HERITABILITY OF LONGEVITY AND THE ROLE OF EARLY AND MID-LIFE ENVIRONMENTS

Heidi A. Hanson
Ken R. Smith
Saundra Hasstedt
University of Utah

We wish to thank the Pedigree and Population Resource of the Huntsman Cancer Institute, University of Utah for providing the data and valuable computing support. This work was also supported by NIH grant AG022095 (Early Life Conditions, Survival and Health; Smith PI).

Introduction

Mortality is the quintessential measure of the health of a population, and large gains in human life expectancy over the past 150 years are evidence of the important relationship between social context and health. While there is certainly variability in longevity between populations, there is also wide variation in longevity within populations. Understanding the determinants of this heterogeneity is essential to understanding the processes of aging and health of a population. But how much of this variation is determined by genetic factors and how much is determined by the environment? While the question of heritability of longevity is not new, with heritability estimates of longevity ranging from 0 to 0.3 (Kerber, O'Brien, Smith, & Cawthon, 2001), we seek to determine if heritability estimates vary between subpopulations and explore the possibility of gene-environment interactions (GxE). By examining sources of variation in heritability estimates, we can illuminate factors that modify the expression of genetic predisposition in a population.

We will investigate heterogeneity in the genetic basis of longevity by assessing the phenotypic correlation between relatives. Variance components and heritability values will be generated using a large genealogical database with information on family structure as well as measures of the broader environment. This study will examine the relationship between social context and the amount of additive genetic variance in adult life-span and exceptional longevity using data from the Utah Population Database (UPDB), a rich source of linked population-based information for demographic, genetic, and epidemiological studies. The sample used in this study consists of 20,120 individuals

from 802 three generation pedigrees. This analysis has two goals: 1) estimate the heritability of longevity after age 30 as well as exceptional longevity in a population using methods designed for use in multigenerational pedigree information; 2) test for differences in heritability estimates of life-span in populations stratified by environmental exposure.

Background

Heritability of Longevity

Over the past few decades, demographers have broadened the focus of work in the demography of aging from a population aging perspective (i.e. measures of change in population age structure) to include a perspective that integrates health and biological explanations with traditional demographic and social theories of aging to explain heterogeneity in health and mortality within and between populations (Olshansky, Carnes, & Brody, 2002; Siegel, 2011; Vasunilashorn & Crimmins, 2008). While it is widely accepted that life-span is determined by a combination of genetic, social and physical environment, and stochastic factors, the interdependent and dynamic role of genes and environment is still not well understood. This may be partially due to fears of genetic determinism within the field of sociology (Shostak & Freese, 2010), the divergent paths of genetics and demography (Adams, 1990), and the difficulty of assessing the role of genes and environment biomarker data.

Longevity is a complex trait, determined by a multiplicity of genetic and environmental factors, each of which contributes to a potentially small amount to phenotypic variation. The genetic variation that is the natural background is a shortened

life-span (relative to exceptional longevity) and exceptional longevity is the result of a mutations. Genes affecting longevity have been parsed into two categories described as gerontogenes: genes that have a negative effect on longevity and longevity-assurance genes that promote longevity (Christensen, Johnson, & Vaupel, 2006). Findings from the New England Centenarian Study (NECS) have suggested that supercentenarians do not lack gerontogenes, but have longevity assurance genes that can counter the deleterious effects of genes and environment as well as slow the rate of aging and lead to delayed onset of age-related disease (Sebastiani et al., 2012). It is also believed that longevity mutations increase the ability to handle stress and robustness (Christensen et al., 2006). The proportion of variation in life-span due to genes is moderate, which can be illustrated by the fact that there is variation in life-span between monozygotic twins (Herskind et al., 1996).

In summary, longevity is determined by a complex relationship between both genes and environment. For a more complete understanding of population heterogeneity in life-span and the forces behind it, one must not only understand the average contribution of genes and environment within a population toward explaining variation in adult mortality, but uncover the factors that influence patterns of variation within the population.

At the most basic level, phenotypic variation can be partitioned into additive genetic variance and general environmental variance. Additive genetic variance is the deviation from the average phenotype, or trait, in the population that is due to the inheritance of a particular allele and that allele's effect on the phenotype. General environmental variance can then be described as the remaining variance that cannot be

attributed to genes. The proportion of variation due to inheritance of a particular allele is not fixed across all environments because the relationship between genotype and phenotype may vary by environment, a phenomenon known as phenotypic plasticity. Narrow sense heritability is a population level statistic that describes the amount of total phenotypic variation (V_T) that can be attributed to additive genetic variation (V_A) in the population ($h^2 = V_A/V_T$). The polygenic model can be used to partition variation into genetic and residual environmental effects. Because it is a population level statistic, it is important to keep in mind that it not a property of individual traits. When h^2equals zero, it indicates that all phenotypic variation within a population can be explained by individual differences, while an h^2 of one indicates that all the phenotypic variation is explained by genetic differences. This is not to say that high heritability suggests little environmental effect on the phenotype. When h^2 is elevated the environment may uniformly contribute to the expression of the trait and therefore contribute little to *differences* between people. It has been shown that heritability of traits can vary across subpopulations (Jason D. Boardman, 2009; Jason D Boardman et al., 2012; D. C. Rowe, Almeida, & Jacobson, 1999). But how much of the population heterogeneity in life-span is determined by genetic factors?

There is evidence of the presence of familial clustering of longevity over many generations and across diverse populations, suggesting that there is a genetic or familial component to successful aging and longevity (Christensen et al., 2006; C. E. Finch & Tanzi, 1997; Herskind et al., 1996; Kerber et al., 2001; T. Perls, Kunkel, & Puca, 2002). The longevity literature has described the genetic and environmental contribution to mortality as being divided into one-third and two-third proportions, respectively (C. E.

Finch & Tanzi, 1997; Siegel, 2011). It has also been suggested that 50% of the variation in life-span after age 30 can be ascribed to attributes (genetic and non-genetic) that are fixed prior to that age (Yashin & Iachine, 1997), and that genetics plays a stronger role with advancing age (Hjelmborg et al., 2006; Montesanto, Dato, Bellizzi, Rose, & Passarino, 2012; Vaupel et al., 1998) (see J. W. Rowe and Kahn (1997) for a dissent from this view). If the proportion of variation in life-span that can be explained by genetic factors varies by age, is it also conditioned by social context? And if so, does this conditioning vary by age?

Conceptualizing the Relationship between Social Environment
And Heritability of Longevity

Attempting to understand the genetic component of longevity without considering how it may be modified by specific environmental factors may not be a fruitful approach to gaining insights into the heritability of this complex trait (Petronis, 2010). The heritability of certain phenotypes may vary throughout the life course (Turkheimer, Haley, Waldron, D'Onofrio, & Gottesman, 2003) and by gender (Visscher, Hill, & Wray, 2008). Given that humans are constantly interacting with the environment and the environment has the ability to alter gene expression, we must also understand how environmental influences might modify the heritability of longevity. Accordingly, in this analysis, comparisons of heritability will be made between three subgroups of the population based on: 1) religious involvement; 2) early disease and nutritional environment; and 3) family environment during childhood.

This study assumes that the same genes affect longevity across environments within a population, but certain attributes of the environment serve to moderate the effect of genes on phenotypic variation. Shanahan and Hofer (2005) have presented a framework for gene and social context interactions that has been used to explain the relationship between the social environment and health behaviors (Jason D. Boardman, 2009; Jason D Boardman et al., 2012). We present a slightly modified version that also utilizes concepts presented by Hoffmann and Merilä (1999) as well as new modifications to help formulate our hypotheses.

Under Shanahan's framework for gene-environment (GxE) interactions, the environment is conceptualized as social context (Shanahan & Hofer, 2005). Four perspectives, described in detail below, can be used to depict how the social environment might affect heritable variation; triggering, compensation, social control, and enhancement. Figure 4.1 shows a modified version of a schema presented by Sebastiani et al. (2012) describing the genetic components of aging. Sebastiani has hypothesized that individuals living to exceptional ages have gerontogenes, but the longevity assurance genes counter the deleterious effects of genetic and environmental factors. Panel A shows the proportion of total phenotypic variance (V_T) that is attributable to additive genetic variance (V_A) and environmental variance (V_E), where the phenotype is longevity after age 30. We show that in a normal environment where there is no GxE interaction (panel A), individuals with shorter life-spans have higher heritability of gerontogenes and individuals with exceptional life-spans have higher heritability of longevity assurance genes.

A triggering effect refers to an environment that interacts with personal predispositions to a diseased state and shortened life-span through, for example, environmental stressors or other factors that induce a biological change. Panel B in figure 4.1 shows the hypothesized triggering GxE interaction in which an adverse environment directly affects the phenotype. When triggering mechanisms are responsible for environmental differences in heritability, we expect to see a decrease in average longevity in adverse environments and an increase in additive genetic variance. This is because the environment leads to phenotypic expression that would otherwise be dormant. This relationship may change with exceptional longevity because selection mechanisms may change the heritability of a trait over time. If the genetically frail individuals are selected out of this population at an earlier age, the surviving population may be comprised of more robust individuals with a genetic predisposition for exceptional longevity (i.e. longevity assurance genes) (Hawkes, Smith, & Blevins, 2012), leading to higher levels of heritability of exceptional longevity in environments detrimental to health. Therefore, under this formulation, we expect that individuals exposed to an unhealthy environment during childhood will have higher heritability of a shorter life-span compared to those living in more advantageous circumstances. This may also translate into higher heritability of exceptional longevity because only the robust in an unhealthy environment survive to exceptional ages.

