Investigating the Genetic and Behavioural Components of the Fertility-Longevity Trade-Off

Abstract

This study investigates the evolutionary trade-off between fertility and lifespan, focusing on how genetic factors, particularly variants like APOE and FOXO3A, mediate this relationship. Survival analysis techniques were applied to data from 12,050 Health and Retirement Study participants to examine the impact of children number on mortality risk, considering socio-demographic factors such as education and marital status, and genetic predisposition to longevity and fertility through polygenic risk scores.

The findings indicated that women with two or three children had significantly lower mortality risks compared to childless women. This pattern was not observed in women with one or more than three children. A similar, albeit weaker, trend was noted in men. Education level consistently correlated with lower mortality risks across genders, while widowhood surprisingly showed higher survival rates compared to being married.

Genetic factors distinctly influenced mortality risks; women with a genetic propensity for longevity experienced enhanced survival, whereas those predisposed to higher fertility faced increased longevity risks. Men's mortality did not significantly correlate with genetic endowment to longevity and fertility.

These results support the fertility-longevity trade-off hypothesis for women, suggesting a more nuanced interplay for men. The study underscores the complex interdependence of socio-demographic, reproductive, and genetic factors in determining lifespan.

Keywords

longevity, fertility, genetic variants, survival analysis, evolutionary trade-off

Table of Contents

Chapter 1 – Introduction

Human longevity, defined as the length of an individual's lifespan, has been changing throughout the centuries due to evolving factors that determine it. Reflecting upon the past hundred years, two medical advancements stand out significantly – the development of vaccinations and the discovery of antibiotics. These, in conjunction with improvements in public sanitation and hygiene, pushed the average lifespan in the United States from less than 50 years at the beginning of the 20th century (Arias and Xu, 2020) to almost 77 years at the start of the second millennium (Arias, 2002). Accompanied by radical changes in lifestyle and nutrition, as well as environmental and social transformations, an increasing number of people are nowadays becoming centenarians or even supercentenarians. In addition to these key contributors, reproductive behaviours have also been found to play an important role in extending the human lifespan. However, the interplay between fertility and longevity is a multifaceted relationship, including physical health and socioeconomic implications.

To provide a more comprehensive portrait of longevity, it is fundamental to consider the largest – and smallest at the same time – differences that characterize every single individual, their genes. For instance, a study by Perls and Theriault (2002) found that genetic variants associated with longevity can account up to 25% of the variation in life expectancy between individuals. Thanks to the advanced genotyping and sequencing technologies, a growing body of research is identifying genetic variants associated with a longer life span.

This thesis aims to explore the relationship between the aforementioned demographic phenomena of mortality and fertility, accounting for the genetic potential of each individual to live a long life and to have a large number of offspring. The intricate web of biological, physical, and social factors that contribute to this relationship will be unravelled to address some questions by leveraging data from a large-scale study, combining genetic, demographic, and clinical information. The Health and Retirement Study (HRS) is indeed a longitudinal study that provides a rich source of data for investigating this field. The study includes approximately 20,000 American participants

per year, older than 50 years at recruitment, with follow-up data on mortality (HRS, 2023). To study the genetic mediation of the relationship between fertility and lifespan, polygenic risk scores will be used as a method of measuring the genetic predisposition for fertility and longevity. They will serve as predictors in the survival analysis, which is the statistical method used to describe how mortality risk is influenced by these factors, accounting for censoring, i.e., the situation in which some individuals are not observed for the entire duration of their lives, making their true lifespan duration unknown.

The rest of this thesis is organized as follows: Chapter 2 provides a review of the main evolutionary theories of aging, followed by evidence regarding the mediation of genetic factors in this relationship, with a focus on the demographic factors known to influence lifespan length. The polygenic score approach for studying this relationship will also be introduced. In Chapter 3, the data sources used for the analysis will be presented, along with the criteria used to select individuals. A more detailed definition of polygenic risk scores computation will be provided, followed by an extensive explanation of the survival models that were implemented. The results of these models will be presented and interpreted in Chapter 4, and then discussed in Chapter 5, highlighting the similarities and inconsistencies with previous evidence and the limitations of this study. Finally, in Chapter 6, the main findings are summarized, pointing to the implications that would arise from a more in-depth understanding of the subject, and outlining potential directions for future research.

Chapter 2 – Theoretical Framework

The Fertility-Longevity Trade-Off

One of the most intriguing ways human longevity is influenced is through the relationship between fertility and lifespan. At the heart of this discourse lie evolutionary theories proposing a profound trade-off between an organism's fertility and its potential for a longer lifespan (Kirkwood, 1977; Williams, 1957). This suggests that individuals with more children might have a shorter lifespan. The framework of these theories posits that allocating limited resources toward reproduction could compromise the body's maintenance and repair mechanisms, consequently influencing life length. Notably, this trade-off appears to have a stronger impact on females compared to males, due to the energy demands and resources required by pregnancy and lactation (Kuningas et al., 2011). Despite studies in the past decade finding no significant associations between higher mortality risk and a shorter lifespan in women with more children (Gavrilova et al., 2004), recent evidence suggests that women with a larger number of offspring tend to have shorter lifespans compared to those with fewer children (Gagnon et al., 2009; Kaptijn et al., 2015).

The mixed findings in the literature (Hsu et al., 2021) might be attributed to a non-linear relationship between fertility and lifespan. For example, Ehrlich (2015) described a U-shaped association between parity and longevity, where both infertility and having many children were associated with decreased longevity. This trajectory was confirmed by Kuningas et al. (2011), who used survival analysis to show that women with two and three children have a lower mortality risk compared to those with no children or more than three children. Therefore, while having a limited number of children might not dramatically influence lifespan, each additional child elevates the demands for energy and resources needed for pregnancy and lactation, progressively impacting longevity (Most et al., 2019).

Furthermore, the association between these two phenomena is mediated by several other factors, such as socioeconomic status, health behaviours, and the presence

of chronic diseases (Bongaarts and Watkins, 1996). For instance, women with more children might more likely come from lower socioeconomic backgrounds, in contrast to women with fewer offspring, as extensively described by Bongaarts and Hodgson (2022). Educational attainment has been identified as a crucial predictor of morbidity and mortality (Christenson and Johnson, 1995), acting through improved impulse control (Kenkel, 1991), a better understanding of health risks (Cutler and Lleras-Muney, 2010), and improved economic positions (Seeman et al., 2004). Additionally, the timing of childbirth can influence a woman's lifespan; Aladdin et al. (2017) noted that women who have children later in life tend to live longer than those who had children earlier, suggesting that older maternal age at birth may represent an important factor in longevity.

Relationship status also plays a role in fertility and is a significant predictor of life expectancy. Balter et al. (2023) highlighted how individuals not living with a partner, whether single, widowed, divorced, or separated, have a shorter life expectancy compared to those cohabiting. The "widowhood effect", highlighting an increased risk of mortality for those who have recently lost a spouse, particularly emphasizes this. A study by Moon et al. (2014) documented this effect, showing a 66% increased chance of death for the bereaved in the initial months following partner loss. Notably, further research indicates that this effect is stronger for men, with an estimated 80% increased risk of mortality in the first year after losing a spouse. However, despite its significant documentation, the widowhood effect is characterized by heterogeneity, with variations in its impact across different studies and populations. The U.S. Medicare Health Outcome Survey (Jia and Lubetkin, 2020) provided additional evidence of marriage's protective effect, particularly for younger individuals compared to the elderly.

Before delving into the analysis, it is important to note that the trade-off between fertility and longevity served as a framework for research in this field, but its dynamics are not absolute. Women with access to healthcare and a health-conscious lifestyle may defy this trade-off, choosing to have children without compromising their longevity. Moreover, while this relationship seems complex for women, it remains relatively unexplored in men. There is a limited body of research confirming the trade-off but not

establishing a robust correlation between male fertility and lifespan (Hsu et al., 2021). This dichotomy could be attributed to the differing physiological demands reproduction imposes on males and the potential influence of lifestyle choices on this association (Balawander and Orkisz, 2020).

To examine the trajectory of this trade-off in both sexes, it is intriguing to investigate whether and how the genetic predispositions for higher fertility and increased life expectancy influences the actualization of this genetic potential. Genetics play a crucial role in both reproductive behaviour and mortality, and their interaction is equally important. For example, a study by Mills et al. (2021) identified 371 genetic variants related to the age at first birth, while Reznick (2005) introduced the genetic basis of aging. A study by Tesi et al. (2020) demonstrated how polygenic risk scores of longevity could predict longer survival among individuals. Although several studies have identified specific genes mediating these phenomena, the intricate web of genetic variants governing their interaction remains enigmatic. Unravelling this might lead to a deeper understanding of the fundamental biological systems and processes that shape human existence and evolution (Tazearslan et al., 2012).

Evolutionary Theories of Aging

Upon laying the groundwork for understanding the interplay between fertility and longevity, attention naturally shifts to a central topic in gerontology: understanding the reasons behind the differing rates of aging among organisms. In pursuit of an answer, the last century has seen the proposal of several evolutionary theories.

A key theory in this field is the disposable soma theory, introduced by Tom Kirkwood in 1977. It posits that organisms possess a finite pool of resources for maintenance and repair, which must be distributed between reproduction and somatic maintenance – the process of repairing damaged cells and tissues to ensure health. Consequently, individuals allocating more resources to reproduction inevitably have fewer resources for somatic maintenance, potentially shortening their lifespan.

Another significant theory is the antagonistic pleiotropy theory, proposed by George C. Williams in 1957. This theory incorporates genetic factors, suggesting that genes that are beneficial in early life may have detrimental effects later. Genes that promote growth and reproduction early on could, over time, damage the body's cells and tissues, thereby reducing lifespan.

Recent decades have seen evidence supporting both theories, indicating a trade-off between fertility and lifespan. Individuals investing heavily in reproduction often have shorter lifespans compared to those who invest more in physical maintenance and potentially enjoy longer lives. The antagonistic pleiotropy theory views aging as a byproduct of natural selection, while the disposable soma theory considers it an adaptive strategy. Empirical support includes studies on model organisms like fruit flies, where genes encouraging early reproduction have been linked to increased age-related disease risk and early death (Sgrò and Partridge et al., 1999). Studies like Reznick et al. (2005) have further confirmed that greater investment in somatic maintenance correlates with longer lifespans in humans.

Additionally, the mutation accumulation theory, proposed by Peter Medawar in 1952, offers another perspective. It argues that a major cause of aging is the accumulation of harmful mutations in body cells over time. As organisms age, they amass mutations in their DNA, potentially leading to age-related diseases like cancer, heart disease, and Alzheimer's, thereby shortening lifespan. Several studies supported this theory, such as research on mice (Stead and Bjedov, 2021) showing those with DNA repair-preventing mutations have shorter lifespans, and findings that people with more genetic mutations tend to live shorter lives (Bin-Jumah et al., 2022).

Besides these three primary theories, numerous other theories have been proposed. It's crucial to recognize that these theories may not be mutually exclusive. Different factors play roles in the aging process, with varying impacts. For example, the epigenetic clock hypothesis, introduced by Horvath in 2013, defined aging as the result of the accumulation of epigenetic changes. These changes, while not altering the DNA sequence, affect gene expression and can lead to gene dysregulation involved in aging,

impacting lifespan. Lastly, aging has been linked to the accumulation of biological damage caused by free radicals – unstable molecules damaging cells and tissues. The oxidative stress theory, proposed by Harman in 1956, suggests that free radicals accumulate with age. The rate-of-living theory, initially by Rubner (1908) and later reinterpreted by Speakman (2005), attributes this accumulation to a slowing metabolism process due to aging.

Genetic Influence on Fertility and Longevity

When talking about the fertility-longevity trade-off, it is extremely meaningful to study the genetic basis of both phenomena. As previously mentioned, there is growing evidence that genetic variants play a role in determining both phenotypes, expressed as numbers of years lived and offspring size. Moreover, some genes might have an impact also on both fertility and longevity, leading to a possible genetic explanation for how these two traits are often seen as being in opposition, as defined by the evolutionary theories that have just been explored.

To explain how genetic variants have been associated with both decreased fertility and increased longevity, several biological pathways have been proposed, spanning from genes actively playing a role in hormones regulation and metabolism. For instance, variants in genes that regulate the production of hormones such as FSH (follicle-stimulating hormone) and TSH (thyroid-stimulating hormone) have been linked to both decreased fertility and exceptional longevity (Corbo et al., 2013; Atzmon et al., 2009). These two hormones are known to play a role in gametogenesis, which is the development and maturation of eggs and sperm, as well as in regulating the body's energy intakes and expenditures (Sendak et al., 2007). Similarly, variants in genes that regulate metabolism, such as those involved in the production of insulin and other hormones that regulate the body’s energy balance, have been associated with both decreased fertility and decreased mortality risk (Kenyon, 2011). This finding is in line with the theoretical framework proposed in the previous subchapter, as energy balance demonstrates to be an important catalyst for both fertility and longevity.

On the other hand, genetic variants that are associated with enhanced fertility may be associated with shorter lifespans. This is because the organism will be allocating more energies to reproduction, leaving fewer resources available for somatic maintenance, slowing the process of repairing damages in cells and tissues, and leading to an early death.

