of evolution in man. II. A first study on the inheritance of longevity and the selective death rate in man. *Proc R Soc Lond.* 1899; 65:290–305.
3. Beeton M, Pearson K. On the inheritance of the duration of life, and on the intensity of natural selection in man. *Biometrika.* 1901; 1:50–89.
4. Carmelli D, Andersen S. A longevity study of twins in the Mormon genealogy. *Prog Clin Biol Res.* 1981; 69(pt C):187–200. [PubMed: 7198241]
5. Cohen BH. Family patterns of Mortality and Life Span. *Q Rev Biol.* 1964; 39:130–181. [PubMed: 14175085]
6. Gavrilov LA, Gavrilova NS, Olshansky SJ, Carnes BA. Genealogical data and the biodemography of human longevity. *Soc Biol.* 2002; 49:160–173. [PubMed: 14652915]
7. Gudmundsson H, Gudbjartsson DF, Frigge M, Gulcher JR, Stefansson K. Inheritance of human longevity in Iceland. *Eur J Hum Genet.* 2000; 8:743–749. [PubMed: 11039573]
8. Houde, L.; Tremblay, M.; Vézina, H. Intergenerational and genealogical approaches for the study of longevity in the Saguenay – Lac-St-Jean population.. Inherited characteristics in populations of the

- past: Exploring intergenerational dimensions of human behaviour; Menorca, Spain. May 19–21, 2005;
9. Iachine IA, Holm NV, Harris JR, et al. How heritable is individual susceptibility to death? The results of an analysis of survival data on Danish, Swedish and Finnish twins. *Twin Res.* 1998; 1:196–205. [PubMed: 10100811]
 10. Matthijs K, Van de Putte B, Vlietinck R. The inheritance of longevity in a Flemish village (18th–20th century). *Eur J Popul.* 2002; 18:59–81.
 11. Mitchell BD, Hsueh WC, King TM, et al. Heritability of life span in the Old Order Amish. *Am J Med Genet.* 2001; 102:346–352. [PubMed: 11503162]
 12. Pearl R. Studies on human longevity IV. The inheritance of longevity: preliminary report. *Hum Biol.* 1931; 3:245–269.
 13. Philippe P. Familial correlations of longevity: an isolate-based study. *Am J Med Genet.* 1978; 2:121–129. [PubMed: 263432]
 14. Sorensen TI. Genetic epidemiology utilizing the adoption method: studies of obesity and of premature death in adults. *Scand J Soc Med.* 1991; 19:14–19. [PubMed: 1925421]
 15. Sorensen TI, Nielsen GG, Andersen PK, Teasdale TW. Genetic and environmental influences on premature death in adult adoptees. *N Engl J Med.* 1988; 318:727–732. [PubMed: 3347221]
 16. Swedlund AC, Meindl RS, Nydon J, Gradie M. Family patterns in longevity and longevity patterns of the family. *Hum Biol.* 1983; 55:115–129. [PubMed: 6601609]
 17. Vaupel JW. Inherited frailty and longevity. *Demography.* 1988; 25:277–287. [PubMed: 3396752]
 18. Williams GC. Pleiotropy, natural selection and the evolution of senescence. *Evolution.* 1957; 11:398–411.
 19. Wyshak G. Fertility and longevity in twins, sibs, and parents of twins. *Soc Biol.* 1978; 25:315–330. [PubMed: 574990]
 20. Cournil A, Legay JM, Schachter F. Evidence of sex-linked effects on the inheritance of human longevity: a population-based study in the Valserine valley (French Jura), 18–20th centuries. *Proc Biol Sci.* 2000; 267:1021–1025. [PubMed: 10874752]
 21. Gavrilov, LA.; Gavrilova, NS. Human longevity and paternal age at conception.. In: Robine, JM.; Kirkwood, TBL.; Allard, M., editors. *Sex and Longevity: Sexuality, Gender, Reproduction, Parenthood.* Springer-Verlag; Berlin: 2000. p. 7-31.
 22. McGue M, Vaupel JW, Holm N, Harvald B. Longevity is moderately heritable in a sample of Danish twins born 1870–1880. *J Gerontol.* 1993; 48:B237–B244. [PubMed: 8227991]
 23. Herskind AM, McGue M, Holm NV, Sorensen TI, Harvald B, Vaupel JW. The heritability of human longevity: a population-based study of 2872 Danish twin pairs born 1870–1900. *Hum Genet.* 1996; 97:319–323. [PubMed: 8786073]
 24. Ljungquist B, Berg S, Lanke J, McClearn GE, Pedersen NL. The effect of genetic factors for longevity: a comparison of identical and fraternal twins in the Swedish Twin Registry. *J Gerontol Med Sci.* 1998; 53A:M441–M446.
 25. Guo G. Use of sibling data to estimate family mortality effects in Guatemala. *Demography.* 1993; 30:15–32. [PubMed: 8440396]
 26. Guo G, Rodriguez G. Estimating a multivariate proportional hazard model for clustered data using the EM algorithm, with and application to child survival in Guatemala. *J Am Stat Assoc.* 1992; 87:969–976.
 27. Hougaard P, Harvald B, Holm NV. Measuring the similarities between the lifetimes of adult Danish twins born between 1881–1930. *J Am Stat Assoc.* 1992; 87:17–24.
 28. Yashin AI, Iachine IA. Genetic analysis of durations: correlated frailty model applied to survival of Danish twins. *Genet Epidemiol.* 1995; 12:529–538. [PubMed: 8557185]
 29. Yashin AI, Vaupel JW, Iachine IA. Correlated individual frailty: an advantageous approach to survival analysis of bivariate data. *Math Popul Stud.* 1995; 5:145–159, 183. [PubMed: 12290053]
 30. Ripatti S, Palmgren J. Estimation of multivariate frailty models using penalized partial likelihood. *Biometrics.* 2000; 56:1016–1022. [PubMed: 11129456]
 31. Therneau TM, Grambsch PM, Pankratz VS. Penalized survival models and frailty. *J Comput Graph Stat.* 2003; 12:156–175.