The second type of GxE interaction is compensation. According to this perspective, in normal and adverse environments the predisposition to a diseased state and shortened life-span is realized but not in enriched settings. The expected change in additive genetic variance in an enriched environment is presented in panel C of figure

4.1. The compensation GxE perspective assumes that the continuous exposure to a healthy environment prevents the expression of a genetic diathesis that predisposes an individual to premature death. Unlike the triggering mechanism, the relationship between environment and phenotype is not causal, but due to environmental variation. Therefore, we would expect to see an increase in average life expectancy in an enriched environment with lower additive genetic variance for the longevity phenotype.

Social control is the third GxE model. This interaction is not presented in figure 1 because the expected outcomes are similar to those presented in panel C. Heritability of longevity may be attenuated in environments with high social control because social norms and structural constraints place limits on choices, and, therefore the environment suppresses phenotypic variance. This is similar to the evolutionary argument of canalization, which argues that selection favors suppression of quantitative traits in constant and structured environments, but the genotype maintains a potential for expressing certain phenotypes under particular environmental conditions (Hoffmann & Merilä, 1999). Thus, involvement with a religious institution that maintains strong social norms for health related behaviors such as alcohol consumption, smoking, social support, and dietary restrictions may lead to increased longevity and exceptional longevity for all members of the group and suppress genetic predispositions for disease. In this situation, we expect to see increased longevity and exceptional longevity for active religious participants with lower levels of heritability compared to non-participants.

The enhancement model of GxE is presented in panel D of figure 1. This is similar to the social control mechanism, but rather than suppressing a predisposition to a shortened life-span, social context can serve to enhance genetic predispositions for

longevity. Individuals in advantaged and organized social settings may be more apt to realize their genetic potential for longevity, while disadvantaged environments lead to unrealized potential. For example, an environment of undernutrition or high levels of exposure to infectious agents may lead to physiological changes that alter an individual's ability to reach their genetic potential (D. Barker, 1995; D. J. P. Barker et al., 1993; Eileen M. Crimmins & Finch, 2006). Here, we would expect to see mean differences in survival between environments and higher heritability of life-span and exceptional longevity in environments more advantageous for health and longevity.

In this paper, we build on a body of literature that examines the heritability of longevity by comparing heritability of longevity and exceptional longevity between sub-populations exposed to different environments that are known to affect adult mortality risks. Using the GxE perspectives discussed above, we compare the heterogeneity of genetic effects by environment. We expect to see differences in the heritability of longevity between environments characterized as salubrious or unhealthy. The GxE categories of triggering, compensation, and social control predict higher levels of heritability of longevity in environments less beneficial to health, while the enhancement typology predicts increased heritability of longevity in healthy environments. We can make generalizations about what type of GxE interaction leads to the observed patterns, but the exact mechanism is not testable under this formulation. Comparing the components of variance between environments will add to the understanding of the relative importance of both genes and environment in determining longevity.

Methods

Data

The majority of life-span epidemiological studies examine health influences of early and adult life conditions with relatively modest sample sizes, particularly given the complexity of the phenomena and the manifold exposures and outcomes. This study utilizes data drawn from the Utah Population Database (UPDB). The UPDB is one of the world's richest sources of linked population-based information for demographic, genetic, and epidemiological studies. UPDB has supported biodemographic studies as well numerous important epidemiological and genetic studies in large part because of its size, pedigree complexity, and linkages to numerous data sources. In the mid-1970s, over 185,000 three-generation families were identified on "Family Group Sheets" from the archives at the Utah Family History Library. These families have been linked into multigenerational families and the full UPDB now contains data on nearly 7 million individuals due to longstanding and on-going efforts to add new sources of data and update records as they become available.

Mortality data are fundamental to the study of exceptional longevity. Information on deaths prior to 1904 comes from genealogical records obtained from the Utah Family History Library and linked to other records within the UPDB. All Utah death certificates are available from 1904 to the present. The UPDB also links to the U.S. Social Security Death Index (SSDI) for the years 1936 – 2011. The SSDI records provide information on deaths based on Social Security records regardless of place of death and are linked to the UPDB. The unique combination of genealogy, death certificates, and SSDI data provide wide spatial and temporal coverage for both the fact and date of death.

The sample used to construct measures of longevity comprises all individuals in the UPDB born between 1850 and 1927. We selected 1927 as the maximum birth year to

allow us to observe mortality to at least age 85 for the youngest members of the cohort. To minimize variability in survival unrelated to aging and based on other evidence of the fixed attributes related to life-span after age 30 (Yashin & Iachine, 1997), we will model mortality beginning at age 30 (Hawkes et al., 2012). We identified 685,949 individuals who met the criteria listed above. Of those, approximately 9% (n=64,258) were right censored and 91% (N=621,961) had vital status follow-up information from family history group sheets, Utah death certificates, or linked Social Security Death Index (SSDI) information. The gender distribution of the sample was 52.5% male and 47.5% female.

Using individuals from the baseline survival analysis, we selected 111,324 three-generation families. Table 4.1 shows the restrictions imposed at the family level. We attempted to select families with the highest data quality and the most complete information. As a result, 31,322 families were excluded from the analysis because at least one grandparent had no information in the UPDB. All founding pedigree members were required to have a birth year greater than 1850 (Utah was settled in 1847) and all members of the family were required to be born before prior to 1928, which allowed us to observe the youngest members of the cohort to age 84. On average, these families had four individuals in the first generation (by definition), 13 individuals in the second generation, and 19 individuals in the third generation (range = 1 to 83). Pedigree size ranged from 7 to 109 members. The final sample consisted of 802 three-generation families with 20,120 members with a calculated longevity measure and information on family of origin. To study exceptional longevity, a nearly deceased cohort is needed.

Therefore, individuals born between 1914 and 1927 were excluded from the exceptional longevity sample, yielding a sample of 14,618 individuals for these analyses.

Measuring Early and Mid-Life Environments

Both early and mid-life conditions will be considered as possible social context that may modify phenotypic expression. To simplify both the measurement and conceptualization of the environment, each environment is treated as a simple dichotomy, comparing salubrious environments to those that are less advantageous to health. We will compare the heritability of longevity by religious participation, infant mortality rate (IMR) in the family of origin, childhood mortality rate (CMR) of the family of origin, and number of siblings. Justifications for each as the basis for deleterious and beneficial environments are described in turn below.

Religious involvement in general is associated with increased life expectancy (Hummer, Rogers, Nam, & Ellison, 1999). It is not surprising that active affiliation with the Church of Jesus Christ of Latter-day Saints (LDS or Mormon) church is also associated with increased life expectancy (Enstrom & Breslow, 2008). Individuals actively affiliated with the LDS church are more likely to abstain from alcohol and tobacco use, fast once a month, and participate in church related social activities (Mineau, Smith, & Bean, 2002). Therefore, affiliation with the LDS church will be treated as a social environment with defined healthy lifestyle norms for men and women. The UPDB contains information on baptism dates from family history records, which were used to classify individuals as followers of the LDS church. Individuals baptized as members of the LDS church before the age of 30 are considered followers of the LDS Church.

Individuals will be parsed into two environments: 1) LDS church involvement; and 2) no LDS church involvement.

Early life health can have long-term consequences on later life health and mortality (Elo & Preston, 1992; Smith, Mineau, Garibotti, & Kerber, 2009). While it is difficult to obtain a measure of early life exposure to disease and other adverse circumstances, we can use mortality outcomes of siblings as a sentinel for early life circumstances. Post-neonatal mortality (the first year of life excluding the first 28 days) for our cohorts of study (1850 – 1927) is closely related to viral and bacterial disease, malnutrition, and income (B. K. Finch, 2003; McKeown, 1976; Preston & Haines, 1991). A similar argument has been made for childhood mortality by Eileen M. Crimmins and Finch (2006), who argue that birth cohorts with lower childhood mortality have increased longevity. As such, we use the death of a sibling during the first year of life (IMR in family of origin) or between ages one and five (CMR in family of origin) as indicators of an adverse childhood environment. Neonatal deaths, deaths within the first 28 days, and stillbirths are not included in our final measure of IMR because these deaths are likely due to endogenous causes and may not represent a family environment marked by disease, an assumption most likely to be true for the years considered here. We consider infant and childhood mortality as distinct environments because it has been suggested that the determinants of infant and childhood mortality decline over time differed during this period (Wolleswinkel-van den Bosch, van Poppel, Looman, & Mackenbach, 2000). Individuals with one or more infant or childhood death in their family of origin (i.e. death of a sibling) were considered to be in an environment of high infant or childhood mortality respectively.

Sibship size (number of siblings) has been shown to be positively associated with lower educational achievement and unhealthy lifestyle choices (Downey, 1995; Hart & Davey Smith, 2003). Sibship size may also be related to exposure to infectious diseases, with children from large sibships having a greater risk of contracting an infectious disease (Hart & Davey Smith, 2003). However, a strong association between sibship size and adult mortality has not been demonstrated in all studies assessing this relationship (Smith et al., 2009). The definition of large sibship was derived empirically as having 7 or more siblings (75^{th} percentile for the sample).

While sex is inherently a biological trait, sex differences in life expectancy are determined by both social and biological factors (E.M. Crimmins & Saito, 2001; Lindahl-Jacobsen et al., 2013; Rieker, Bird, & Lang, 2010). Sex differences may exist in the effects of early life conditions on later life health. Male fetuses have higher mortality rates than female fetuses, a disadvantage that continues throughout the life course (Kraemer, 2000). Earlier studies have found slight differences in the heritability of longevity between males and females, with males having higher heritability than females (Herskind et al., 1996). Accordingly, we test for environmental differences in the heritability of longevity by sex.