The APOE gene is an example of a gene that has been linked to both fertility and longevity (Robinson et al., 2020; Tesi et al., 2021). This gene codes for a protein called apolipoprotein E, which is involved in the metabolism of cholesterol and other lipids. There are three different variants of the APOE gene, named APOE-ε2, APOE-ε3, and APOE-ε4. Regarding the effects of this gene on mortality, the APOE-ε2 variant is associated with increased longevity, and people with this variant have a lower risk of developing Alzheimer's disease and other age-related diseases (Raulin et al., 2022). The APOE-ε3 variant is the most common variant in the population and is not associated with any major health risks (NIA, 2023). Lastly, the APOE-ε4 variant is associated with an increased risk of Alzheimer's disease and other age-related diseases, leading people with this variant to have a shorter lifespan (Tesi et al., 2021). A study by Jasienska et al. (2015) showed that the APOE gene is also associated with fertility, as it is a major supplier of cholesterol precursors for the production of ovarian oestrogen and progesterone. Individuals with the APOE-ε2 variant have been shown to have lower cholesterol levels (Mahley, 2016) and higher fertility compared to people with any of the other variants (Corbo et al., 2004). Considering APOE-ε2, the study by Tesi et al. (2021) showed that this variant increases the number of eggs that are released during ovulation, and the meta-analysis by Deelen et al. (2019) reported that also the quality of eggs was higher, making it easier for women to become pregnant. On the other hand, women with the APOE-ε4 variant have been shown to have a lower chance of getting pregnant and a higher risk of miscarriage (Li et al., 2014). Moreover, the APOE gene may also be involved in the development of ovarian cancer as it contributes to the production of ovarian oestrogen and progesterone.

Another gene that has been strongly linked to fertility is FOXO3A (Forkhead Transcription Factor O Subfamily Member 3a), which is a gene that encodes a protein

that helps to protect cells from damage. Hence, people with mutations in the FOXO3A gene were found to have lower fertility than people without mutations (Bao et al., 2014). As for APOE, also this gene is thought to be involved in fertility by regulating the expression of genes involved in the growth and differentiation of eggs and the progression of ovarian-related diseases (Zhang et al., 2020). Tesi et al. (2021) further explored the effects of this gene and related it to a decrease in the expression of genes involved in the apoptosis of eggs, i.e., their death.

Since both longevity and fertility are highly polygenic traits, meaning that a lot of different genes have an impact on their determination, several more variants have been discovered that influence mechanisms related to the two traits. For example, individuals with a variation in the SIRT1 gene (Sirtuin 1) tended to have shorter lifespans than people without mutations (Kilic et al., 2015). This gene is indeed involved in metabolism regulation as it activates genes that entail DNA reparation and the removal of damaged cells. It also increases the expression of genes needed for the production of antioxidants, which are chemicals that can neutralize free radicals. This biological mechanism is coherent with the oxidative stress theory and the rate-of-living theory that were proposed as theoretical pathways to unravel these mechanisms. Even if this gene did not result to be strongly related with fertility, it still interacts with it as its activity is required for optimal fertility (Alam et al., 2023).

Focusing on males, another gene has been found to interact both with their fertility potential and their longevity. The IGF-1 (insulin-like growth factor 1) is a hormone produced by the IGF1 gene, which stimulates the maturation of spermatozoa, and increases sperm count and motility, leading to an enhanced potential to generate offspring (Rodriguez et al., 2019). According to the trade-off with longevity, people whose IGF1 gene encodes for higher levels of IGF-1, also tend to be more stimulated to produce growth hormone, and this can lead to an increased risk of cancer (Grimberg, 2014) and other age-related diseases (Vitale et al., 2019). Finally, Tesi et al (2021) pointed out that IGF1 decreases the expression of genes that are involved in the repair of genetic damages, directing individuals with high production of the IGF1 hormone to shorter lifespans.

The Polygenic Risk Scores Approach

Given the fact that longevity and fertility are two complex traits, it is not very useful for the purposes of this research to consider individual genetic variants, known as Single Nucleotide Polymorphisms (SNPs), since they are weak predictors of the traits being analysed. Additionally, since reproductive behaviours are being considered, the Fourth Law of Behavioural Genetics will be recalled. This law states that a typical human behavioural trait is associated with an enormous quantity of genetic variants, each accounting for only a very small percentage of the behavioural variability (Chabris et al., 2015). For this reason, it is more convenient to use an aggregate measure of individual genetic predisposition to these traits. This measure, defined as a polygenic risk score (PRS), can be utilized for various purposes, such as identifying risk categories across individuals, measuring genetic propensity for a trait or risk condition, and studying interactions between genes and the environment (GxE). However, PRSs are population-specific, and they suffer from the drawback of not providing information on the underlying biological process that leads to a natural predisposition to a specific trait or risk.

The polygenic risk score approach offers a comprehensive way to study the genetic effects on complex traits by incorporating information from multiple genetic variants across the human genome. Additionally, they can be used to predict the risk of developing a certain trait. A polygenic risk score for an individual is defined as the weighted sum of a person's genotypes at J different loci, such that a PRS for individual i can be calculated as the sum of the risk allele counts a_{ij} (0, 1, or 2) for each SNP $j = 1, \dots, J$ related to a phenotype of interest, multiplied by a weight w_j derived as the effect size estimate from the most powerful Genome-Wide Association Study (GWAS) on the phenotype. The formula to compute PRS is as follows:

$$PRS_i = \sum_{j=1}^{J} a_{ij} w_j$$

A polygenic risk score is therefore a linear combination of the effects of multiple SNPs on the trait of interest, with the underlying model in a PRS usually being additive since the number of risk alleles are summed for each SNP included in the score.

Choi et al. (2020) provided a thorough guide to compute PRS, but the approach by Mills et al. (2020) was also used to define PRS. The first step to obtaining them is to acquire GWAS summary statistics and genome-wide data from large independent samples to avoid result inflation. Next, SNPs common between the two samples are aligned to ensure consistency in SNP measurement. This is crucial as SNPs from the genotype sample may be reported differently or erroneously. Defining each SNP as a difference in a single DNA nucleotide, they are expressed in terms of the four bases that constitute human DNA: A for adenine, C for cytosine, T for thymine, and G for guanine. The DNA molecule consists of two strands wound around each other, with adenine pairing with thymine and cytosine with guanine. For instance, if a SNP is A/C in the base sample and T/G in the target, a flip of the alleles in the base sample is required, swapping A with T and C with G, as the bases were collected on different strands. However, for ambiguous SNPs, such as A/T or C/G pairs, it is not possible to determine whether the base and target data are referring to the same allele, so these need to be removed from the analysis. In such cases, nearby SNPs can be used to capture the genetic signal. To avoid oversampling a densely genotyped genome region, linkage disequilibrium-pruning is applied, filtering variants so that the remaining ones have a linkage disequilibrium (LD) below a specific threshold. LD measures the degree to which an allele of one genetic variant is correlated with an allele of a nearby variant within a given population (Bush and Moore, 2012). The LD window, typically set to 1 Mb, defines the distance after which variants are presumed statistically independent. After this step, the most significant variant in each locus is selected through clumping, computing correlations between the target variant and nearby variants within a predefined genetic distance, and removing all nearby variants with a correlation index r^2 higher than 0.7, thus avoiding dropping and double counting of potential causal SNPs. Finally, only SNPs with an association p-value within the summary statistics lower than a certain threshold are kept. As both traits under consideration—offspring

size and longevity—are highly polygenic, a more lenient threshold is appropriate. However, including many variants in the PRS calculation increases the risk of including non-causal variants, creating noise in the PRS.

The last step in constructing a PRS is to sum up risk alleles weighted by parameters from the summary statistics. It is then possible to evaluate the strength of the explanatory or predictive association between the PRS and the trait under study by regressing the phenotype on the PRS. To understand how much variability can be explained by including the PRS in a model, it is necessary to compute the difference in the R^2 of a model that includes them and a model that includes only principal components.

Recent studies have exploited PRS to predict human lifespan and to identify genetic variants associated with exceptional longevity. For instance, Sebastiani et al. (2017) reviewed four GWAS that identified genetic variants linked to exceptional longevity, including one (Sebastiani et al., 2012) that used polygenic scores based on 150 different genetic variants to predict exceptional longevity in a cohort of Ashkenazi Jewish centenarians. The study found that the polygenic score was a strong significant predictor, with individuals in the top quintile of the score having a 2.5-fold higher odds of becoming centenarians compared to those in the bottom quintile. Another study by Joshi et al. (2017) used polygenic scores based on more than 8 million genetic variants related to exceptional longevity in a sample of over 50,000 individuals, identifying influential variants of the genes APOE and FOXO3A as expected, and XDKN2A/B, which showed a consistent association also with parental lifespan in the large meta-analysis of human longevity by Deelen et al. (2019). Lastly, a large cohort study by Marioni et al. (2018) found that the polygenic score computed on over 500,000 genetic variants was associated with mortality risk, with individuals in the top 10% of the score living on average 5 years longer than those in the bottom 10%.

Regarding fertility behaviours, it is interesting to note that most mutations undergone in the past two centuries have had a relevant impact on public policies and economic theories of family and human capital formation. However, none of the fertility

models utilized in previous literature take into account biological factors, despite the potential significant role of genetic endowment in fertility outcomes. It has been proven that genes matter for many fertility processes (Mathieson et al., 2023), both by explaining an important part of the variation observed in family outcomes and by significantly interacting with the environment. This interaction helps to understand the transformations undergone in the past 200 years. Indeed, a study by Barban et al. (2021) revealed how fertility is powerfully shaped by both genetic predispositions and societal environmental contexts, suggesting that genetic influences on fertility may be more important when social norms and economic conditions allow a broad range of life-course alternatives. In the following analysis, polygenic risk scores for the number of children were exploited to understand the strength of genetic prefisposition to fertility in shaping individuals' longevity.

Chapter 3 - Materials and Methods

In this chapter, the databases and relative variables under analysis will be listed. Subsequently, the statistical method of survival analysis will be thoroughly explained, with a particular emphasis on the construction and interpretation of the models. Next, the paper will define the three models used to investigate the relationship between mortality and reproductive behaviour. These models also consider the potential mediation by individual genetic endowment for longevity and fertility, as well as other socioeconomic predictors like marital status and education level.

Data Sources

The data source for this analysis is the Health and Retirement Study (HRS) administered by the Institute for Social Research of the University of Michigan. This is a longitudinal panel survey conducted on a representative sample of approximately 20,000 Americans over the age of 50 every two years. The study is sponsored by the National Institute of Aging (NIA) and the Social Security Administration (SSA). The HRS collects information about health, family structure, and various socioeconomic details through different surveys. Additionally, genetic data were collected from almost 20,000 respondents between 2006 and 2012. As stated by Richard Hodes, director of the NIA, "The addition of genetic data provides a major new dimension for the study and is expected to result in much deeper insights into how we age."

The HRS Cross-Wave Tracker File (HRS, 2023) contains demographic information for every participant who was interviewed in any wave from 1992 to 2020, totalling $N = 43{,}559$ uniquely identified records. This source has been used to retain data regarding participants' demographic details. Gender was categorized as females and males, and the analyses were conducted separately for each gender. Regarding marital status, the categories in the surveys that were included in the analysis were "Married," "Separated or divorced," "Widowed," and "Never married." Education level was categorized into four groups: "No degree" for those who did not complete their

education, “High school diploma” for those with a high school diploma or equivalent, such as a General Education Development (GED) diploma, “College degree” for individuals with a 2-year or 4-year college degree, and “Advanced degree” for those with a master’s or professional degree.

Death occurrence and date were recorded by the HRS through contacting the next-of-kin of non-responders for an exit interview, and by querying the National Death Index at the end of each wave, ensuring data completeness. For data regarding participants’ reproductive behaviour, demographic data were merged with Rand HRS Family Data 2020 (HRS, 2023), which includes information on HRS respondents’ children and spouses, totalling $N = 42{,}406$ uniquely identified records from 1992 to 2020. Information about the number of children of each participant was retrieved and grouped into 5 categories: “No children,” “1 child,” “2 children,” “3 children,” and “More than 3 children.” Considering previous evidence on the relationship between the number of children and longevity (Kuningas et al., 2011), it was suggested to consider separately the groups of people without children and those with one child.

As previously mentioned, the HRS provides genetic data from over 19,000 respondents who agreed to be genotyped. For this study, the Polygenic Score Data (HRS, 2021) was utilized. The dataset contains Polygenic Risk Scores (PRS) computed for several traits, including the number of children and longevity, for $N = 12{,}090$ HRS respondents who provided salivary DNA between 2006 and 2012.

To ensure the integrity of the study, participants of European ancestry were carefully selected in order to minimize population stratification. Population stratification refers to systematic variations in allele frequencies among subpopulations within a larger group, which can introduce bias (Price et al., 2006; Hamer and Sirota, 2000). It's important to note that the majority of Genome-Wide Association Studies (GWAS), including those referenced in this analysis, primarily rely on European-ancestry samples, limiting their applicability to other ancestral backgrounds. Therefore, the study adhered to the established definition of European ancestry in genetic epidemiology (Lu et al., 2022).