32. Pankratz VS, de Andrade M, Therneau TM. Random-effects Cox proportional hazards model: general variance components methods for time-to-event data. *Genet Epidemiol.* 2005; 28:97–109. [PubMed: 15532036]
33. Wei LJ, Lin DY, Weissfeld L. Regression analysis of multivariate incomplete failure time data by modeling marginal distributions. *J Am Stat Assoc.* 1989; 84:1065–1073.
34. Kerber RA, O'Brien E, Smith KR, Cawthon RM. Familial excess longevity in Utah genealogies. *J Gerontol Biol Sci.* 2001; 56A:B130–B139.
35. Mineau GP, Smith KR, Bean LL. Adult mortality risks and religious affiliation: the role of social milieu in biodemographic studies. *Annales de Demographie Historique.* 2004; 2:85–104.
36. Gavrilov, LA.; Gavrilova, NS. *The Biology of Life Span: A Quantitative Approach.* Harwood; New York: 1991.
37. Juckett DA, Rosenberg B. Comparison of the Gompertz and Weibull functions as descriptors for human mortality distributions and their intersections. *Mech Ageing Dev.* 1993; 69:1–31. [PubMed: 8377524]
38. Heckman, J.; Singer, B. Social science duration analysis.. In: Heckman, J.; Singer, B., editors. *Longitudinal Analysis of Labor Market Data.* Cambridge University Press; Cambridge, U. K.: 1985. p. 39-110.
39. Horowitz JL. Semiparametric estimation of a proportional hazard model with unobserved heterogeneity. *Econometrica.* 1999; 67:1001–1028.

Table 1

Descriptive Statistics

Variable	Members of LL Kindreds			Members of Control Kindreds				
	Mean	SD	Minimum	Maximum	Mean	SD	Minimum	Maximum
Birth year	1879	24.92	1748	1953	1876	27.09	1763	1953
Age at death	79.00	12.38	50	107	74.76	11.08	50	105
FEL	5.35	1.83	-2.27	14.55	2.77	1.84	-3.82	18.03

Note: LL = Long-lived; FEL = Familial Excess Longevity; SD = standard deviation.

Table 2

Parameter Estimates and 95% Confidence Intervals for Four Frailty Models

Characteristic	Proportional Hazard	Shared Frailty	Correlated Frailty	Shared & Correlated Frailty
Birth year	0.989 (0.987, 0.989)	0.988 (0.987, 0.989)	0.987 (0.986, 0.988)	0.988 (0.987, 0.989)
Gender (Male = 1)	1.234 (1.176, 1.296)	1.253 (1.193, 1.317)	1.269 (1.204, 1.337)	1.259 (1.197, 1.323)
LDS (Yes = 1)	1.034 (0.983, 1.088)	1.011 (0.959, 1.066)	1.005 (0.950, 1.063)	1.008 (0.955, 1.064)
LL kindred = 1, Control kindred = 0	0.858 (0.807, 0.912)	0.810 (0.745, 0.880)	0.819 (0.759, 0.884)	0.806 (0.740, 0.877)
Familial Excess Longevity	0.938 (0.926, 0.951)	0.951 (0.936, 0.966)	0.943 (0.928, 0.958)	0.951 (0.936, 0.967)
σ_f^2		0.036 (0.023, 0.053)		0.035 (0.022, 0.053)
σ_p^2			0.114 (0.067, 0.175)	0.020 (0.000, 0.079)

Note: LDS = Latter-day Saints; LL = long lived.