Definition of Longevity

The mortality schedule for individuals born between 1850 and 1927 has changed considerably. Longevity, therefore, defined simply as years lived after age 30, is not appropriate because it is not directly comparable across birth cohorts. While much of the improvement in life expectancy seen during this period was due to improvements in

infant and child mortality, there were also gains in adult mortality. Cohort life tables for Utah show that individuals born in 1850 and surviving to age 30 could expect to live an additional 40.5 and 39.5 years for females and males respectively, compared to 53.1 and 48.5 years of additional life after 30 for individuals born in 1920 (Lindahl-Jacobsen et al., 2013). Therefore, to de-trend the data, we define longevity as the difference between an individual's attained age (y) and the age to which that individual was expected to live (median predicted age of death conditioned on surviving to 30, $\hat{y}$) according to a model that incorporates two basic determinants of life-span: gender and birth year. Therefore, a longevity score (LS) is simply the difference between these two values, $y - \hat{y}$. The baseline survival models used to determine $\hat{y}$ are described below. This approach is similar to one taken by Kerber et al. (2001) in calculating a measure of familial excess longevity using Utah genealogies.

Previous studies have suggested that the heritability of longevity increases with age (Hjelmborg et al., 2006; Yashin & Iachine, 1997), and may perhaps be the strongest for those surviving to the latest ages (Atzmon et al., 2004; Gudmundsson, Gudbjartsson, Frigge, Gulcher, & Stefánsson, 2000; T. T. Perls, Bubrick, Wager, Vijg, & Kruglyak, 1998). Exceptional longevity can be defined as an exceptionally long life-span compared to other individuals experiencing the same historical influences (birth cohort) (Michael Anson et al., 2012). As done in previous studies, we will define the exceptional longevity (EL) phenotype as living to exceptional age, and explore differences in heritability of survival to the 90^{th} and 95^{th} percentile based on the baseline hazard models (Kerber, O'Brien, Boucher, Smith, & Cawthon, 2012).

Constructing Baseline Survival Models

We assume a parametric form for the survival distribution and a generalized class of accelerated failure time (AFT) models, the extended family of generalized Gamma models. Unlike proportional hazard models, AFT models assume that the effect of covariates is multiplicative with respect to survival time. We test the fit of the exponential, Weibull , log-normal, log-logistic, and gamma models. These models were selected because they provide a simple point estimate for duration that generally fits the observed data for adult mortality. While the Gompertz model is appropriate for modeling human mortality between 30 and 85 (Olshansky and Carnes, 1997), this study is concerned with exceptional longevity (past the age of 85) and therefore this model was not considered. The nested structure of the family of generalized Gamma models (exponential, Weibull, log-logistic, and gamma) allows for use of the likelihood ratio test to assess model fit. The Akaike information criterion, or AIC, can be used to test the fit of non-nested models. Final models were selected for the construction of the longevity measures based on model fit.

The full sample of 685,949 individuals born between 1850 and 1927 meeting the sample criteria described in the data section above were used to estimate survival time for individuals surviving to age 30. Models were stratified by gender and included two covariates, birth year, and birth year squared. All models showed a significant positive relationship between birth year and survival. The generalized Gamma model proved to be the best fit based on likelihood ratio test ($p<0.001$) and AIC. The shape and scale parameters in the generalized gamma model are also significantly different than "0" and "1", implying that the fitted distribution is different from the Weibull, log-normal, and

exponential models. The change in the predicted 50th and 90th percentile by model and year is displayed in figure 4.2. Panels A and B show the trend in predicted median and 90th percentile longevity respectively for men and panels C and D display the estimates for women. The exponential model does not fit the data well and therefore the results are not shown. These figures show that the log-normal and log-logistic models, which provided the worst fit based on the AIC statistic, also predict out of range values for the 90th percentile. Both the Weibull and generalized gamma model provide sensible estimates. Therefore, the generalized gamma model was used to estimate $\hat{y}$ and consequently LS.

LS was defined as the observed minus the expected life-span for all deceased individuals. The UPDB contains multiple sources of linked records that can be used to create a last observed date. Therefore, we know that individuals without a death record were alive until their last observed date in UPDB. The observed life-span for individuals born after 1905, not deceased, and with a known follow-up date that exceeded the median predicted survival time for their gender and birth cohort is calculated by subtracting the birth year from the last observed date in UPBD. Therefore, censored individuals that are likely still living were used in the LS analyses and have a positive LS score by definition. To test for biased results created by this specification, we ran sensitivity analysis using the nearly deceased cohort (N=14,618).

Heritability Estimates

Several forms of analysis of variance (ANOVA) are available to measure heritability of a phenotypic trait, such as parent-offspring regressions and sibling

analyses. While these models have useful features, they are limited because they do not use information from multigenerational relationships and they require that sample sizes be well-balanced. Unlike other forms of analysis of variance (ANOVA), maximum likelihood (ML) estimators do not place special demands on the design or balance of the data, providing a powerful approach to estimating variance components using large pedigrees (Lynch & Walsh, 1998) and minimizing the inflation of estimates of additive genetic variance due to shared environments between relatives. To allow for use of information on multigenerational relationships, heritability is estimated with a polygenic model using PAP v. 7.1 (Hasstedt, 2005).

Genotypic variance can be decomposed into additive (V_A), dominance (V_D), and epistatic (V_I). V_D and V_I are, however, extremely difficult to estimate in non-experimental settings (Kruuk, 2004). The polygenic model specifies the expected genetic relationship between relatives as a function of the coefficient of relationship, allowing for the estimation of variation due to genetic and residual environmental effects. The coefficient of relationship is $(1/2)^p$, where p is the degree of relationships (it is also commonly described as two times the probability that two individuals will share a common gene by descent (IBD)). For example, for a parent-child relationship the coefficient of relationship is 0.5, which equals 2 x 0.25, where 0.25 is the probability that parent and child share a common allele. The polygenic model allows us to partition the total phenotypic variance (σ_T^2) into the following components:

$$\sigma_T^2 = \sigma_A^2 + \sigma_E^2 \qquad \text{(eq. 4.1)}$$

where σ_A^2 is the additive genetic variance and σ_E^2 is the residual variance, which includes environmental, dominance, and epistatic effects. These components are used to calculate heritability, with narrow sense heritability (h^2) being defined as the proportion of phenotypic variance, σ_T^2, that can be attributed to the additive genetic effects, σ_A^2:

$$h^2=\sigma_A^2/\sigma_T^2 \qquad \text{(eq. 4.2)}$$

The polygenic model is similar to a mixed model with fixed and random effects. The general model in matrix form is:

$$y = \boldsymbol{X}\beta + \boldsymbol{Z}u + e \qquad \text{(eq.4.3)}$$

where *y* is a column vector containing the phenotypic values for a trait measured in *n* individuals; β is a vector of fixed effects; *u* is a vector of random effects; **X** and **Z** are known incidence matrices; and *e* is a column vector of random residual effects. We assume that *u* follows a multivariate normal (MVN) distribution with mean zero and variance **G**, and that *e* also follows a MVN distribution with mean zero and variance R. Note that **G** = σ_A^2A, where A is an *n x n* matrix of kinship coefficients describing the genetic correlation between all individuals in the sample, and R=σ_E^2I, where *I* is the identity matrix. This general model can be used to estimate the variance components for a single trait (univariate model) and has been extended to allow for joint modeling of multiple traits (bivariate for two traits and multivariate for multiple traits). The univariate model was used to estimate the heritability of LS and exceptional longevity in the population. We then use the multivariate model to estimate h^2 by environment.

The multivariate model provides a means for estimating covariance and, therefore, correlation between traits. Falconer (1952) suggested that traits measured in two environments can be treated as two different traits. This allows comparisons

between discrete environments of different types, where the bivariate model defines a trait as being expressed in environment one or two. For example, if individual *i* is in environment one, they have a value for trait one and are missing trait two. Conversely, if individual *j* is in environment two, they are missing trait one and have a value for trait two. This approach is more appropriate than stratification, because it allows for the joint estimation of heritability in two subpopulations. It is also slightly different than a normal bivariate trait model, which jointly models two phenotypes measured on the same individual because no individual expresses a trait in both environments. In this approach, *k* traits (in our case *k*=2) are combined to form a vector $Z = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = (y_{11}, \ldots, y_{1n,\ \ldots,} y_{2n})$ with mean μ_z and variance **G**. The model in matrix notation is,

$$\mu_z = \begin{bmatrix} X_1 & 0 \\ 0 & X_2 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix} + \begin{bmatrix} W_1 & 0 \\ 0 & W_2 \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \qquad \text{(eq. 4.4)}$$

where y_1 and y_2 are vectors of phenotypic values in environment one and two, respectively; β_1 and β_2 are the vectors of the fixed effects in environment one and two, respectively; a_1 and a_2 are the vectors of the random additive genetic effects in environment one and two, respectively; e_1 and e_2 are the vectors of random residual effects for environment one and two, respectively; $\mathbf{X}_1$ and $\mathbf{X}_2$ are the known incidence matrices relating the observations to the respective fixed effects in environments one and two; and $\mathbf{W}_1$ and $\mathbf{W}_2$ relate the observations to the random effects in environments one and two.