The final sample consisted of self-reported non-Hispanic whites who exhibited principal component (PC) loadings within ± one standard deviation of the mean for the first two eigenvectors in Principal Component Analysis (PCA) among all unrelated study subjects. Subsequently, an additional PCA was conducted within the European sample to generate eigenvectors for covariates in the statistical models, effectively controlling for potential population stratification (Price et al., 2006). This rigorous approach ensured that any genetic elements related to common ancestry that could artificially correlate with PRS and longevity were appropriately accounted for.

Regarding PRS for each trait, specifically the number of children and longevity, the analysis was based on a replicated GWAS involving genotyping of approximately 2.4 million SNPs. Stringent quality controls (HRS, 2013) were meticulously implemented to maintain the reliability and accuracy of genetic data. Initially, first-degree relatives were excluded to mitigate potential bias arising from close family relationships. Subsequently, individuals with a high rate of missing genetic data (call rates exceeding 2%) were excluded to ensure data quality. Additionally, SNPs with a call rate below 98% were removed to address incomplete genotyping. SNPs showing significant deviations from Hardy-Weinberg Equilibrium (HWE) with a p-value < 0.0001 were also systematically excluded to prevent issues related to population substructure. Moreover, SNPs associated with chromosomal anomalies were rigorously eliminated. As a result, approximately 21 million SNPs were imputed to the 1000 Genome Project reference panel, utilizing the original 1,905,968 SNPs that met quality control criteria.

The construction of longevity PRS, as described by Broer et al. (2015), relied on summary statistics from a 2015 GWAS conducted by the Cohorts for Heart and Aging Research in Genomic Epidemiology (CHARGE) consortium. This comprehensive GWAS meta-analysis incorporated data from 6,036 individuals who reached the age of 90 years or older, along with 3,757 controls who passed away between the ages of 55 and 80. All participants shared European ancestry, and their genetic information was imputed based on approximately 2.5 million SNPs, referencing the HapMap sample. Utilizing an additive model, the analysis sought associations between individual SNPs

and longevity while adjusting for variables such as sex and genetic PCs. Subsequently, PRS were computed and standardized. However, it's important to note that due to the removal of the HRS cohort from the summary statistics, the estimates presented in the PRS may not align precisely with those in the original publication by Broer et al. (2015).

For the assessment of the number of children ever born, PRS were grounded in the findings of the 2016 investigation conducted by the Sociogenome consortium, led by Barban. This extensive meta-analysis synthesized data from a large cohort of 343,072 individuals of European ancestry. The dataset encompassed approximately 2.4 million SNPs that successfully passed stringent quality controls. Associations were meticulously adjusted to account for principal components (PCs), thereby minimizing confounding effects related to population stratification. Furthermore, considering the diverse origins of participants across 62 different cohorts, adjustments were made for respondents' birth years to mitigate cohort effects. The analysis focused on individuals who had completed their reproductive period, as delineated by an age threshold of 45 years for women and 55 years for men. Following computation, PRS were standardized, following a similar approach to previous analyses.

Given the primary focus of this study on exploring the relationship between longevity and fertility and how genetic factors mediate this connection, data from the three HRS datasets were merged. Consequently, only individuals with both available genetic data and information about the number of children were retained, resulting in a sample size of $n = 12{,}050$ individuals for this comprehensive analysis.

Survival Analysis

To explore the factors influencing people's longevity, survival analysis techniques were employed. It's important to highlight our deliberate focus on providing a comprehensive definition of the fundamental concepts of survival analysis. This emphasis on theoretical foundations is purposeful, as it forms the basis for the subsequent sections on model fitting and interpretation. Survival analysis encompasses

powerful tools for describing and modelling the time until a specific event, such as death, occurs. It's crucial to acknowledge that, for some individuals, this event may remain unobservable for various reasons, such as loss to follow-up, withdrawal from the study, or occurrence of a different event that precludes observation of the primary event, known as censoring in survival analysis.

To define the three key functions in survival analysis, let's consider the age at which respondents pass away as a non-negative continuous random variable T. In survival analysis, this variable is referred to as survival time, representing the duration until the event of interest - in this case death - occurs. Its distribution function is defined as:

$$F(t) = P(T \leq t)$$

Furthermore, the probability density function $f(t)$ allows to define the probability that a randomly selected individual will survive beyond a certain age t through the survival function. This function illustrates how the probability of survival decreases from 1 (when all individuals are alive) to 0 (when all subjects in the sample have passed away), and it is expressed as:

$$S(t) = P(T > t) = 1 - F(t)$$

From the survival function, the cumulative hazard function can be derived, which represents the accumulated risk of death as age progresses. It is calculated as:

$$H(t) = -\log S(t)$$

Subsequently, the hazard function is obtained, also known as the hazard rate, which measures the instantaneous risk of an individual dying immediately after reaching a certain age t, given that they were alive at age t:

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} = \frac{d}{dt} H(t)$$

Therefore, if the age of death were known for each HRS respondent, the estimation of the three aforementioned survival functions would have been straightforward. However, for some individuals in the sample, death was not observed, resulting in what are known as censored observations. In the case of random right censoring, as in this study, only a lower bound of the true survival time is observed for certain individuals. This is because some were lost at a certain wave of the survey or were still alive in 2020. Consequently, only one of the two variables, either T (the survival time) or C (the censoring time), is recorded for each subject in the sample, depending on whether the respondent was deceased or censored. Therefore, for each subject in the sample, a set of (Y_i, Δ_i) for $i = 1, \ldots, 12{,}050$ was calculated, where Y_i represents the time until the first event (death or censoring) occurred, and Δ_i is an indicator variable equal to 1 if the subject was deceased and equal to 0 if the subject was censored:

$$Y_i = \min(T_i, C_i)$$

$$\Delta_{\mathrm{i}} = I(T_i \leq C_i) = \begin{cases} 0 \; if \; T_i > C_i \\ 1 \; if \; T_i \leq C_i \end{cases}$$

For deceased subjects, the age at death was computed as the difference between the date of birth and the best available information about the death date. For censored individuals, the most recent year in which the respondent was known to be alive was used to estimate the censored survival time.

In 1958, Kaplan and Meier (JASA, 1958) introduced a nonparametric estimation method for the survival function with right-censored data. This estimator is based on the order in which deaths and censored observations occur. Given r distinct deaths, with $r \leq n$, the event times can be ordered as $Y_{(1)}, \ldots, Y_{(r)}$, along with the corresponding number of deaths $d_{(1)}, \ldots, d_{(r)}$, and the size of the risk set $R_{(j)}$, which represents the subsample that is still alive at that time. To obtain an estimate for the survival function, the likelihood function is utilized:

$$Likelihood = \prod_{i=1}^{n} \left(h(Y_i)\right)^{\Delta_i} S(Y_i)$$

This likelihood is maximized to derive the estimator of the survival function:

$$\hat{S}(t) = \prod_{j:Y_{(j)} \leq t} (1 - h_{ij})$$

Where the hazard rate can be estimated by the ratio of the number of deaths to the population at risk at each age:

$$\hat{h}_{(j)} = \frac{d_{(j)}}{R_{(j)}}$$

The Kaplan-Meier estimator of the survival function is then obtained by substituting this estimate into $\hat{S}(t)$:

$$\hat{S}(t) = \prod_{j:Y_{(j)} \leq t} \frac{R_{(j)} - d_{(j)}}{R_{(j)}}$$

The estimated function $\hat{S}(t)$ is a step function that changes its value at the time of each event. The Kaplan-Meier (KM) survival curve, which is a plot of the KM survival probability against time, provides a useful summary of the data. It can be utilized to estimate measures such as the median survival time. However, since the survival function often exhibits a right-skewed distribution, calculating the mean survival time estimate is highly sensitive to outliers and necessitates certain assumptions, particularly regarding the largest observation. When the largest observation is censored, the KM estimator becomes inconsistent and unreliable in the right tail. Therefore, the estimation of the median survival time is usually preferred and is computed as the age at which the estimated probability of survival is 0.5:

$$\hat{x}_{0.5} = \inf\left\{t \middle| \hat{S}(t) \leq 0.5\right\}$$

In this analysis, to elucidate the impact of different variables on survival probability or mortality risk, the initial step involved univariate analysis to compare KM survival curves among the various categories of the variables under investigation. First, survival curves based on the number of children were compared to provide an overview of the relationship between fertility and mortality. Subsequently, differences based on demographic covariates, including sex, marital status, and education level, were evaluated to gain insights into how they might interact within the subsequent models. Lastly, a comparison was made between individuals in the highest and lowest deciles of the polygenic risk score distributions for longevity and the number of children ever born. This comparison aimed to investigate differences in survival based on the genetic predisposition to a longer lifespan and having more children.

To assess whether differences among categories were statistically significant, the log-rank test was employed to compare the survival curves of each category of the aforementioned variables. Essentially, this test compares the observed number of deaths in each group to the expected number of deaths if the null hypothesis were true, assuming that the survival curves were identical. The test statistic follows an approximate χ^2 distribution.

Cox Proportional Hazards Models

Using univariate analysis, the impact of each predictor on mortality was examined, disregarding the influence of other factors. An alternative method for examining the relationship between individuals' survival time and one or more predictor variables is the Cox proportional-hazards model, as introduced by Cox in 1972. Unlike the Kaplan-Meier approach, which is limited to studying categorical variables, this regression model is suitable for both quantitative and categorical predictors. Moreover, the Cox regression model allows for the simultaneous assessment of the effects of multiple risk factors on survival time. In situations like this, where several covariates may potentially impact an individual's longevity, it is crucial to account for the influence of others when investigating survival in relation to any particular factor.

The Cox model is defined by the hazard function $h(t)$ which represents the risk of death at age t, and is expressed as follows:

$$h(t) = h_0(t) \cdot \exp\left(b_1 x_1 + b_2 x_2 + \cdots + b_p x_p\right)$$

In this equation, the term h_0 denotes the baseline hazard, corresponding to the hazard's value when all covariates $(x_1, x_2, \ldots, x_p)$ are set to zero. This quantity exhibits an exponential relationship with the covariate values and the coefficients $(b_1, b_2, \ldots, b_p)$ which quantify the effect size of each covariate.

Cox proportional hazards models rely on two primary assumptions. The first assumption is that the relationship between the log hazard and each covariate must be linear, a condition that can be confirmed through residual plots. The second crucial assumption is that the hazard curves for different groups of observations should remain proportional over time, indicating constant relative hazards. In fact, when considering two respondents, denoted as k and k', with distinct sets of covariate values x and x', their respective hazard functions can be expressed as follows:

$$h_k(t) = h_0(t) \cdot \exp\left(\sum_{i=1}^{n} \beta x\right)$$

and

$$h_{k'}(t) = h_0(t) \cdot \exp\left(\sum_{i=1}^{n} \beta x'\right)$$

Since the baseline hazard is the same in both equations, the hazard ratio for these two patients becomes independent of time:

$$\frac{h_k(t)}{k'(t)} = \frac{\exp(\sum_{i=1}^{n} \beta x)}{\exp(\sum_{i=1}^{n} \beta x')}$$

Consequently, the hazard of the event in any group is a constant multiple of the hazard in any other group, establishing the Cox model as a proportional-hazards model.

This assumption implies that, as mentioned earlier, the hazard curves for different groups should remain proportional. In simpler terms, if one individual has a risk of death at an initial time point that is twice as high as another individual, this relative risk should remain consistent at all later time points.

It is essential to assess whether the Cox regression model adequately describes the data. In this analysis, three types of diagnostics were conducted: testing the proportional hazards assumption, detecting nonlinearity in the relationship between the log hazard and the covariates, and examining influential observations or outliers. To assess these model assumptions, residuals were analysed. To check the proportional hazard assumption, Schoenfeld residuals were calculated. Martingale residuals were employed to assess nonlinearity, and deviance residuals, derived from a transformation of Martingale residuals, were used to examine influential observations. Additionally, to assess the overall significance of the models, the likelihood ratio test, the Wald test, and the score log-rank test were evaluated. It's worth noting that these three methods are asymptotically equivalent for large sample sizes and are expected to yield similar results.

If the assumptions are not violated, interpreting the direction and strength of coefficients is more convenient when calculating hazard ratios (HR), obtained as the exponential of each coefficient:

$$HR_j = \exp(\beta_j)$$

For continuous covariates, a $\beta_j > 0$ or equivalently, $\beta_j > 0$ indicates that as the value of the j^{th} predictor increases, the hazard of death increases, leading to a shorter survival duration. In other words, $HR_j > 1$ suggests a covariate that is positively associated with mortality risk and negatively associated with lifespan duration. Conversely, when $\beta_j < 0$ or equivalently, $HR_j < 1$, the variable is negatively associated with mortality risk, making it a protective factor as its value increases. For binary or categorical covariates, the coefficients are interpreted similarly, comparing the increase or decrease in mortality risk associated with belonging to a particular category of the predictor relative to a baseline category of the same variable.