Table 3Estimates of β , σ_f^2 , and σ_p^2 When the Fixed Covariates Are Uncorrelated With Frailty

Parameter	Model		
	Proportional Hazards	Shared Frailty	Correlated Frailty
$\beta = 2.0, \sigma_p^2 = 0.5$			
β	1.600 (0.052)	1.707 (0.053)	1.930 (0.076)
$SD(\hat{\beta})$	0.044 (1.2e-4)	0.048 (1.2e-4)	0.054 (1.9e-4)
σ_f^2	–	0.119 (0.034)	–
σ_p^2	–	–	0.412 (0.094)
$\beta = 2.0, \sigma_p^2 = 1.0$			
β	1.378 (0.050)	1.521 (0.051)	1.846 (0.077)
$SD(\hat{\beta})$	0.040 (1.0e-4)	0.044 (1.1e-4)	0.053 (1.9e-4)
σ_f^2	–	0.189 (0.046)	–
σ_p^2	–	–	0.712 (0.129)
$\beta = 0.7, \sigma_p^2 = 0.5$			
β	0.556 (0.032)	0.593 (0.034)	0.669 (0.042)
$SD(\hat{\beta})$	0.029 (5.9e-5)	0.031 (6.1e-5)	0.035 (9.5e-5)
σ_f^2	–	0.110 (0.036)	–
σ_p^2	–	–	0.390 (0.105)
$\beta = 0.7, \sigma_p^2 = 1.0$			
β	0.477 (0.031)	0.527 (0.033)	0.634 (0.043)
$SD(\hat{\beta})$	0.028 (5.5e-5)	0.030 (5.7e-5)	0.036 (1.0e-4)
σ_f^2	–	0.173 (0.047)	–
σ_p^2	–	–	0.652 (0.137)
$\beta = 0.1, \sigma_p^2 = 0.5$			
β	0.079 (0.028)	0.084 (0.030)	0.094 (0.033)
$SD(\hat{\beta})$	0.028 (6.2e-5)	0.029 (6.3e-5)	0.032 (9.6e-5)
σ_f^2	–	0.103 (0.039)	–
σ_f^2	–	–	0.363 (0.120)
$\beta = 0.1, \sigma_f^2 = 1.0$			
β	0.068 (0.027)	0.075 (0.029)	0.089 (0.034)
$SD(\hat{\beta})$	0.027 (5.6e-5)	0.028 (5.9e-5)	0.034 (9.8e-5)
σ_f^2	–	0.165 (0.050)	–
σ_p^2	–	–	0.604 (0.148)

Note: Empirical expected values and standard deviations (SD) of parameter estimates were based on 1000 simulations.

Table 4Estimates of β , σ_f^2 , and σ_p^2 When the Fixed Covariates Are Correlated With Frailty

Parameter	Model		
	Proportional Hazards	Shared Frailty	Correlated Frailty
$\beta = 2.0, \sigma_p^2 = 0.5$			
β	1.699 (0.054)	1.812 (0.056)	2.041 (0.080)
$SD(\hat{\beta})$	0.046 (1.3e-4)	0.050 (1.4e-4)	0.056 (2.2e-4)
σ_f^2	–	0.119 (0.035)	–
σ_p^2	–	–	0.403 (0.100)
$\beta = 2.0, \sigma_p^2 = 1.0$			
β	1.463 (0.053)	1.615 (0.055)	1.956 (0.083)
$SD(\hat{\beta})$	0.041 (1.1e-4)	0.046 (1.2e-4)	0.055 (2.2e-4)
σ_f^2	–	0.190 (0.048)	–
σ_p^2	–	–	0.707 (0.134)
$\beta = 0.7, \sigma_p^2 = 0.5$			
β	0.646 (0.033)	0.689 (0.034)	0.776 (0.044)
$SD(\hat{\beta})$	0.030 (6.1e-5)	0.032 (6.4e-5)	0.036 (1.1e-4)
σ_f^2	–	0.112 (0.037)	–
σ_p^2	–	–	0.391 (0.113)
$\beta = 0.7, \sigma_p^2 = 1.0$			
β	0.554 (0.033)	0.612 (0.035)	0.738 (0.046)
$SD(\hat{\beta})$	0.028 (5.7e-5)	0.031 (6.0e-5)	0.037 (1.1e-4)
σ_f^2	–	0.178 (0.049)	–
σ_p^2	–	–	0.666 (0.145)
$\beta = 0.1, \sigma_p^2 = 0.5$			
β	0.163 (0.028)	0.174 (0.030)	0.195 (0.035)
$SD(\hat{\beta})$	0.028 (6.1e-5)	0.029 (6.6e-5)	0.033 (1.1e-4)
σ_f^2	–	0.104 (0.040)	–
σ_p^2	–	–	0.370 (0.129)
$\beta = 0.1, \sigma_p^2 = 1.0$			
β	0.141 (0.028)	0.155 (0.030)	0.185 (0.035)
$SD(\hat{\beta})$	0.027 (5.9e-5)	0.029 (6.4e-5)	0.034 (1.1e-4)
σ_f^2	–	0.168 (0.051)	–
σ_p^2	–	–	0.615 (0.157)

Note: Empirical expected values and standard deviations (SD) of parameter estimates were based on 1000 simulations.