The variance-covariance matrix for **Z** can be expressed as $\mathbf{V} = \mathbf{G} + \mathbf{R} = \boldsymbol{C} \otimes \boldsymbol{A} + \boldsymbol{E} \otimes \boldsymbol{I}_n$, where G is the Kronecker product of C and A ($C \otimes A$). **C** is the *k* x *k* matrix of additive genetic covariances, and **E** is the *k* x *k* residual covariance matrix. **A** and **I** are

respectively the *n x n* kinship coefficient and identity matrices, with $c_{ij}=\sigma_A(i,j)$ being the additive genetic covariance between characters *i* and *j* within an individual and cross-covariance $c_{ij}A_{lm}$ being the additive genetic value of character *i* in individual *l* and the additive genetic value of character *j* in individual *m* (Lynch & Walsh, 1998, p. 777). In a bivariate analysis, C is a 2 x 2 matrix of the form:

$$C = \begin{bmatrix} \sigma_A^2(1) & \sigma_A(1,2) \\ \sigma_A(1,2) & \sigma_A^2(2) \end{bmatrix} \qquad \text{(eq. 4.5)}$$

where $\sigma_A^2(1)$ and $\sigma_A^2(2)$ are the additive genetic variances for traits 1 and 2, respectively, and $\sigma_A(1,2)$ is the additive genetic cross covariance.

Defining the environment at the individual level and estimating heritability using the multivariate model without defining the genetic correlation between traits leads to biased estimates of heritability because heritability estimates from an environment only include information about family members in the same environment. To correct for this problem, we assume perfect genetic correlation between the trait values. A bivariate analysis that explicitly models genetic correlations exploits more information content of the data (Amos, de Andrade, & Zhu, 2001). The genetic correlation between traits can be defined as:

$$\rho_{x12} = \frac{\sigma_A(1,2)}{\sqrt{\sigma_{A1}^2 \sigma_{A2}^2}} \qquad \text{(eq. 4.6)}$$

where $\sigma_A(1,2), \sigma_{A1}^2$, and σ_{A2}^2 are all components of variance mentioned above and $\sigma_A(1,2) = \rho_{x12}\sigma_{A1}\sigma_{A2}$. By constraining ρ_{x12} to one, we are requiring the covariance between traits to equal the square-root of the product of the variances and forcing the model to include information from both environments. Constraining the genetic correlation to unity allows for heritability and additive genetic variance to vary in both

environments, but requires them to be dependent. In a bivariate trait analysis, where both phenotypes are measured for an individual, the genetic correlation is often estimated and used to describe the pleiotropic nature of the traits. However, estimating the genetic correlation across environments would be erroneous in our situation because when $\rho_x < 1$, we are only using partial information from the pedigree because the covariance is weighted by the correlation coefficient ($\rho_{x12}\sigma_{A1}\sigma_{A2}$). Algebraically, this solves the bias problem because it forces the measure of additive genetic variance for each environment to include information about family members from both environments. It is also conceptually plausible because a genetic correlation of one indicates the effect of the same polygenes on the trait in both environments.

LS was Box-Cox transformed and standardized (μ=0, σ=1) to improve computational performance and abide by distributional assumptions of the variance components models. The transformation was performed using Proc transreg in SAS, which uses a maximum likelihood approach to find the optimal transformation, which in this cases was λ=1.75. This transformation reduced the skewness coefficient from -0.85 to -0.26. The simple correlation between the transformed variable and the original measure of LS was 0.98.

To test the hypothesis of heterogeneity in heritability, the likelihood ratio statistic was used. Models were estimated, allowing heterogeneity in heritability estimates between environments, and compared to models where the heritability estimates were constrained to be equal across environments. Sex and birth year were not considered as covariates in the model because they were controlled for when creating the measures of longevity.

Results

Descriptive Statistics

Figure 4.3 shows the distribution of LS for the baseline survival cohort and the sample selected for the heritability estimates. Both distributions are slightly skewed with means of -1.2 and -1.7 for the full cohort and the heritability cohort, respectively. The skewed distribution reflects the change in the rate of mortality between ages 30 and the median predicted survival time for an individual's sex and cohort. Cohort life tables for Utah show that the q_x for females at age 30 in the 1900 birth cohort is 0.02, compared to 0.05 at age 60 and 0.29 at age 80 (Lindahl-Jacobsen et al., 2013). Therefore, it is not unexpected to see the long left tail in the LS distribution. The distributional skew is due to a combination of factors including model fit (the fit provides a good approximation of the survival curve, but does not fit the data exactly) and censoring of the youngest cohort.

Table 4.2 shows the descriptive statistics for individuals in the heritability samples. The longevity sample includes all 20,120 individuals with calculated longevity, LS, from the 802 selected pedigrees. Individuals in the exceptional longevity sample were required to be born before 1914 so that we could observe survival in the UPDB to age 99. Approximately 8% of males and females in this sample survive to the 90th percentile for their cohort and sex. This number is slightly smaller than 10% because the cut point for the 90th percentile is derived from the baseline survival models. 48% of the sample is female and approximately three-fourths of the sample was affiliated with the LDS church. All members of a family with a sibling that died during infancy or childhood are counted as having an infant death in their family of origin and in historical cohorts. Children from large families experience excess infant and childhood mortality

rates (Bean, Mineau, & Anderton, 1990; Knodel & Hermalin, 1984), so this percentage is slightly higher than the 17.4% and 18.5% percent of nuclear families with an infant or childhood death, respectively. There is not a substantial amount of overlap in these measures, with 6.4% of nuclear families having both an infant and childhood death.

Figure 4.4 shows the effect of environment on LS without considering family structure. Significant differences in LS exist in all environments. Panel A shows the distribution of LS by religious status, with individuals not affiliated with the LDS church on average having a two point reduction in longevity score ($p<0.001$). The distribution of LS by infant and childhood mortality in family of origin is displayed in Panels B and C respectively, with individuals having one or more sibling die during the post-neonatal period having a two point reduction in LS ($p<0.001$), and individuals with one or more sibling deaths during childhood having a 1.5 reduction in LS ($p<0.001$). Panel D shows the distribution of LS by sibship size and illustrates the nearly two-point reduction of LS for individuals having seven or more siblings.

Heritability estimates

The overall heritability of LS in the sample is 0.18, which is within the range of previously reported estimates. We find that in the four environments considered when not conditioned on sex, the mean LS is lower in unhealthy environments but there are no significant differences in h^2. The pattern of heritability of LS by environment is somewhat mixed, with higher heritability of LS in environments with low IMR and CMR, but lower heritability of LS in the other two healthy environments. It is important to note that heritability is a population statistic, thus we are comparing subpopulations

defined by an environment and not average individual differences in phenotype. The addition of environment-specific means and variances significantly improve model fit for all environments, with lower means and environmental variances in environments that are considered beneficial to longevity.

To further investigate sex differences in heritability and GxE interactions, we considered models separately by sex. In a bivariate model, considering only sex differences, we find that heritability of LS is significantly lower for females compared to males, h^2_{LSf}= 0.14 and h^2_{LSm}=0.22 (LR χ^2= 9.03, p=0.003), and there is little difference between the mean and environmental variances by gender. The lack of difference in the mean LS is by design. LS was constructed as a gender specific measure (i.e., the baseline survival models were stratified by sex), and therefore one would not expect to see gender differences in the average LS.

Multivariate models were used to calculate the heritability estimates for LS by environment and sex (results in Table 4.3). When considering the differences in heritability of LS by sex and environment, the mean differences in LS by environment are similar, with lower mean LS in environments considered unhealthy. We find no significant differences in the heritability of LS by environment with the exception of female environments classified by CMR, which show a nine point difference in h^2_{LS} between the healthful and unhealthy environments. The heritability of LS is lower in female environments with high CMR when compared to female environments with low sibling CMR (LR χ^2=5.88, *p*=0.015). This is in contrast to the higher heritability of LS clustered about a lower mean LS in the male environment with high CMR compared to an environment with low CMR, although these differences are not significant. For

females, there is little difference in total phenotypic variance between the two CMR subpopulations (σ_T^2 is approximately 1.30 and 1.29 in high CMR and low CMR subpopulations respectively). This is supportive of the enhancement hypothesis, which suggests that individuals are unable to realize there genetic potential in adverse environments.

Sensitivity analyses using the nearly deceased cohort (n=14,618) were run for the LS models. We found that heritability estimates were slightly smaller (0.17 vs. 0.18 in the larger sample), but the observed differences by gender and environment were all in the same direction. The differences in heritability by CMR environment remained significant.

We considered defining EL as survival to the 90th or 95th percentile conditioned on birth year and sex. The sample for these analyses is smaller than the sample used to obtain estimates of heritability of LS because observing EL requires a nearly extinct cohort (N_{LS}=20,120, N_{EL}= 14,618). Heritability estimates for the two phenotypes were very similar, with h^2_{EL}=0.352 when EL is defined as survival to the 90th percentile (shown in Table 4), and h^2_{EL}=0.345 (95% CI= 0.244, 0.447) when EL is defined as survival to the 95th percentile (results not shown). The small decline in heritability between the 90th and 95th percentiles suggests that heritability does not increase linearly with age, and that perhaps there is an upper limit to increases in heritability of longevity. However, the differences are negligible and not relevant to the main hypotheses of this paper. Therefore, we show results for survival to the 90th percentile conditioned on age and sex.

Table 4.4 shows the heritability estimates for EL by environment and gender. We find that heritability for EL is nearly twice the heritability of LS (0.18 vs. 0.35). Bivariate models were used to test for environmental differences in the heritability of EL. We find that allowing the prevalence to vary by environment significantly improves model fit, with higher prevalence of EL in healthful environments. We find no difference in heritability of EL by environment when not conditioned on gender. There are also no gender differences in the heritability of EL (LR χ^2= 0.552, p=0.46), which differs from the LS findings.

When using the multivariate model to test for environmental differences in heritability of EL by gender and environment, we do not find evidence of significant differences in with the exception of the male CMR environment. The heritability of EL is 31 points higher in male environments with high CMR compared to male environments with low CMR (LR χ^2=4.25, p=0.04), and there is no difference in the prevalence of exceptional longevity between environments. This suggests that a triggering GxE interaction may be operating through selection mechanisms, where the frail are selected out of the adverse populations at faster rates and only the genetically robust individuals with longevity assurance genes that are able to thwart the effects of gerontogenes survive to exceptional ages.