Therefore, mortality risk serves as a measure to assess how survival probability is distributed in the sample, reflecting longevity. In this thesis, three different survival models were fitted separately for men and women. First, to describe the effect of the number of children on survival, then adding demographic covariates to account for education level, marital status, and year of birth. Finally, polygenic risk scores for longevity and the number of children ever born were utilized to understand the contribution of the genetic component to explain this phenomenon.

The first model was defined using only the number of children as a variable to predict mortality risk, separately for men and women. Since one of the categories will serve as the reference category, there will be just four parameters in the model, corresponding to the groups with one, two, three, or more than three children, while the reference category are those without children. The equation for the hazard function is as follows:

$$h(t) = h_0(t) \cdot \exp\left(\sum_{j=1}^{4} \beta_j \cdot \text{Children}_j\right)$$

Next, a second model was fitted, adding the effect of demographic covariates to control for potentially influential factors such as education level, marital status, and year of birth. There is evidence from the previous chapter that these factors impact longevity. Both education level and marital status are divided into four groups, with high school diploma holders and married individuals serving as the reference categories for each. For each of these two variables, three parameters were added to the model, along with one parameter for the year of birth. The equation for the second model is as follows:

$$h(t) = h_0(t) \cdot \exp\left(\sum_{j=1}^{4} \beta_\text{j} \cdot \text{Children}_j + \beta_5 \cdot \text{Birth Year} + \sum_{j=6}^{8} \beta_\text{j} \cdot \text{Education}_j + \sum_{j=9}^{11} \beta_j \cdot \text{Martial Status}_j\right)$$

Finally, to gain insights into how much residual variance was attributed to genetic constitution, both the PRS for longevity and the PRS for fertility were added as continuous covariates. This was done in addition to the inclusion of the ten main genetic PCs, the computation of which was described in the previous chapter. These PCs control for any genetic aspects of European ancestry that might be spuriously correlated with the PRS and longevity. The model is defined by the following equation:

$$h(t) = h_0(t) \cdot \exp\left(\sum_{j=1}^{4} \beta_{\mathrm{j}} \cdot \mathrm{Children}_j + \beta_5 \cdot \mathrm{Birth\ Year} + \sum_{j=6}^{8} \beta_{\mathrm{j}} \cdot \mathrm{Education}_j + \sum_{j=9}^{11} \beta_j \cdot \mathrm{Martial\ Status}_j + \beta_{12} \cdot PRS\ Longevity + \beta_{13} \cdot PRS\ Fertility\ + \sum_{j=14}^{24} \beta_j \cdot PCA_j\right)$$

The results of the three models and their interpretation will be provided in the following chapter. It is important to note that the assumptions of the three Cox models were checked for each of them before interpreting the coefficients.

Chapter 4 - Results

Beginning with a concise description of the distribution of the variables used in the analysis, the Kaplan-Meier estimates' results are then presented through plots to compare survival among various categories of the primary covariates. Following a discussion of potential violations of model assumptions, the coefficients of the three Cox models are provided and subsequently interpreted.

Descriptive Statistics

After applying inclusion and exclusion criteria to the full datasets, a sample of $n = 12{,}050$ individuals was retained for analysis. Table 1 provides a comprehensive overview of the sample, presenting the range, mean, and relative standard deviation (SD) for continuous variables. For categorical variables, it includes the frequency, both in absolute numbers and as percentages for each group.

Table 1: Characteristics of the study sample.

Variable	**Min**	**Max**	**Mean (SD)**
Year of birth	1905	1980	1941 (12)
Age at death/last follow-up	40	105	77.19 (10.66)
PRS Longevity	-3.50	4.63	0 (1)
PRS Fertility	-4.65	3.50	0 (1)

Variable	Status	N (frequency %)
Mortality	Deceased	4,226 (35.07%)
	Alive	7,824 (64.93%)
Sex	Male	5,175 (42.95%)
	Female	6,875 (57.05%)
Number of children	No children	1,470 (12.20%)
	1 child	1,481 (12.29%)
	2 children	3,850 (31.95%)
	3 children	2,714 (22.52%)
	More than 3 children	2,535 (21.04%)
Education level	No degree	1,447 (12.01%)
	High school diploma	6,791 (56.36%)
	College diploma	2,517 (20.89%)
	Advanced degree	1,295 (10.75%)
Marital status	Married	6,343 (52.64%)
	Never married	395 (3.28%)
	Separated/Divorced	1,672 (13.88%)
	Widowed	3,640 (30.21%)
PRS for longevity and fertility	Both high	2,543 (21.10%)
	High longevity, low fertility	3,204 (26.52%)
	Low longevity, high fertility	2,766 (22.9%)
	Both low	3,564 (29.50%)

The selected respondents were born between 1905 and 1980, resulting in a diverse age distribution. The average age at death in this sample was 77 years. The composition of the sample included a nearly equal number of males (43%) and females (57%). Among these individuals, 35% had deceased before 2020, while the remaining 65% were still alive at the last wave of interviews.

The individuals exhibited a range of reproductive behaviours, with 12% choosing not to have children and another 12% having just one child. Among the remaining respondents, 32% had two children, and the rest were equally divided between those with three or more than three children.

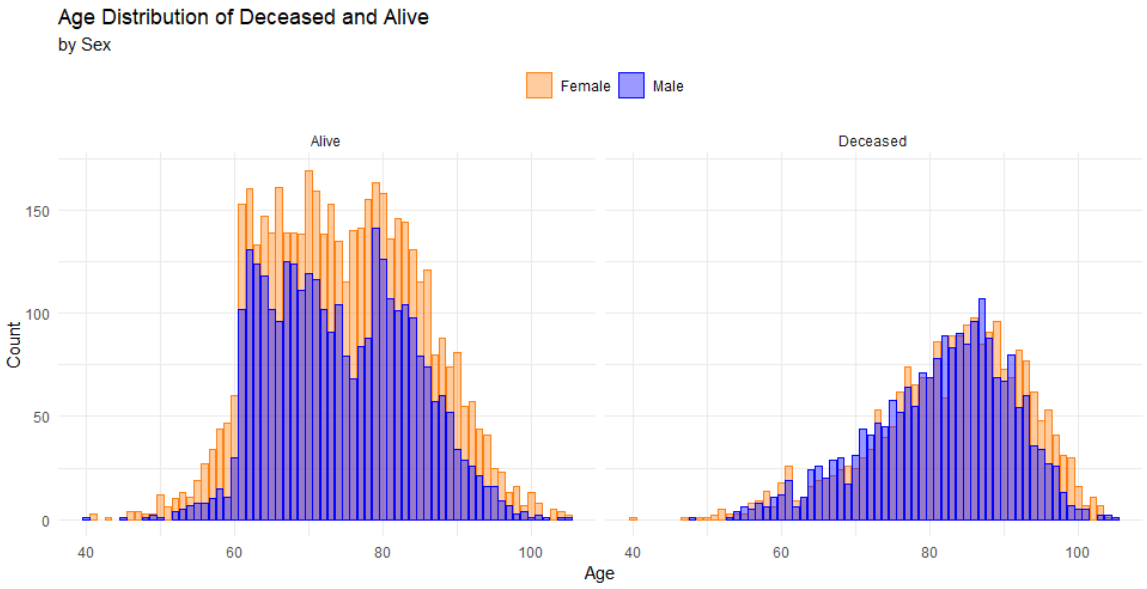
Regarding educational backgrounds, 12% did not possess any degree, while more than 55% had obtained a high school diploma, 21% had completed college, and the remaining 11% held advanced degrees. When categorizing by marital status, over half of the sample was married, with only 3% being unmarried, nearly 14% being separated or divorced, and 31% being widowed.

Finally, the distributions of the polygenic scores for both longevity and the number of children were standardized. However, in Table 1, their distribution is reported based on their sign, considering the negative scores as low and the positive scores as high. Individuals were almost evenly divided into four categories: those with both high or both low PRS and those with a major genetic endowment for one trait rather than the other.

Although the sample overview is presented for males and females together, all analyses were conducted separately by sex. Therefore, all results will be evaluated independently and compared. To investigate the distribution of age at death or at the last follow-up, the observed distributions of survival and censored times were plotted. This was done to compare differences between deceased respondents and those who were censored. The comparison is depicted in Figure 1, where it is evident that the two groups of people exhibit different distributions. Those who died before 2020 have a distribution skewed to the right, with a median age at death of 84 years for women and 83 years for

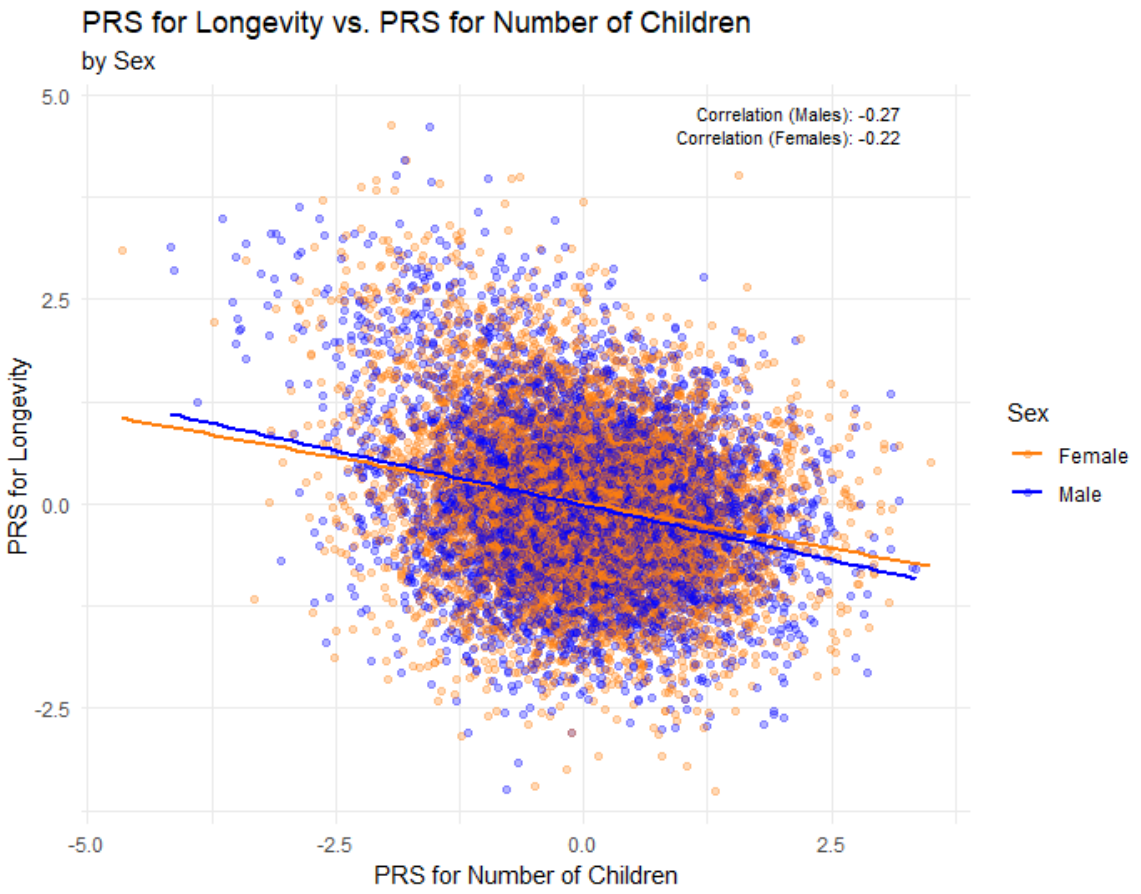
men, compared to those who were still alive at the last wave of the survey, who have a median age of 74 years at censoring.

Figure 1: Distribution of age at death or at last follow-up for deceased and survivors divided by sex.



As both the PRS for longevity and the PRS for the number of children are components of the third survival model, it is of interest to examine their correlation. The correlation between the genetic predisposition to these two traits is negative and significant for both sexes, with a value of $r^2 = -0.22$ for females and $r^2 = -0.27$ for males.

Figure 2: Correlation among polygenic risk scores for longevity and fertility divided by sex.