Discussion

Our analysis of longevity is based on information from 20,120 individuals from 802 three-generation families used to examine the heritability of longevity, defined as survival after age 30. We also estimated the heritability of exceptional longevity using

information from a subset of that sample (n=14,618) that is nearly extinct. Our findings support previous studies suggesting a moderate heritable component to longevity that increases with age (Herskind et al., 1996; Hjelmborg et al., 2006; Kerber et al., 2001), although the adult ages at which this assessment is made varies across analyses. We find little difference between the heritability of survival to the 90^{th} and 95^{th} percentiles, suggesting that the increase in proportion of variance due to genetic factors may not be a constant linear increase as suggested by other studies (Hjelmborg et al., 2006). We find that sex differences in the heritability of longevity after age 30 support other studies showing higher heritability of longevity for males (Herskind et al., 1996), but no sex differences in the heritability of exceptional longevity. We investigated the heterogeneity of longevity and exceptional longevity by early and mid-life social environments. We find some evidence that the heritability of longevity varies between environments, but overall there is not strong support of a gene-environment interaction for the selected environments.

We find evidence that childhood environments marked by high child mortality, indicative of exposure to infectious disease and undernutrition for the surviving members, may affect the proportion of phenotypic variation attributable to genetic factors. The sex and age differences of the effects suggest an enhancement GxE interaction because adverse childhood circumstances limit the genetic potential of individuals to survive to older ages. Conceptually, CMR is used to identify environments with excess exposure to infectious disease and undernutrition. For females, genetic factors contribute little to the total variance in longevity in such environments, which suggests that genetic potential is

not reached in such environments. While a similar pattern exists for EL, the difference in heritability between environments is not significant.

We see the opposite effect for male environments, although the observed patterns do not necessarily conflict with the female results. Males have a mortality disadvantage relative to females throughout the life course that is partially due to biological factors (Kraemer, 2000). Therefore, they may be more susceptible to environmental conditions that trigger genetic predispositions for disease and lead to higher mortality selection compared to females reared in the same environment. The difference in the direction of the effect for males suggests that the adverse environment may actually trigger genetic diatheses, with higher heritability clustered about a lower mean longevity in deleterious environments, but these differences are not significant. This results in higher heritability of exceptional longevity because individuals surviving to this age have some predisposition or genetic robustness that prevented them from being selected out of the population at earlier ages.

It is interesting that we find heterogeneity in CMR environments, but not in environments characterized by IMR. This may be partially due to differences in specific causes of death for the two groups, as suggested by Wolleswinkel-van den Bosch et al. (2000). We did consider variations of our definition of IMR, which included neonatal deaths, although this did not change the substantive conclusion that heritability of longevity does not vary between subpopulations with different rates of infant mortality.

While we find some evidence of heterogeneity in the heritability of longevity between environments, heritability estimates seem to be fairly impervious to early and mid-life circumstances. Herskind et al. (1996) reported stability of heritability estimates

over sex and cohort during periods of rapid change in living conditions. However, the birth cohorts selected for that study would still be children during periods with higher childhood mortality (1870 – 1900) than experienced during modern times. Our results suggest that improvements in social and health conditions that have caused declines in childhood mortality may lead to a higher proportion of variability in longevity attributable to genetic factors. More research needs to be done to test for other environmental differences in the heritability of longevity, including socioeconomic status and fertility history.

The nearly twofold increase in heritability of exceptional longevity compared to the heritability of longevity after age 30 suggests that selection mechanisms may affect the heritability of longevity throughout the life course. Individuals without longevity assurance genes may be selected out of the population at early ages, leaving a subset of the population that is made up of a higher proportion of robust individuals. While the heritability of longevity increases with age, exceptional longevity is still only moderately heritable, and the environment explains the largest amount of phenotypic variation. It is also remarkable that there are gender differences in heritability of longevity after age 30, but not with respect to exceptional survival. This is suggestive that individuals surviving to exceptional ages have survived mortality selection because they have a genetic variant that increases the ability to handle stress and/or counteract deleterious effects of the environment or generontogenes. This is further supported by other research suggesting the buffering role of longevity genes (Bergman, Atzmon, Ye, MacCarthy, & Barzilai, 2007; Sebastiani et al., 2012).

Epigenetics is one of several possible biological mechanisms that allow social circumstances to get "under the skin", and it recently has been suggested that epigenetic changes have the propensity to persist across subsequent generations (Feinberg, 2007). This is a provocative idea that lends support to mutligenerational transmission of social disparities. More research needs to be done to uncover the possible mechanisms leading to phenotypic variation across social environments and the possibility of transmitting the adverse effects to subsequent generations. We suggest further study into the possibility of GxE interactions and health and longevity outcomes. While we did not find an association between all environments, there is a suggestion that the social environment may play an important role in modifying the heritability of longevity.

In this paper, we assessed variation in heritability estimates of longevity after age 30. However, other cutoffs, such as post-reproductive aging, should also be considered. **This will be done before the PAA 2014 conference.** Further modeling of heterogeneity of variance and the variance of longevity across other environments could be valuable in understanding how the social environment moderates the genetic component to aging. **Socioeconomic status and adult fertility will be two environments that will also be considered for the PAA 2014 conference. Our preliminary work suggests that heritability of longevity varies by SES.**

Care should be taken when interpreting polygenic heritability when the genetic correlation has been fixed to unity, because it is assumed that the same genes affect longevity across environments. While we feel this is a valid assumption for subgroups of a single population, and the reader should be aware of this constraint.

To our knowledge, this is the first study using multigenerational pedigree information to investigate heterogeneity in heritability of longevity across multiple early and mid-life environments. Studies in other fields have examined heterogeneity in variance components by gender and age using a similar method (Giolo, Pereira, de Andrade, Krieger, & Soler, 2010; Pilia et al., 2006), lending validity to this approach.

.

References

Adams, J. (1990). *Convergent issues in genetics and demography*: Oxford University Press, USA.

Amos, C. I., de Andrade, M., & Zhu, D. K. (2001). Comparison of Multivariate Tests for Genetic Linkage. *Human Heredity, 51*(3), 133-144.

Atzmon, G., Schechter, C., Greiner, W., Davidson, D., Rennert, G., & Barzilai, N. (2004). Clinical Phenotype of Families with Longevity. *Journal of the American Geriatrics Society, 52*(2), 274-277. doi: 10.1111/j.1532-5415.2004.52068.x

Barker, D. (1995). Fetal origins of coronary heart disease. *BMJ, 311*(6998), 171-174.

Barker, D. J. P., Godfrey, K. M., Gluckman, P. D., Harding, J. E., Owens, J. A., & Robinson, J. S. (1993). Fetal nutrition and cardiovascular disease in adult life. *The Lancet, 341*(8850), 938-941. doi: 10.1016/0140-6736(93)91224-a

Bean, L. L., Mineau, G. P., & Anderton, D. L. (1990). *Fertility change on the American frontier: Adaptation and innovation* (Vol. 4): Univ of California Press.

Bergman, A., Atzmon, G., Ye, K., MacCarthy, T., & Barzilai, N. (2007). Buffering Mechanisms in Aging: A Systems Approach Toward Uncovering the Genetic Component of Aging. *PLoS Comput Biol, 3*(8), e170. doi: 10.1371/journal.pcbi.0030170

Boardman, J. D. (2009). State-Level Moderation of Genetic Tendencies to Smoke. *American Journal of Public Health, 99*(3), 480-486. doi: 10.2105/AJPH.2008.134932

Boardman, J. D., Roettger, M. E., Domingue, B. W., McQueen, M. B., Haberstick, B. C., & Harris, K. M. (2012). Gene–environment interactions related to body mass: School policies and social context as environmental moderators. *Journal of Theoretical Politics, 24*(3), 370-388. doi: 10.1177/0951629812437751

Christensen, K., Johnson, T. E., & Vaupel, J. W. (2006). The quest for genetic determinants of human longevity: challenges and insights. *Nat Rev Genet, 7*(6), 436-448.

Crimmins, E. M., & Finch, C. E. (2006). Infection, inflammation, height, and longevity. *Proceedings of the National Academy of Sciences of the United States of America, 103*(2), 498-503. doi: 10.1073/pnas.0501470103

Crimmins, E. M., & Saito, Y. (2001). Trends in healthy life expectancy in the United States, 1970-1990: gender, racial, and educational differences. *Social Science and Medicine, 52*(11), 1629-1642.

Downey, D. B. (1995). When Bigger Is Not Better: Family Size, Parental Resources, and Children's Educational Performance. *American Sociological Review, 60*(5), 746-761. doi: 10.2307/2096320

Elo, I. T., & Preston, S. H. (1992). Effects of Early-Life Conditions on Adult Mortality: A Review. *Population Index, 58*(2), 186-212. doi: 10.2307/3644718

Enstrom, J. E., & Breslow, L. (2008). Lifestyle and reduced mortality among active California Mormons, 1980–2004. *Preventive Medicine, 46*(2), 133-136. doi: 10.1016/j.ypmed.2007.07.030

Falconer, D. (1952). The problem of environment and selection. *American Naturalist*, 293-298.

Feinberg, A. P. (2007). Phenotypic plasticity and the epigenetics of human disease. *Nature, 447*(7143), 433-440.

Finch, B. K. (2003). Early origins of the gradient: the relationship between socioeconomic status and infant mortality in the United States. *Demography, 40*(4), 675-699. doi: 10.2307/1515203

Finch, C. E., & Tanzi, R. E. (1997). Genetics of aging. *Science, 278*(5337), 407-411.