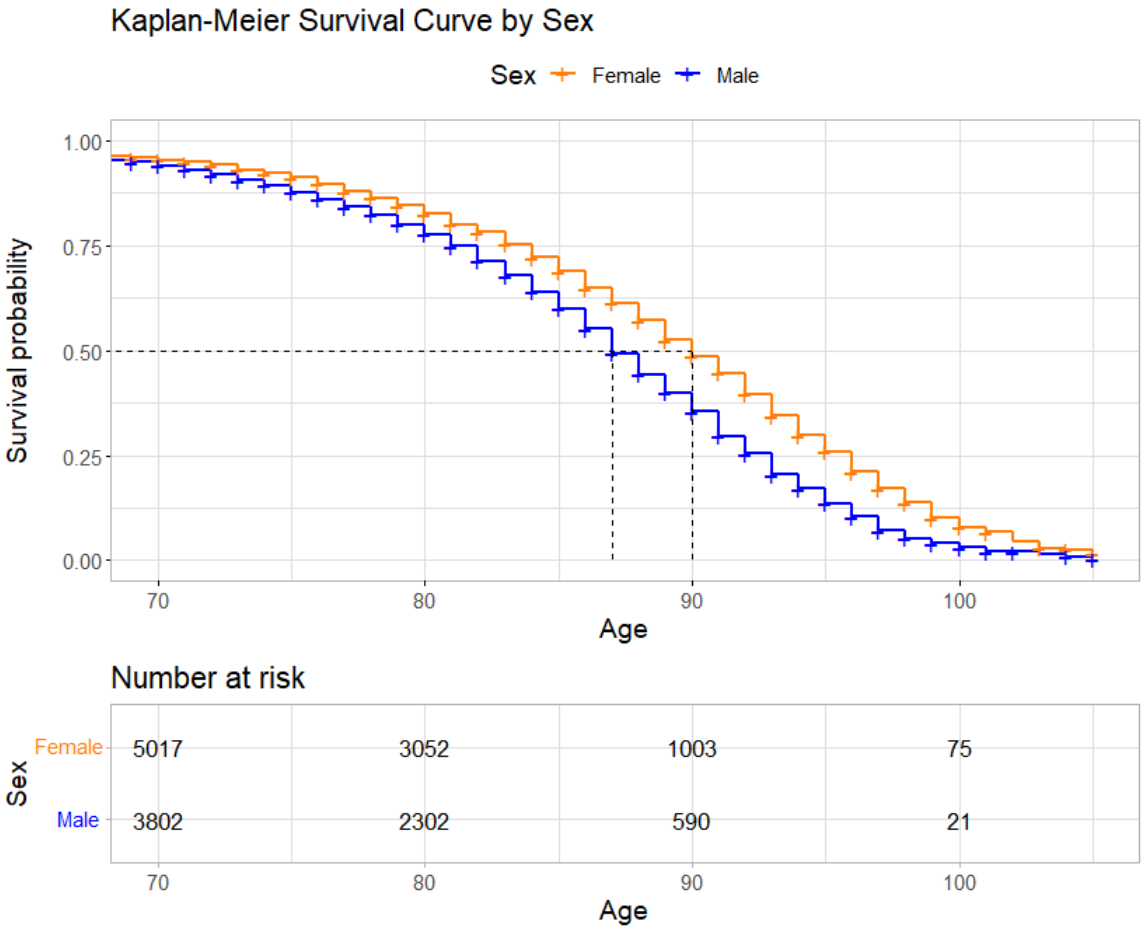


Kaplan-Meier Survival Estimates

To begin describing the impact of different characteristics on survival, Kaplan-Meier curves were estimated and plotted for each of the variables included in the analysis. In the following figures, the horizontal axis represents the age at which individuals died or were censored, while the vertical axis tracks the survival probability, i.e., the proportion of people surviving. The figures presented focus on survival probabilities after 70 years, as there were no remarkable differences before this age. Consequently, the curves in the following figures do not start with a survival probability of 1 because some individuals had already exited the population at risk.

Since the analysis was conducted separately for men and women, it was interesting to visualize the overall difference in survival trajectories between the two sexes. In Figure 2, the two lines represent the survival curves of females and males, with each plus sign "+" indicating a different year in which events occurred. The median survival time, which is the age at which the survival probability $S(t) = 0.5$, is approximately 87 years for males and 90 years for females, suggesting longer survival for females compared to males. In this case, the log-rank test for testing the difference in survival yields a p-value < 0.001, indicating a significant difference between men and women.

Figure 2: Kaplan-Meier survival curves for males and females with risk table.

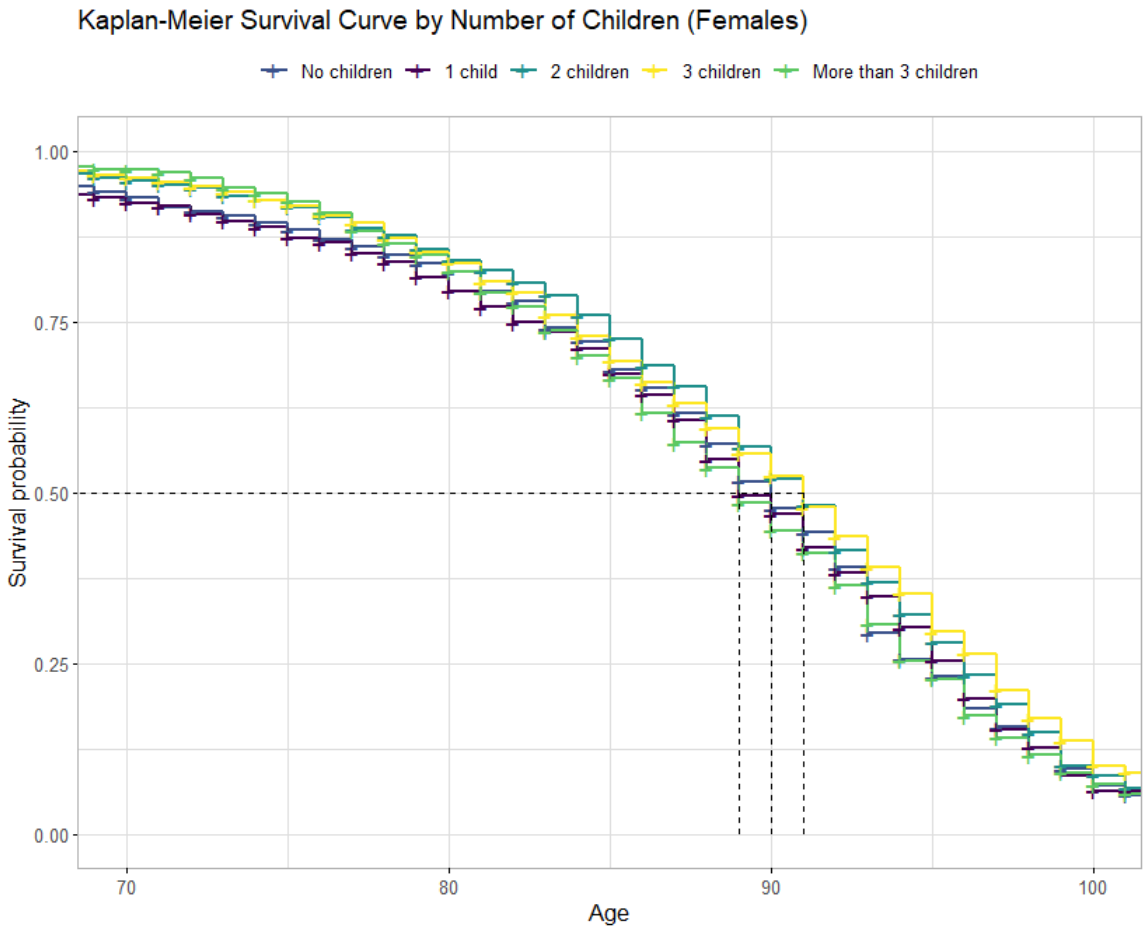


To understand how the primary variable of interest, the number of children, impacts survival, Kaplan-Meier curves were plotted separately for women and men based on their number of children. Women with no children have a median survival age

of 90 years, which is one year higher than women with either one child or more than three children but one year less than women with two or three children. For men, except for those with two children, who have a median survival age of 88 years, all other groups have a median survival age of 87 years. The survival probability is significantly different between women with different numbers of children, with a p-value of 0.003 for the log-rank test. However, the difference is not significant for men, as the log-rank test yields a p-value of 0.6.

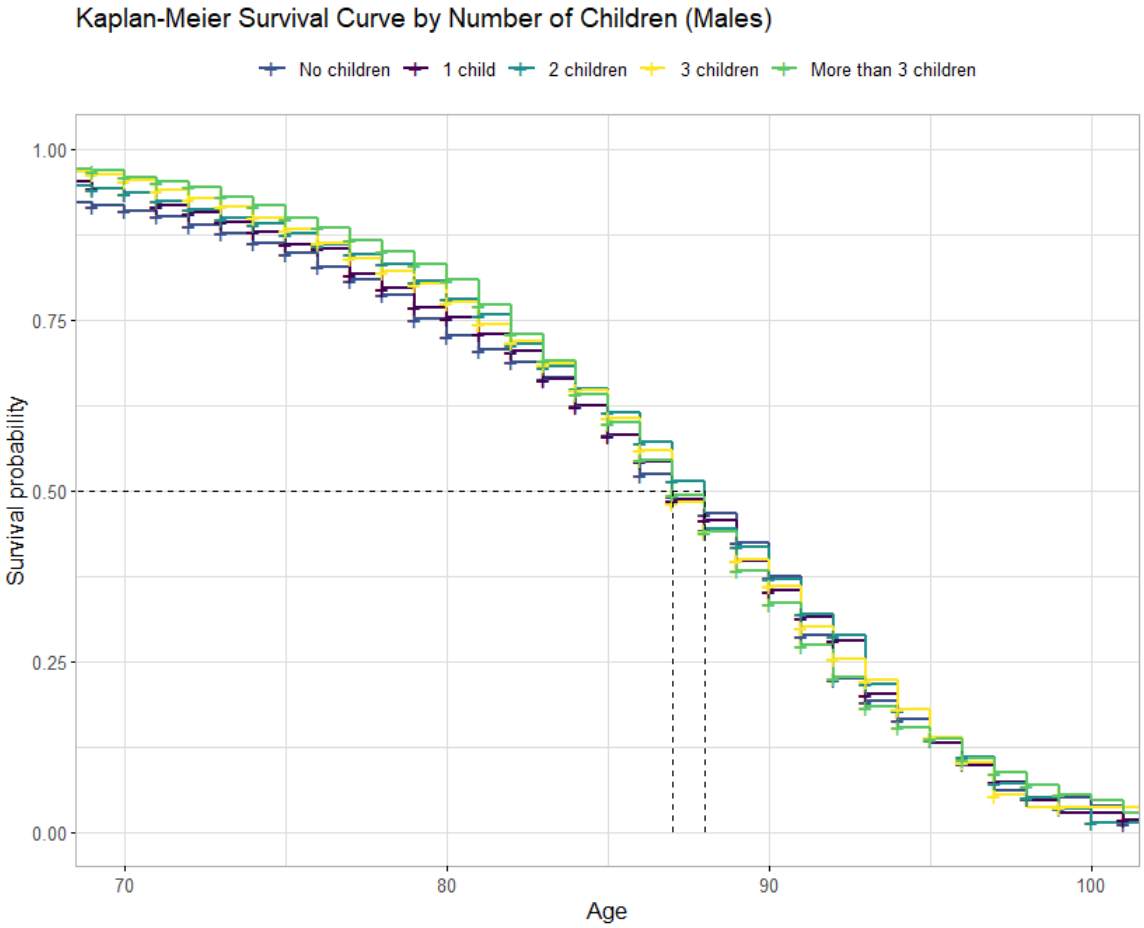
In the case of women, the survival trajectories of the five groups represented in Figure 2 are complex, but it appears that the survival curve of individuals with two children is among the highest, along with the curve of women with three children. On the other hand, women with no children, just one child, or more than three children have a lower probability of surviving at all ages compared to those with two or three children.

Figure 2: Kaplan-Meier survival estimator by number of children for females.



Men's survival trajectories, as shown in Figure 3, are even more intricate than those of women. While men with two children seem to have a slight survival advantage, the other four groups do not show any evident differences between each other, although men with fewer than two children may have a slightly lower probability of surviving than those with more than two children.

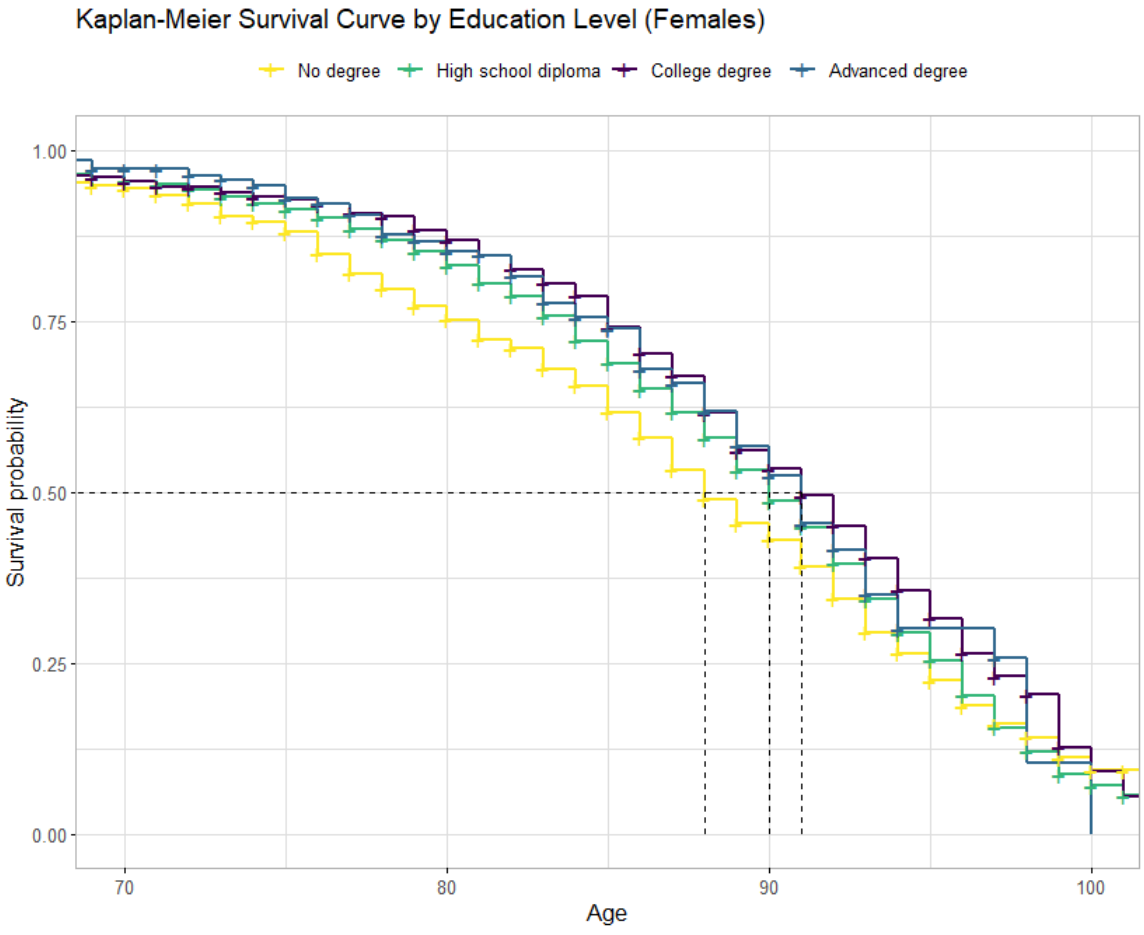
Figure 3: Kaplan-Meier survival estimator by number of children for males.