Giolo, S., Pereira, A., de Andrade, M., Krieger, J., & Soler, J. (2010). Evaluating gene by sex and age interactions on cardiovascular risk factors in Brazilian families. *BMC Medical Genetics, 11*(1), 132.

Gudmundsson, H., Gudbjartsson, D. F., Frigge, M., Gulcher, J. R., & Stefánsson, K. (2000). Inheritance of human longevity in Iceland. *European journal of human genetics : EJHG, 8*(10), 743-749.

Hart, C. L., & Davey Smith, G. (2003). Relation between number of siblings and adult mortality and stroke risk: 25 year follow up of men in the Collaborative study. *Journal of Epidemiology and Community Health, 57*(5), 385-391. doi: 10.1136/jech.57.5.385

Hasstedt, S. (2005). jPAP: document-driven software for genetic analysis. *Genet Epidemiol, 29*(3), 255.

Hawkes, K., Smith, K. R., & Blevins, J. K. (2012). Human actuarial aging increases faster when background death rates are lower: a consequence of differential heterogeneity? *Evolution, 66*(1), s103-114.

Herskind, A., McGue, M., Holm, N., Sørensen, T., Harvald, B., & Vaupel, J. (1996). The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Human genetics, 97*(3), 319-323. doi: 10.1007/bf02185763

Hjelmborg, J. B., Iachine, I., Skytthe, A., Vaupel, J. W., McGue, M., Koskenvuo, M., Christensen, K. (2006). Genetic influence on human lifespan and longevity. *Human genetics, 119*(3), 312-321.

Hoffmann, A. A., & Merilä, J. (1999). Heritable variation and evolution under favourable and unfavourable conditions. *Trends in Ecology & Evolution, 14*(3), 96-101. doi: http://dx.doi.org/10.1016/S0169-5347(99)01595-5

Hummer, R., Rogers, R., Nam, C., & Ellison, C. (1999). Religious involvement and U.S. adult mortality. *Demography, 36*(2), 273-285. doi: 10.2307/2648114

Kerber, R. A., O'Brien, E., Boucher, K. M., Smith, K. R., & Cawthon, R. M. (2012). A Genome-Wide Study Replicates Linkage of 3p22-24 to Extreme Longevity in Humans and Identifies Possible Additional Loci. *PLoS ONE, 7*(4), e34746. doi: 10.1371/journal.pone.0034746

Kerber, R. A., O'Brien, E., Smith, K. R., & Cawthon, R. M. (2001). Familial Excess Longevity in Utah Genealogies. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences, 56*(3), B130-B139. doi: 10.1093/gerona/56.3.B130

Knodel, J., & Hermalin, A. I. (1984). Effects of birth rank, maternal age, birth interval, and sibship size on infant and child mortality: evidence from 18th and 19th century reproductive histories. *American Journal of Public Health, 74*(10), 1098-1106. doi: 10.2105/AJPH.74.10.1098

Kraemer, S. (2000). The fragile male. *Clinical Medicine NetPrints*, 1.

Kruuk, L. E. B. (2004). Estimating genetic parameters in natural populations using the 'animal model'. *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences, 359*(1446), 873-890. doi: 10.1098/rstb.2003.1437

Lindahl-Jacobsen, R., Hanson, H. A., Oksuzyan, A., Mineau, G. P., Christensen, K., & Smith, K. R. (2013). The male–female health-survival paradox and sex differences in cohort life expectancy in Utah, Denmark, and Sweden 1850–1910. *Annals of Epidemiology, 23*(4), 161-166. doi: http://dx.doi.org/10.1016/j.annepidem.2013.02.001

Lynch, M., & Walsh, B. (1998). Genetics and analysis of quantitative traits.

McKeown, T. (1976). *The modern rise of population* (Vol. 11): Edward Arnold London.

Mineau, G. P., Smith, K. R., & Bean, L. L. (2002). Historical trends of survival among widows and widowers. *Social Science & Medicine, 54*(2), 245-254. doi: 10.1016/s0277-9536(01)00024-7

Montesanto, A., Dato, S., Bellizzi, D., Rose, G., & Passarino, G. (2012). Epidemiological, genetic and epigenetic aspects of the research on healthy ageing and longevity. *Immun Ageing, 9*(1), 6.

Olshansky, S. J., Carnes, B. A., & Brody, J. (2002). A biodemographic interpretation of life span. *Population and Development Review, 28*(3), 501-513.

Perls, T., Kunkel, L. M., & Puca, A. A. (2002). The Genetics of Exceptional Human Longevity. *Journal of the American Geriatrics Society, 50*(2), 359-368. doi: 10.1046/j.1532-5415.2002.49283.x

Perls, T. T., Bubrick, E., Wager, C. G., Vijg, J., & Kruglyak, L. (1998). Siblings of centenarians live longer. *The Lancet, 351*(9115), 1560.

Petronis, A. (2010). Epigenetics as a unifying principle in the aetiology of complex traits and diseases. *Nature, 465*(7299), 721-727.

Pilia, G., Chen, W.-M., Scuteri, A., Orrú, M., Albai, G., Dei, M., . . . Schlessinger, D. (2006). Heritability of Cardiovascular and Personality Traits in 6,148 Sardinians. *PLoS Genet, 2*(8), e132. doi: 10.1371/journal.pgen.0020132

Preston, S. H., & Haines, M. R. (1991). *Fatal Years: Child Mortality in Late Nineteenth-Century America*: National Bureau of Economic Research, Inc.

Rieker, P. P., Bird, C., & Lang, M. (2010). Understanding gender and health: Old patterns, new trends and future directions. *Handbook of medical sociology, Nashville*.

Rowe, D. C., Almeida, D. M., & Jacobson, K. C. (1999). School Context and Genetic Influences on Aggression in Adolescence. *Psychological Science, 10*(3), 277-280. doi: 10.1111/1467-9280.00150

Rowe, J. W., & Kahn, R. L. (1997). Successful Aging. *The Gerontologist, 37*(4), 433-440. doi: 10.1093/geront/37.4.433

Sebastiani, P., Solovieff, N., DeWan, A. T., Walsh, K. M., Puca, A., Hartley, S. W., . . . Wilk, J. B. (2012). Genetic signatures of exceptional longevity in humans. *PLoS ONE, 7*(1), e29848.

Shanahan, M. J., & Hofer, S. M. (2005). Social Context in Gene–Environment Interactions: Retrospect and Prospect. *The Journals of Gerontology Series B: Psychological Sciences and Social Sciences, 60*(Special Issue 1), 65-76. doi: 10.1093/geronb/60.Special_Issue_1.65

Shostak, S., & Freese, J. (2010). Gene-Environment Interaction and Medical Sociology. *Handbook of Medical Sociology. Nashville: Vanderbilt UP*, 418-434.

Siegel, J. S. (2011). *The demography and epidemiology of human health and aging*: Springer Verlag.

Smith, K. R., Mineau, G. P., Garibotti, G., & Kerber, R. (2009). Effects of childhood and middle-adulthood family conditions on later-life mortality: Evidence from the Utah Population Database, 1850–2002. *Social Science & Medicine, 68*(9), 1649-1658. doi: 10.1016/j.socscimed.2009.02.010

Turkheimer, E., Haley, A., Waldron, M., D'Onofrio, B., & Gottesman, I. I. (2003). Socioeconomic Status Modifies Heritability of IQ in Young Children. *Psychological Science, 14*(6), 623-628. doi: 10.1046/j.0956-7976.2003.psci_1475.x

Vasunilashorn, S., & Crimmins, E. (2008). Biodemography: Integrating disciplines to explain aging. *Handbook of theories of aging*, 63-85.

Vaupel, J. W., Carey, J. R., Christensen, K., Johnson, T. E., Yashin, A. I., Holm, N. V., . . . Curtsinger, J. W. (1998). Biodemographic Trajectories of Longevity. *Science, 280*(5365), 855-860. doi: 10.1126/science.280.5365.855

Visscher, P. M., Hill, W. G., & Wray, N. R. (2008). Heritability in the genomics era [mdash] concepts and misconceptions. *Nat Rev Genet, 9*(4), 255-266.

Wolleswinkel-van den Bosch, J. H., van Poppel, F. W., Looman, C. W., & Mackenbach, J. P. (2000). Determinants of infant and early childhood mortality levels and their decline in The Netherlands in the late nineteenth century. *International journal of Epidemiology, 29*(6), 1031-1040. doi: 10.1093/ije/29.6.1031

Yashin, A., & Iachine, I. (1997). How frailty models can be used for evaluating longevity limits: Taking advantage of an interdisciplinary approach. *Demography, 34*(1), 31-48. doi: 10.2307/2061658

Table 4.1. Pedigree Selection

3 Generation Families from 1850 - 1927 Cohort*	**111,324**
Exclusions	
Families missing information on at least one grandparent (These people have placeholder genealogy records)	31,322
At least one of the grandparents is born before 1850	52,780
A member of G3 born after 1927	26,420
Total Number of 3 Generation Families for Analysis	**802**

*This is calculated by taking any member of the BC and ascending 3 generations. The result is 111,324 distinct treetops (defined by the unique combination of maternal and paternal grandparents). Families that did not meet our selection criteria were then excluded.