In this investigation, an examination was conducted to explore survival differences among individuals with varying marital statuses and educational backgrounds. The purpose of this analysis was to gain an initial understanding of the relationships between these factors and their impact on survival. This step was crucial for comprehending the expected direction of the covariate effects in the estimated models.

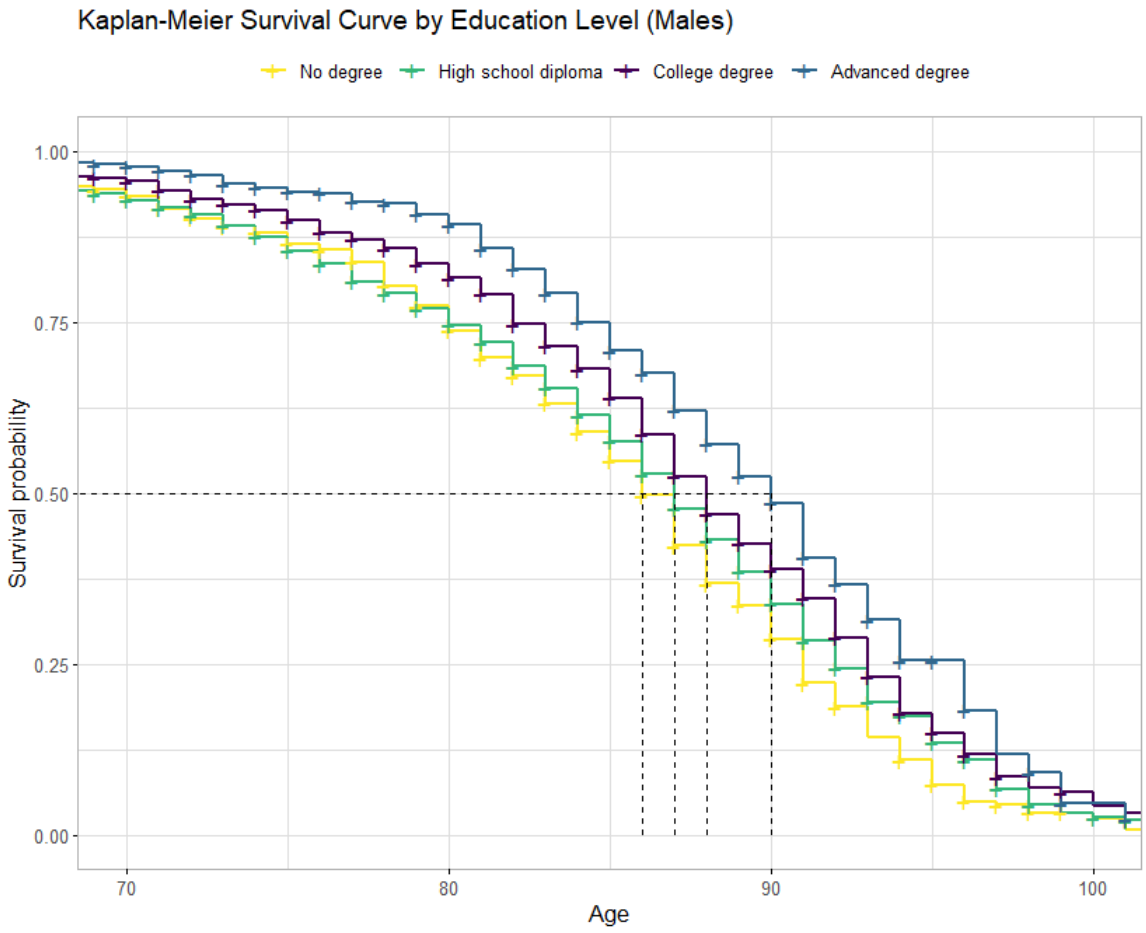
When considering the highest level of education among women, it becomes evident that those with a college degree or a more advanced degree tend to have the highest median survival age, which is typically around 91 years. Women who completed high school typically exhibit a median survival age of 90 years, while women without a degree generally have a median survival age of 88 years. Examining Figure 4, one can observe not only variations in median survival age but also differences in the survival curves. The survival curves of women with college and advanced degrees tend to closely align, suggesting similar survival trends. Conversely, women with a high school diploma tend to have a slightly lower curve, and those without a degree display an even lower one. This notable and statistically significant gap (with a log-rank test p-value < 0.001) underscores a substantial difference in survival between women without a degree and those with a higher level of education.

Figure 4: Kaplan-Meier survival estimator by education level for females.



A similar hierarchy existed for men, where overall survival probability increased with higher levels of education, as depicted in Figure 5. However, men with an advanced degree exhibited a significantly higher survival probability at all ages, with a median survival age of 90 years. The other three groups differed in median survival age by one year each, resulting in a median survival age of 88 for men with a college degree, 87 for men with a high school diploma, and 86 for those without any degree. The differences between these groups were statistically significant based on the log-rank test (p-value <0.001).

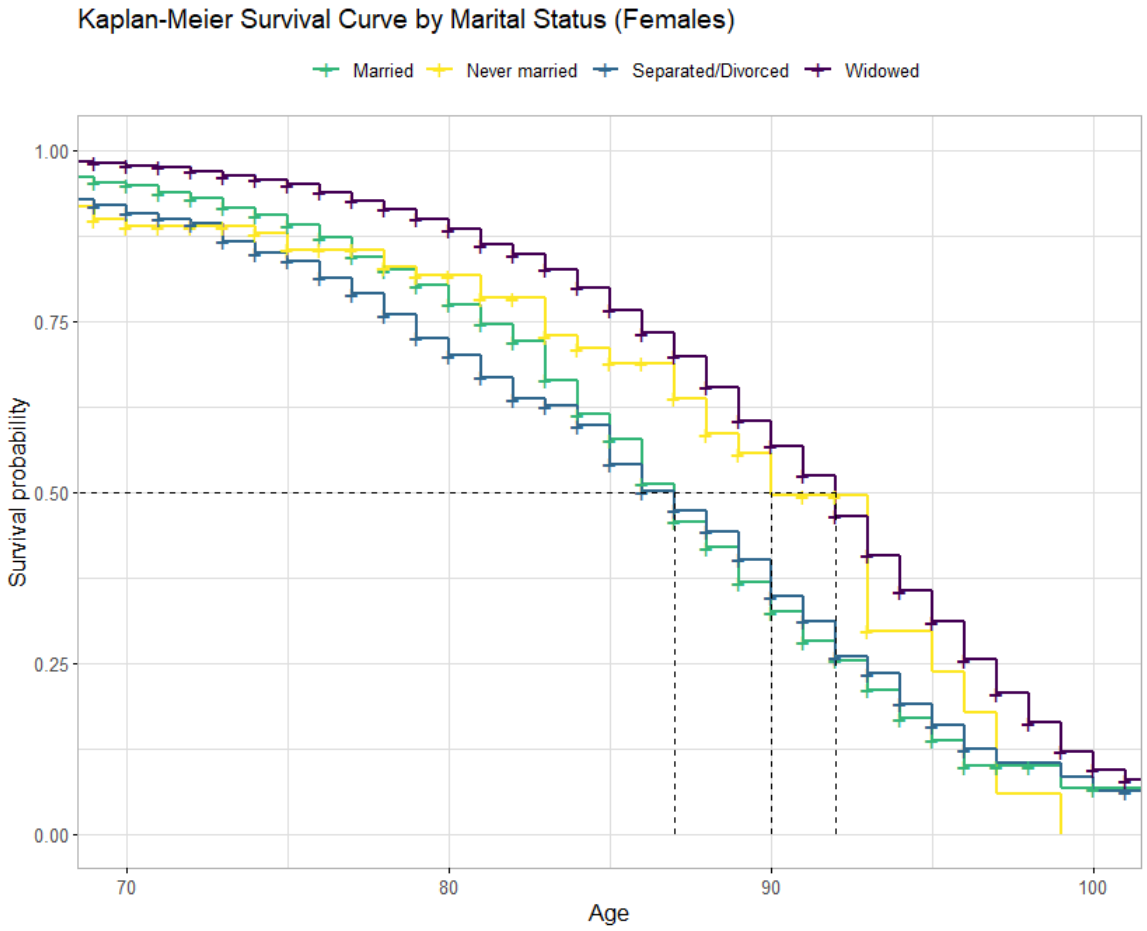
Figure 5: Kaplan-Meier survival estimator by education level for males.

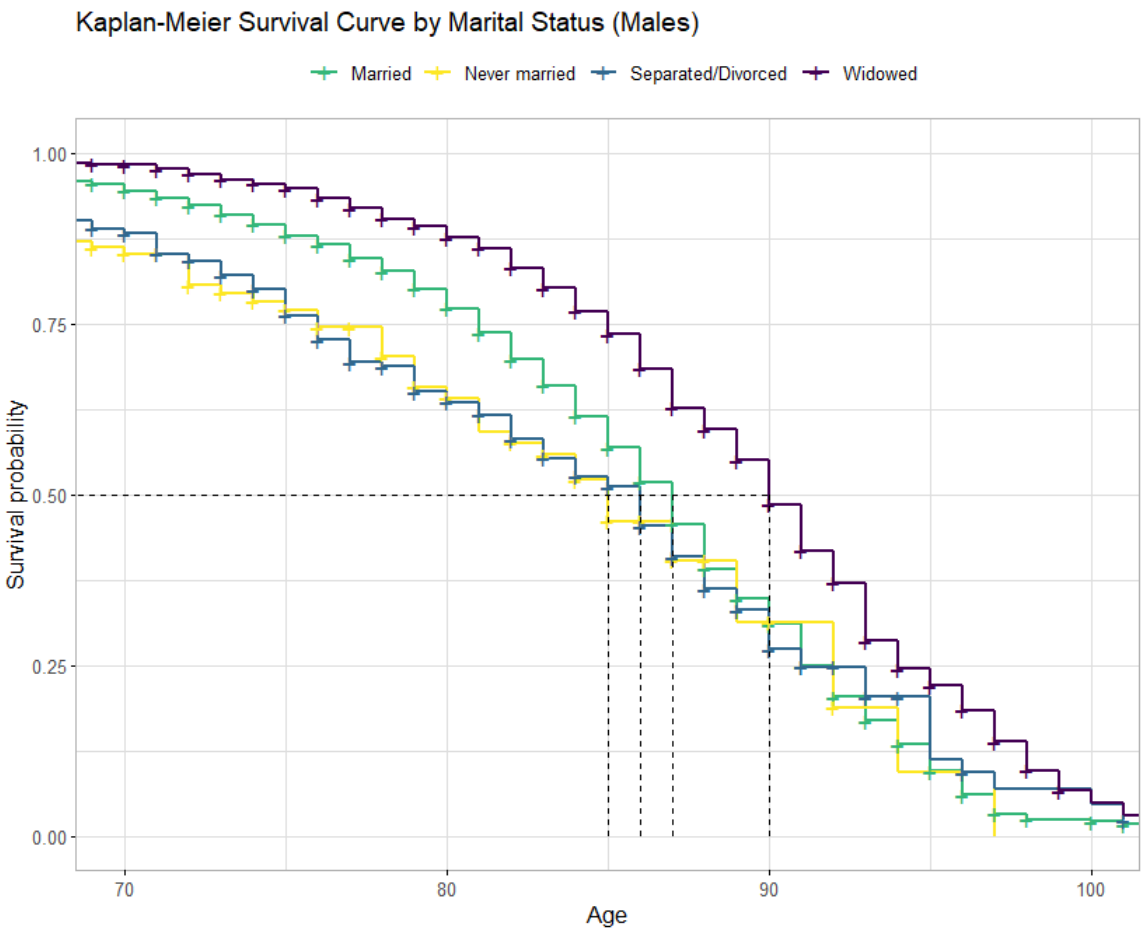


The results obtained by grouping respondents based on their marital status also revealed clear differences among categories for both sexes, as shown one above the other in Figure 6 to facilitate visual comparison. Particularly, widowed women exhibited the highest survival probability at all ages, with a median survival age of 92 years, two years higher than women who had never been married, who also had a relatively high

survival probability. Both married and separated or divorced women exhibited a lower survival probability at all ages, with a median survival age of 87 years. With the exception of the advantage of widowed individuals in terms of years of life, which was also evident for men, with a median survival age of 90 years, the trajectories for men differed. In fact, while not getting married appeared to be a protective factor for women, men who got married demonstrated higher survival probabilities, with a median survival age of 87 years, compared to their counterparts who had never been married or who were separated or divorced, with respective median survival ages of 85 and 86 years. It was noteworthy that differences in marital status among individuals of different genders were significant, with a p-value < 0.001 based on the log-rank test.