Table 4.2. Descriptive Statistics for Individuals from 802 Utah families

	Longevity (N=20,120)		Exceptional Longevity (N=14,618)	
	Male (N=10,393)	Female (N=9,727)	Male (N=7,532)	Female (N=7,086)
Birth Year	1897 (21.1)	1897 (20.3)	1889 (18.8)	1889 (18.0)
Longevity Score (LS)	-1.7 (15.2)	-1.9 (15.5)	-2.3 (15.4)	-2.7 (16.5)
Survived to the 90th Percentile (EL)			8.4%	8.0%
Survived to the 95th Percentile			4.2%	3.9%
Baptized Latter-Day Saint	72.5%	74.4%	75.6%	77.7%
One or more Post-Neonatal Infant Deaths in Family of Origin	19.5%	19.7%	23.9%	23.1%
One or more Childhood Deaths in Family of Origin	19.5%	19.9%	23.1%	24.2%
Large number of Siblings	29.6%	29.2%	35.6%	35.6%

Table 4.3. Summary of the Results Obtained for Polygenic Models of LS (N=20,120)

		Male and Female			Male			Female		
Environment		μ (95% CI)	h^2 (95% CI)	$\widehat{\sigma^2}_e$	μ (95% CI)	h^2 (95% CI)	$\widehat{\sigma^2}_e$	μ (95% CI)	h^2 (95% CI)	$\widehat{\sigma^2}_e$
	All	-0.10 (-0.12, -0.08)	0.18 (0.15,0.20)	1.10	-0.09 (-0.12,-0.07)	0.22 (0.18,0.26)	1.10	-0.09 (-0.11,-0.07)	0.14 (0.10,0.17)	1.11
LDS	Not Affiliated with LDS Church	-0.17 (-0.20, -0.14)	0.20 (0.14,0.25)	1.14	-0.21 (-0.25,-0.16)	0.27 (0.18,0.35)	1.15	-0.13 (-0.18,-0.09)	0.14 (0.07,0.21)	1.09
	Affiliated with LDS Church	-0.06 (-0.08, -0.03)	0.16 (0.13,0.19)	1.08	-0.04 (-0.07,-0.02)	0.19 (0.15,0.23)	1.07	-0.07 (-0.10,-0.04)	0.13 (0.09,0.17)	1.09
Infant Mortality Post Neonatal	High Infant Mortality	-0.16 (-0.20, -0.12)	0.16 (0.11,0.22)	1.13	-0.15 (-0.20,-0.10)	0.18 (0.10,0.26)	1.07	-0.17 (-0.23,-0.12)	0.15 (0.07,0.23)	1.19
	Low Infant Mortality	-0.07 (-0.10, -0.05)	0.18 (0.15,0.21)	1.10	-0.08 (-0.11,-0.05)	0.23 (0.19,0.27)	1.11	-0.07 (-0.10,-0.04)	0.13 (0.09,0.17)	1.08
Childhood Mortality	High Childhood Mortality	-0.19 (-0.23, -0.15)	0.15 (0.09,0.20)	1.17	-0.18 (-0.23,-0.13)	0.27 (0.17,0.37)	1.13	-0.20 (-0.25,-0.15)	0.07 (0.01,0.12)	1.21
	Low Childhood Mortality	-0.07 (-0.09, -0.05)	0.19 (0.16,0.22)	1.08	-0.07 (-0.10,-0.05)	0.21 (0.17,0.25)	1.09	-0.07 (-0.09,-0.04)	0.16 (0.12,0.20)	1.08
Sibship Size	Large Sibship	-0.21 (-0.25, -0.18)	0.19 (0.15,0.24)	1.17	-0.19 (-0.23,-0.15)	0.22 (0.16,0.29)	1.09	-0.24 (-0.29,-0.20)	0.16 (0.10,0.22)	1.25
	Small Sibship	-0.05 (-0.07, -0.03)	0.18 (0.15,0.21)	1.07	-0.06 (-0.09,-0.03)	0.22 (0.18,0.27)	1.10	-0.04 (-0.06,-0.01)	0.13 (0.09,0.17)	1.04

Table 4.4. Summary of the Results Obtained for Polygenic Models of EL

		Male and Female		Male		Female	
Environment		Prevalence (95% CI)	h^2 (95% CI)	Prevalence (95% CI)	h^2 (95% CI)	Prevalence (95% CI)	h^2 (95% CI)
All		0.08 (0.08, 0.09)	0.35 (0.28,0.43)	0.08 (0.08,0.09)	0.32 (0.21,0.43)	0.08 (0.07,0.09)	0.39 (0.27,0.51)
LDS	Non LDS	0.07 (0.06, 0.08)	0.43 (0.21,0.65)	0.06 (0.05,0.07)	0.55 (0.21,0.90)	0.07 (0.06,0.09)	0.28 (-0.01,0.58)
	LDS	0.09 (0.08, 0.09)	0.33 (0.25,0.42)	0.09 (0.08,0.10)	0.28 (0.16,0.39)	0.08 (0.07,0.09)	0.40 (0.26,0.55)
Infant Mortality Post Neonatal	High Infant Mortality	0.07 (0.06, 0.08)	0.28 (0.13,0.43)	0.07 (0.06,0.08)	0.28 (0.04,0.51)	0.08 (0.06,0.09)	0.28 (0.05,0.51)
	Low Infant Mortality	0.09 (0.08, 0.09)	0.37 (0.28,0.46)	0.09 (0.08,0.10)	0.32 (0.19,0.45)	0.08 (0.07,0.09)	0.43 (0.28,0.58)
Childhood Mortality	High Childhood Mortality	0.07 (0.06, 0.08)	0.44 (0.27,0.62)	0.08 (0.08,0.09)	0.57 (0.29,0.84)	0.07 (0.06,0.08)	0.33 (0.09,0.57)
	Low Childhood Mortality	0.09 (0.08, 0.09)	0.33 (0.24,0.42)	0.08 (0.08,0.09)	0.26 (0.14,0.38)	0.08 (0.08,0.09)	0.41 (0.27,0.56)
Sibship Size	Large Sibship	0.07 (0.07, 0.08)	0.33 (0.20,0.46)	0.07 (0.06,0.08)	0.30 (0.12,0.47)	0.08 (0.06,0.09)	0.37 (0.18,0.56)
	Small Sibship	0.09 (0.08, 0.09)	0.37 (0.27,0.47)	0.09 (0.08,0.10)	0.35 (0.19,0.50)	0.08 (0.07,0.09)	0.40 (0.23,0.57)

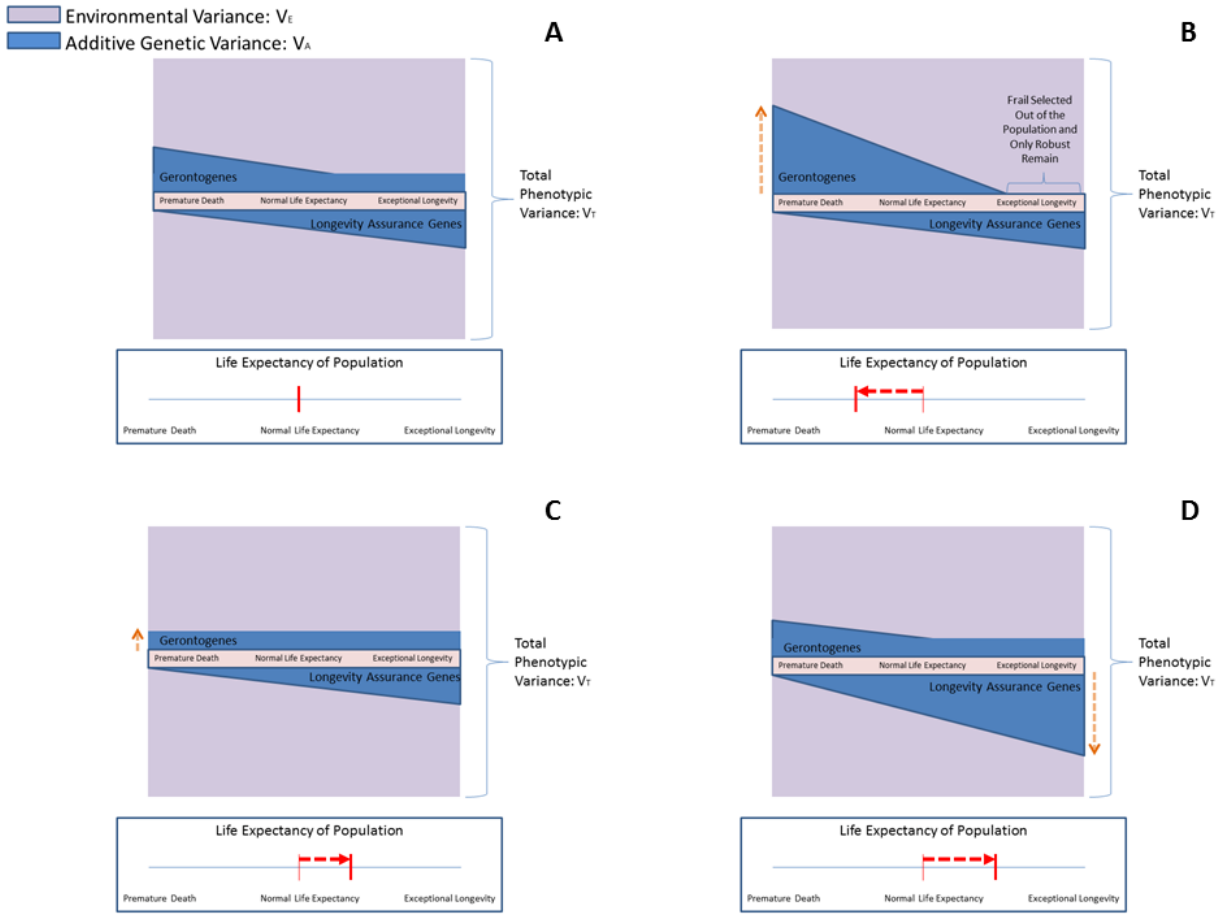

Figure 4.1. Hypotheses for GxE Interactions. Panel A shows the expected phenotypic variation in a normal environment. Panel B shows a triggering GxE interaction in an adverse environment. Panel C shows a compensation GxE interaction in an enriched environment. Panel D shows an enhancement interaction in an enriched environment.

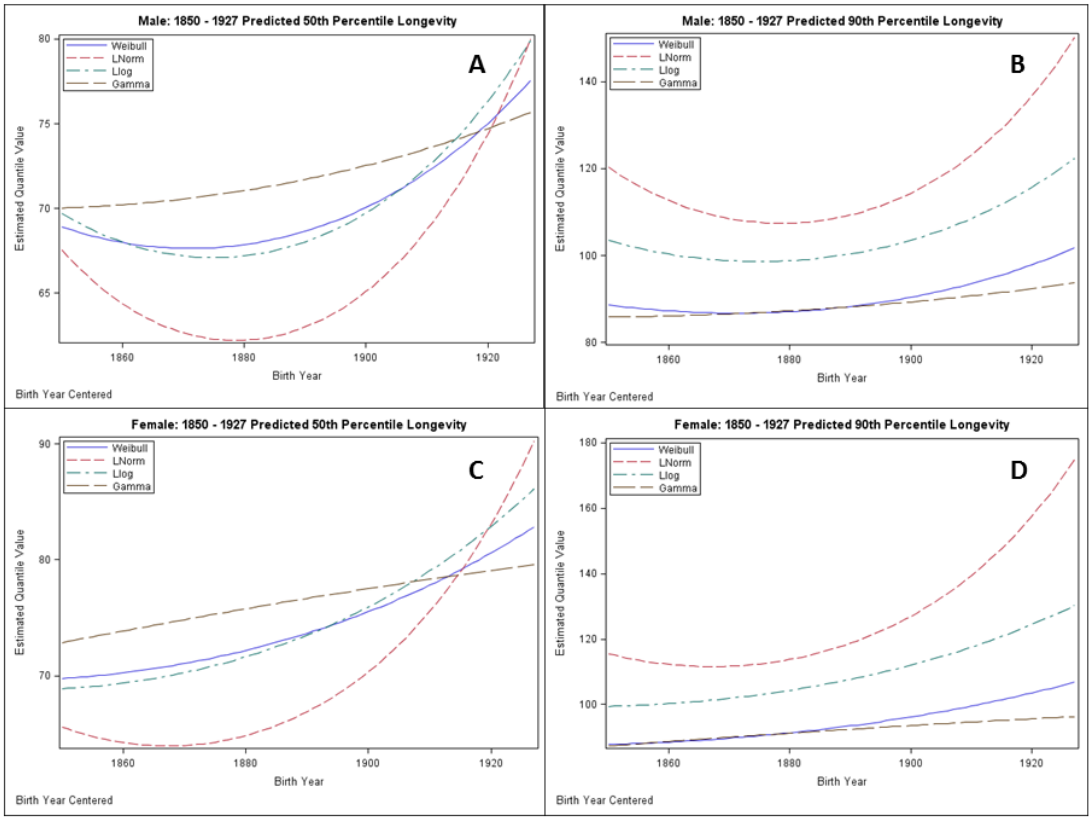

Figure 4.2. Predicted Values of Survival to the 50th and 90th Percentiles by Gender and Birth Year. Panels A and B show the estimates for male 50th and 90th percentile estimates respectively. Panels C and D show estimates for female 50th and 90th percentiles respectively.

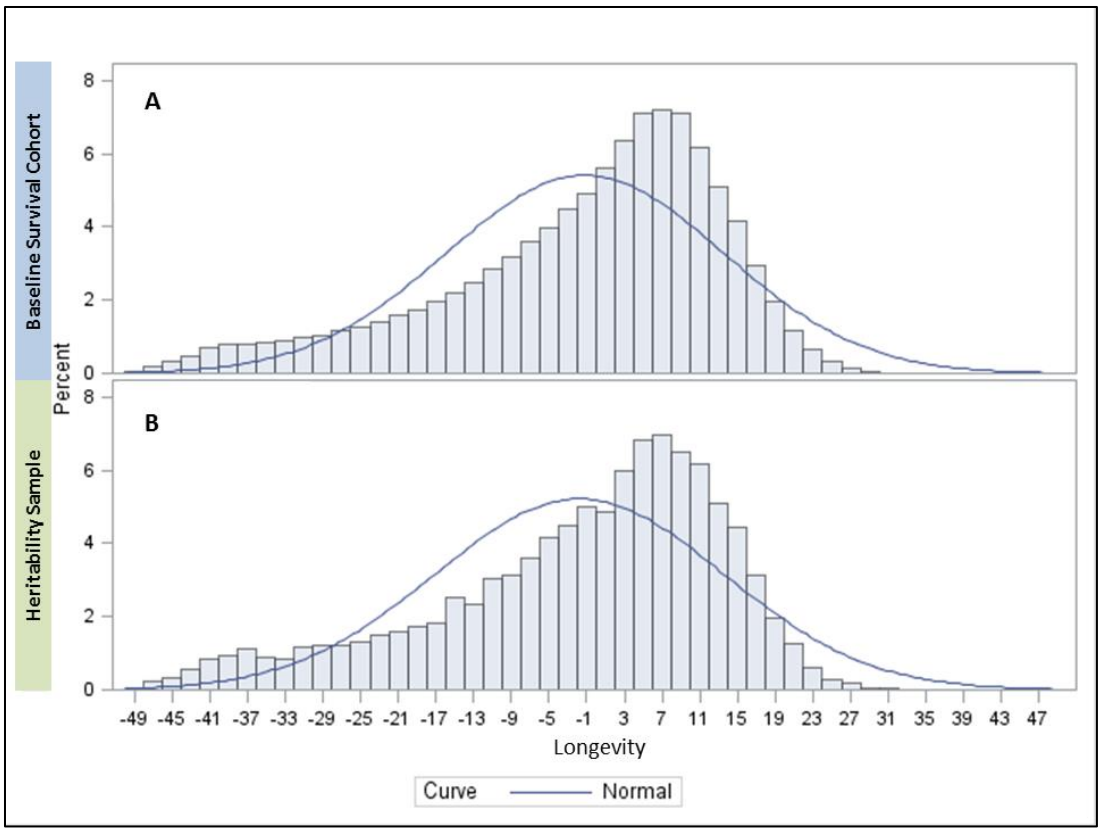

Figure 4.3. Distribution of Calculated Longevity for Individuals Born between 1850 and 1927 and Surviving to Age 30. Panel A shows the distribution for the cohort used in the baseline survival analysis (N=685,949). Panel B shows the distribution for the heritability sample (N=20,120)

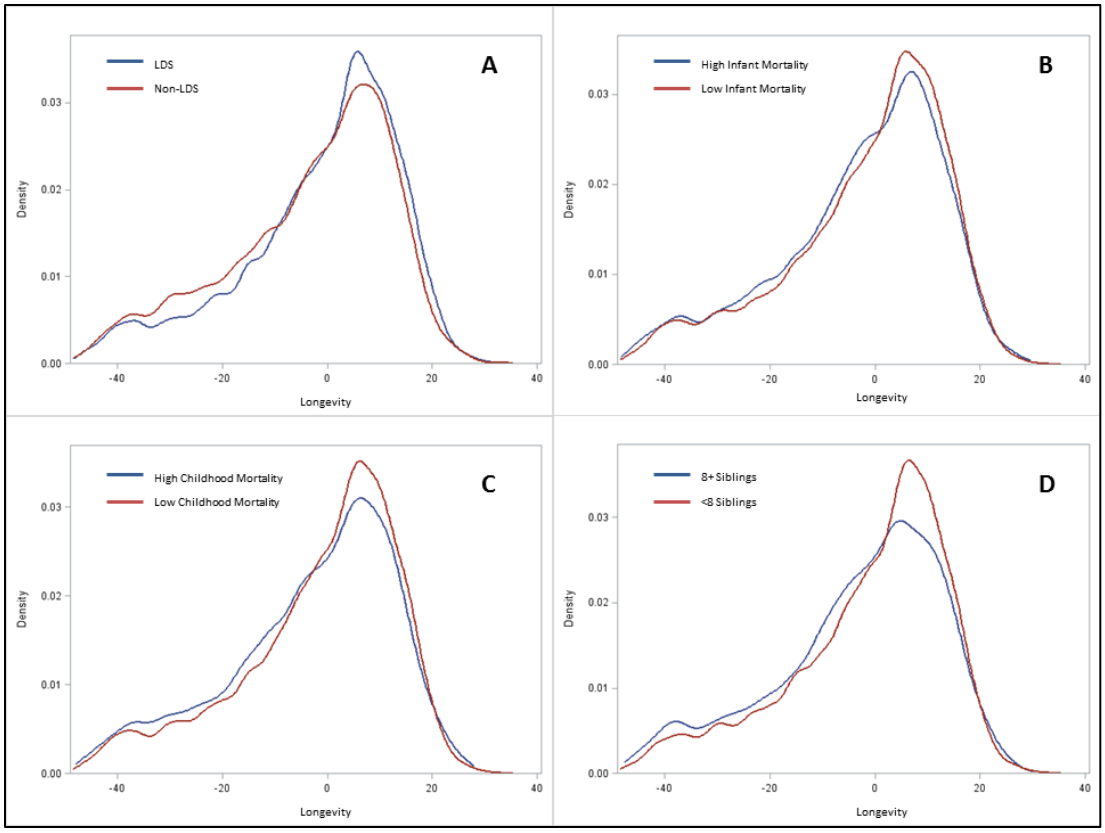

Figure 4.4. Distribution of Longevity by Environment. Empirical densities of longevity are plotted by environment.