Figure 6: Kaplan-Meier survival estimator by marital status for females and males.

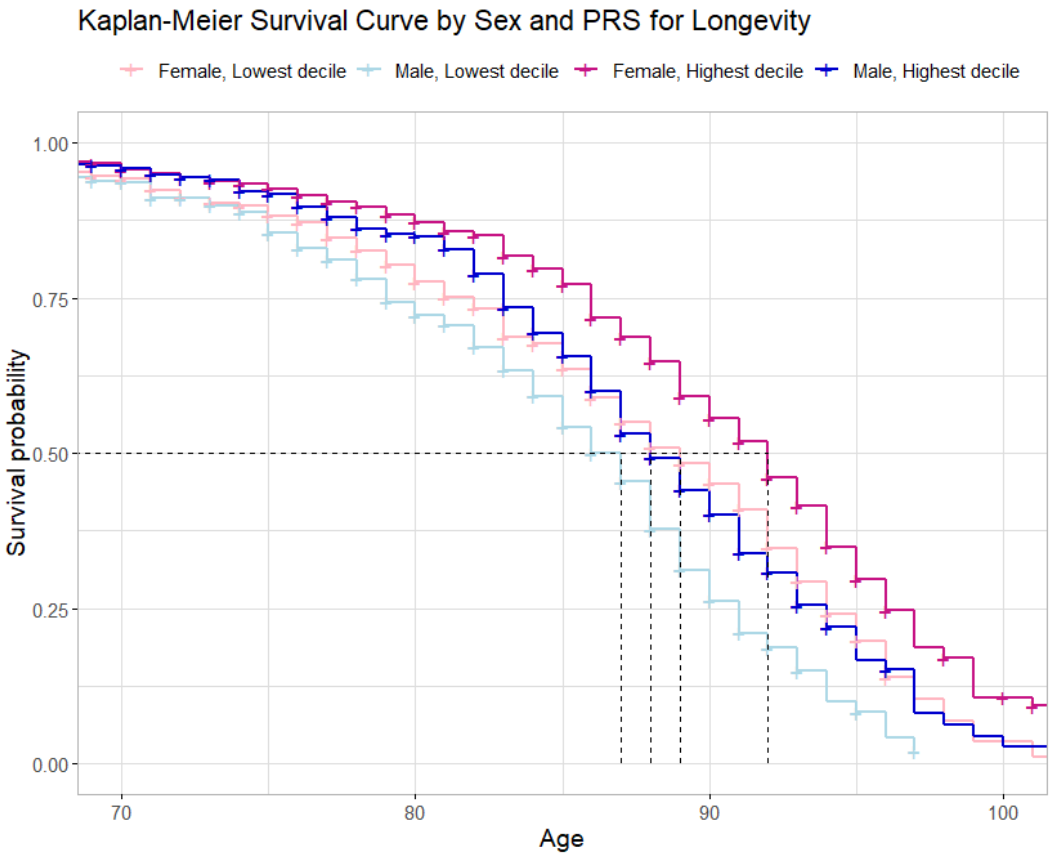


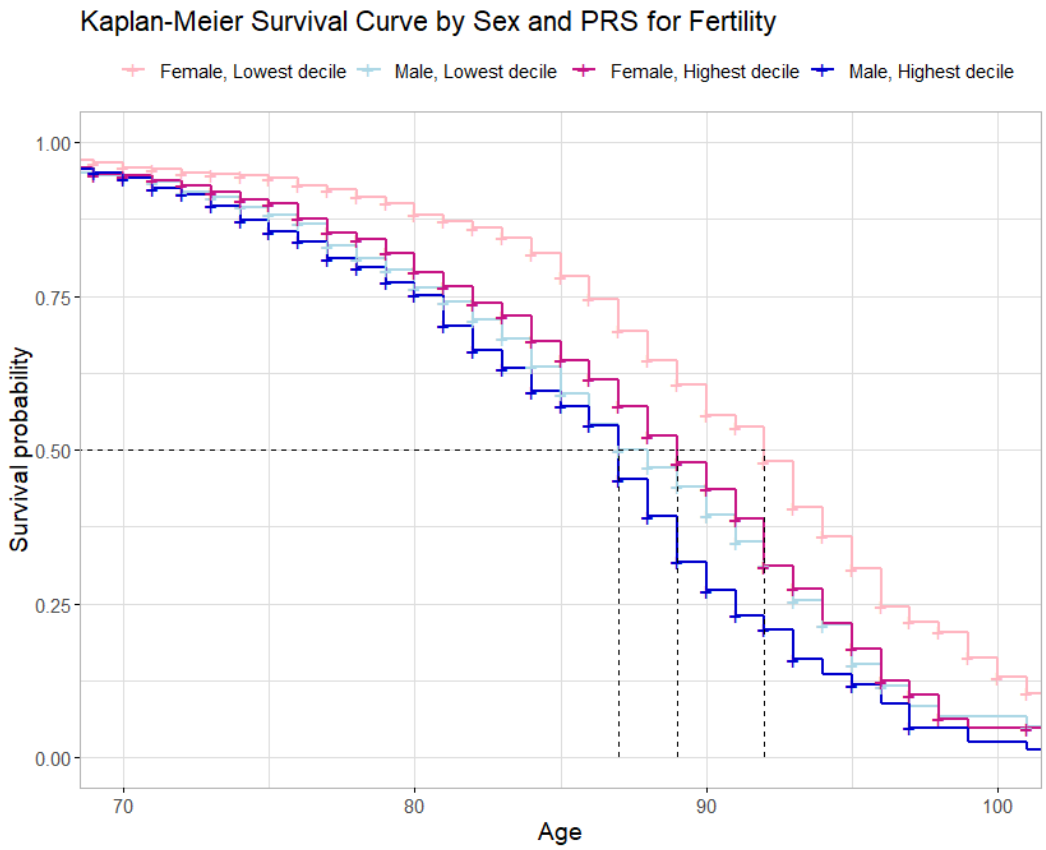


Finally, the effect of genetic predisposition for longevity and fertility was visualized in Figure 7 by comparing the survival curves of individuals in the sample who respectively had a PRS in the highest decile of the distribution, represented by a darker tone, or in the lowest decile, represented by a lighter tone. When comparing the two plots with a focus on women, it became evident that those with a PRS for longevity in the highest decile experienced advantages in terms of survival, with a median survival age of 92 years, which was 3 years higher than those in the lowest decile of the distribution. However, concerning women's distribution of the PRS for fertility, the situation was reversed, with women who were genetically predisposed to have more children exhibiting lower survival compared to those in the lowest decile of the PRS for fertility distribution. For men, the difference between being more or less genetically predisposed to longevity did not constitute a strong protective factor, as it added only one year to the median survival age, shifting it from 87 years for those in the lowest decile of the distribution to 88 years for those in the highest decile. Regarding the PRS for fertility, the median survival age remained the same at 87 years for both men in the

highest and lowest PRS deciles of the distribution, with only a slight difference between the two curves.

Figure 7: Kaplan-Meier survival estimator for males and females with a PRS respectively in the lowest or highest decile of the distribution of the PRS of longevity and of fertility.





Model Estimation

To simultaneously assess the effects of all predictors, the next step involved fitting the three previously defined Cox proportional hazards models. The first Cox model included only the number of children as a covariate, with the group having no children as the reference category. Consequently, the coefficients were interpreted in comparison to individuals with no children. The results of the models are summarized by sex in the following two tables. For each group, the summary includes the estimated coefficient (β) and its standard error, the corresponding hazard ratio (obtained through exponentiation) along with its 95% confidence interval, and the values of z and p-values from the Wald statistic. The Wald statistic assesses whether the beta coefficient of a specific variable significantly differs from zero.

The summary of the results from the first Cox model for women in Table 2 indicates that the three overall tests suggest a significant effect of the number of children

on mortality. The sign of the estimated coefficients and their respective hazard ratios reveal that, compared to those with no children, individuals with one child have a slightly higher mortality risk, although the difference is not significant. Conversely, having two or three children is a significantly protective factor that reduces the mortality risk by 17% and 19%, respectively, compared to women with no children. Lastly, having more than three children does not significantly differ from not having children in terms of mortality risk.

Table 2: Results of the Cox model for women with only the number of children as covariate.

Variable	**Coefficient (Std. Error)**	**HR (95% CI)**	**z**	**p-value**
1 Child	0.011 (0.092)	1.011 (0.845 − 1.211)	0.121	0.903
2 Children	−0.162 (0.080)	0.851 (0.727 − 0.996)	−2.011	0.044 *
3 Children	−0.176 (0.082)	0.839 (0.714 − 0.984)	−2.153	0.031 *
More than 3 children	−0.001 (0.078)	0.999 (0.857 − 1.164)	−0.015	0.988
Likelihood ratio test				0.003
Wald test				0.003
Score logrank test				0.003

Table 3 presents the results of the model fitted for men. However, the number of children does not emerge as a significant predictor of mortality risk. In fact, none of the coefficients is statistically significant, even though having children appears to have a protective effect, slightly reducing the risk of death, particularly for men with two children. The three overall tests also do not yield significance, indicating that this predictor does not significantly influence mortality. The proportional hazard assumption, for both the models for females and males, was confirmed by testing if the Schoenfeld residuals remained independent of time and through graphical diagnostics.

Table 3: Results of the Cox model for men with only the number of children as covariate.

Variable	**Coefficient (Std. Error)**	**HR (95% CI)**	**z**	**p-value**
1 Child	-0.070 (0.097)	0.933 $(0.881 - 1.129)$	0.474	0.474
2 Children	-0.118 (0.081)	0.889 $(0.759 - 1.041)$	0.145	0.145
3 Children	-0.110 (0.083)	0.896 $(0.762 - 1.054)$	0.186	0.186
More than 3 children	-0.110 (0.082)	0.896 $(0.763 - 1.052)$	0.179	0.179
Likelihood ratio test				0.6
Wald test				0.6
Score logrank test				0.6

For both the model for females and males, the proportional hazard assumption was confirmed by testing if the Schoenfeld residuals were independent of time and also by using graphical diagnostics.

Since the number of children did not prove to be a significant predictor for men's longevity, a second model was fitted, adjusting the estimates by adding some demographic variables, separately for females and males. All three tests for both models had a p-value <0.001, confirming the overall model significance. However, the proportional hazard assumption did not hold perfectly for all variables, as the effect of some was not independent of time. Recognizing this limitation, we present and interpret the results of the two models, with a more in-depth discussion in the next chapter to evaluate the estimated direction and magnitude of the effect of each predictor based on previous evidence.

Table 4 demonstrates that after controlling for three demographic variables (year of birth, marital status, and education level), the protective effect of having two or three children on women's survival compared to those with no children is upheld. Specifically, with the same demographic variables, women with two children have a mortality risk 17% lower than women without children, while women with three children have a mortality risk 20% lower. As in the previous simpler model, the difference between women without children and those with one or more than three children is not significant. As expected from the Kaplan-Meier survival curves, having a higher level of education significantly enhances survival. Women with no degree face a 35% higher mortality risk than those with a high school diploma, and those with a college degree have a 35% lower mortality risk than high school graduates. Finally, women with an advanced level of education experience a mortality risk more than 50% lower than those with a high school diploma. The differences in survival among individuals with different education levels are highly significant. Regarding marital status, women who never married have a similar mortality risk to married women, while separated or divorced ones exhibit a 15% higher mortality risk at all ages compared to those who were married. Widowed women, on the other hand, show a 30% difference in mortality risk from those whose partner is still alive.

Table 4: Results of the Cox model for women with the number of children and other demographic variables as predictors.

Variable	Coefficient (Std. Error)	HR (95% CI)	z	p-value
1 Child	0.068 (0.096)	1.071 (0.888 − 1.292)	0.714	0.475
2 Children	−0.187 (0.085)	0.829 (0.702 − 0.980)	−2.194	0.028 *
3 Children	−0.225 (0.087)	0.799 (0.674 − 0.947)	−2.587	0.009 **
More than 3 children	−0.016 (0.084)	0.984 (0.835 − 0.160)	−0.193	0.847
Year of birth	0.116 (0.004)	1.123 (1.113 − 1.132)	26.590	< 0.001 ***
Advanced degree	−0.413 (0.106)	0.662 (0.538 − 0.815)	−3.887	< 0.001 ***
College degree	−0.297 (0.065)	0.743 (0.654 − 0.844)	−4.587	< 0.001 ***
No degree	0.302 (0.056)	1.353 (1.213 − 1.509)	5.429	< 0.001 ***
Never married	−0.094 (0.164)	0.910 (0.660 − 1.257)	−0.571	0.568

Separated/Divorced	0.142	1.152	1.911	0.056
	(0.074)	(0.996 − 1.332)		
Widowed	−0.267	0.765	−4.891	< 0.001
	(0.055)	(0.688 − 0.852)		***

The results for males are presented in Table 5, with fewer variables emerging as significant predictors of longevity. Specifically, the number of children was confirmed to have a non-significant impact on survival. In terms of education, the effect is similar to that observed for women in terms of direction. Men with an advanced degree have a 78% lower mortality risk than those with a high school diploma, while those with a college degree have a 27% lower mortality risk than high school graduates. This indicates that, for men, the gain in terms of survival for those with an advanced degree is greater than for women. Men without any degree, on the other hand, face a 37% higher mortality risk at any age compared to those with a high school diploma. Additionally, marital status has a weaker influence on men's survival functions, with only widowed men showing a significant difference from married men, having a 34% lower mortality risk, all else being equal.

Table 5: Results of the Cox model for men with the number of children and other demographic variables as predictors.

Variable	**Coefficient (Std. Error)**	**HR (95% CI)**	**z**	**p-value**
1 Child	0.063	1.065	0.607	0.544
	(0.104)	(0.869 − 1.305)		
2 Children	−0.090	0.914	−1.014	0.311
	(0.089)	(0.767 − 1.088)		

3 Children	-0.048	0.953	-0.528	0.598
	(0.091)	$(0.797 - 1.140)$		
More than 3 children	-0.051	0.950	-0.564	0.573
	(0.090)	$(0.796 - 0.134)$		
Year of birth	0.109	1.115	26.114	< 0.001
	(0.004)	$(1.105 - 1.124)$		***
Advanced degree	-0.574	0.563	-7.234	< 0.001
	(0.079)	$(0.482 - 0.658)$		***
College degree	-0.237	0.789	-3.861	< 0.001
	(0.061)	$(0.700 - 0.890)$		***
No degree	0.316	1.372	5.307	< 0.001
	(0.060)	$(1.221 - 1.542)$		***
Never married	0.146	0.158	0.961	0.336
	(0.152)	$(0.859 - 1.561)$		
Separated/Divorced	0.119	1.126	1.571	0.116
	(0.076)	$(0.971 - 1.301)$		
Widowed	-0.297	0.743	-5.332	< 0.001
	(0.056)	$(0.667 - 0.829)$		***

A third model was separately fitted for females and males, incorporating genetic information from polygenic risk scores for longevity and fertility while adjusting for the top ten genetic principal components. Similar to the previous model, the overall significance was confirmed with a p-value < 0.001 in the likelihood ratio test, the Wald

test, and the score log-rank test. However, the proportional hazards assumption was rejected for some covariates in these models but confirmed for the polygenic risk scores, as their effect remained independent of age. While coefficients were estimated for each of the ten genetic principal components, they are not reported in the following tables for brevity, as they were included to capture any remaining genetic noise.

Table 6 presents results for women, which closely resemble those in Table 4 for the model without genetic predictors. The effects of demographic predictors, education attainment, and marital status exhibit the same strength and direction as in the second model. Notably, mortality risk was inversely correlated with education level. Among marital status groups, married women did not display significant differences from women who never married, while separated or divorced women showed a higher mortality risk. Widowed women, on the other hand, had a survival advantage compared to women whose partners were alive. Regarding the effect of having children on longevity, the third model aligns with the previous one. When holding all other predictors constant, women with no children or more than three children did not exhibit significant differences in mortality risk compared to those without children. Instead, the protective effect of having two or three children versus none remained significant, with a respective widening of the difference in mortality risk at every age by 23% and 28%. For women, genetic potential for fertility expressed through the PRS for fertility was significantly associated with mortality risk at a significance level of 10%, with a 4% increase in mortality risk for each additional unit in the polygenic score, implying that women biologically inclined to have more children might have shorter lives. Conversely, women with a higher PRS for longevity tended to have lower mortality risk at all ages, with a 6% decrease in mortality risk for each additional unit in the polygenic score.

Table 6: Results of the Cox model for women with the number of children, demographic variables, and polygenic risk scores for longevity and fertility as predictors.

Variable	Coefficient (Std. Error)	HR (95% CI)	z	p-value
1 Child	0.059 (0.096)	1.061 (0.878 − 1.281)	0.613	0.540
2 Children	−0.205 (0.086)	0.815 (0.688 − 0.964)	−2.381	0.017 *
3 Children	−0.250 (0.089)	0.779 (0.655 − 0.926)	−2.823	0.005 **
More than 3 children	−0.065 (0.087)	0.937 (0.791 − 1.111)	−0.745	0.456
Year of birth	0.115 (0.004)	1.121 (1.112 − 1.131)	26.428	< 0.001 ***
Advanced degree	−0.430 (0.107)	0.651 (0.528 − 0.802)	−4.033	< 0.001 ***
College degree	−0.307 (0.065)	0.735 (0.647 − 0.836)	−4.720	< 0.001 ***
No degree	0.304 (0.056)	1.355 (1.215 − 1.512)	5.443	< 0.001 ***
Never married	−0.027 (0.165)	0.973 (0.704 − 1.344)	−0.165	0.869

Separated/Divorced	0.149 (0.074)	1.161 (1.003 − 1.342)	2.003	0.045 *
Widowed	−0.269 (0.055)	0.764 (0.686 − 0.851)	−4.919	< 0.001 ***
PRS of longevity	−0.058 (0.026)	0.944 (0.897 − 0.993)	−2.215	0.027 *
PRS of fertility	0.040 (0.024)	1.041 (0.961 − 0.993)	1.651	0.099 .

After controlling for genetic predisposition for longevity and fertility, even though neither had a significant impact on men's mortality risk, the effect the number of children also disappeared. However, the signs of the estimated coefficients, as shown in Table 7, remained consistent across the three models, confirming that men with children might have a survival advantage compared to childless men. The inclusion of genetic covariates did not alter the direction or magnitude of the effects of education attainment and widowhood, both of which were still negatively associated with mortality risk.

Table 7: Results of the Cox model for men with the number of children, demographic variables, and polygenic risk scores for longevity and fertility.

Variable	**Coefficient (Std. Error)**	**HR (95% CI)**	**z**	**p-value**
1 Child	0.068 (0.105)	1.071 (0.872 − 1.314)	0.652	0.514
2 Children	−0.080 (0.090)	0.924 (0.774 − 1.102)	−0.882	0.377

3 Children	−0.057	0.945	−0.614	0.539
	(0.093)	(0.788 − 1.133)		
More than 3 children	−0.083	0.920	−0.881	0.378
	(0.094)	(0.765 − 1.107)		
Year of birth	0.108	1.114	24.836	< 0.001
	(0.004)	(1.104 − 1.123)		***
Advanced degree	−0.579	0.560	−7.270	< 0.001
	(0.080)	(0.479 − 0.655)		***
College degree	−0.240	0.786	−3.901	< 0.001
	(0.062)	(0.697 − 0.887)		***
No degree	0.311	1.365	5.186	< 0.001
	(0.060)	(1.214 − 1.536)		***
Never married	0.157	0.170	1.022	0.307
	(0.153)	(0.866 − 1.579)		
Separated/Divorced	0.117	1.124	1.532	0.126
	(0.076)	(0.968 − 1.305)		
Widowed	−0.292	0.747	−5.236	< 0.001
	(0.056)	(0.669 − 0.833)		***
PRS of longevity	−0.002	0.998	−0.072	0.943
	(0.028)	(0.944 − 1.055)		
PRS of fertility	0.038	1.039	−1.468	0.142
	(0.026)	(0.987 − 1.093)		

Chapter 5 - Discussion

The number of children, education level, marital status, and genetic factors all exhibited varying effects on survival, with differences between men and women. Women with two or three children had a lower mortality risk, higher education was associated with improved survival, and being widowed was linked to a longer life. Genetic factors also played a role, with higher polygenic scores for longevity associated with lower mortality risk for women. For men, the impact of having children on survival was inconclusive, and genetic factors did not significantly influence their survival, suggesting that their longevity might be influenced by different factors.

This chapter aims to discuss and compare these findings with previous research, identify limitations and potential sources of bias in the study, and explore the implications of these findings in the context of public health and personalized healthcare.

Interpretation of Findings with Previous Research

The results of this study align with previous research on the Fertility-Longevity Trade-Off. Evolutionary theories, as proposed by Kirkwood and Williams, suggested that individuals with more children may experience a shorter lifespan due to the allocation of limited resources toward reproduction. The findings of this study corroborate this theory, particularly in the case of women, as those with two or three children displayed a significantly lower mortality risk (Kirkwood, 1977; Williams, 1957).

Kirkwood's theory, which introduced the concept of the "disposable soma," posits that an organism's resources are limited and must be allocated between reproduction and somatic maintenance. In essence, more offspring come at the expense of individual longevity. Williams' theory of "antagonistic pleiotropy" highlights the idea that genes promoting fertility early in life may have detrimental effects on longevity. These

foundational theories continue to influence research into the Fertility-Longevity Trade-Off.

The non-linear relationship between fertility and longevity, as suggested by Ehrlich and Kuningas et al., was also observed in this study. The U-shaped association found in the analysis emphasizes that having a limited number of children may not dramatically affect lifespan, but each additional child progressively consumes energy and resources, impacting longevity (Ehrlich, 2015; Kuningas et al., 2011).

Ehrlich's work on the non-linear relationship between fertility and longevity has gained prominence in recent years. He proposed that the trade-off is not linear and may be better described as a curve, with a subtle decline in longevity with an increasing number of children, which becomes steeper as the number of offspring grows. Kuningas et al. further emphasized this non-linear relationship, highlighting that a limited number of children may have little impact on lifespan but that each additional child exacts a progressively greater toll on the individual's resources, ultimately impacting their longevity.

The role of genetic factors in longevity, as outlined in this study, has been previously explored. Various genes, including APOE, FOXO3A, SIRT1, and IGF-1, have been implicated in affecting both fertility and longevity. The study's observation that higher polygenic scores for longevity were associated with lower mortality risk in women supports previous research indicating that genetics plays a substantial role in determining lifespan. The APOE gene, with its ε2 and ε4 variants, has been extensively studied for its influence on both fertility and longevity. Research highlights that the ε2 variant is associated with increased longevity at fertility expenses, while the ε4 variant is linked to a shorter lifespan but higher fertility. FOXO3A, a gene protecting cells from damage, has been investigated for its effects on fertility and longevity, with mutations in this gene resulting in lower fertility (Bao et al., 2014). SIRT1, a gene influencing metabolism regulation, has been associated with decreased fertility and longevity (Kilic et al., 2015). The IGF-1 gene has been linked to male fertility but may also increase the risk of cancer and age-related diseases. However, it is worth noting that the inconclusive

findings regarding genetic factors in men highlight the complexity of genetic influences on longevity and emphasize that this area of research is far from fully understood (Corbo et al., Atzmon et al., Bao et al., Kilic et al., Rodriguez et al.). The complexity of genetic interactions and their influence on longevity and fertility has also been highlighted in other studies. Chabris et al. (2015) emphasized that these polygenic traits are influenced by both genetics and the environment, with genetic endowment playing a substantial role in shaping individual outcomes.

Additionally, the influence of educational attainment on survival, echoes previous findings (Christenson and Johnson, 1995) where it was associated with lower mortality risk through better pulse control, risk understanding, and economic positions, which ultimately influence longevity.

The unexpected finding regarding widowed individuals in this study challenges the conventional wisdom that being in a relationship is associated with higher life expectancy. Jia and Lubetkin (2020) and Balter et al. (2023) have suggested that partnership status, such as marriage or cohabitation, is linked to increased life expectancy. The supportive and collaborative nature of partnerships is believed to contribute to better health and, consequently, greater longevity. However, for widowed individuals in this study, the experience of loss and adaptation to a new life situation might have introduced a selection bias, where only the individuals with greater resilience and resources could maintain their health and longevity after the loss of a partner. Further research is necessary to better understand the factors influencing the longevity of widowed individuals.

Limitations and Potential Sources of Bias

While the results provide valuable insights, it is crucial to acknowledge the study's limitations and potential sources of bias. These considerations are essential for interpreting the findings accurately and for guiding future research in this field.

First and foremost, the study's sample was drawn from a specific time frame, encompassing individuals born between 1905 and 1980. This temporal limitation may impact the generalizability of the findings. The characteristics and behaviours of individuals from this time period may differ from those in more recent generations, making it essential to exercise caution when applying the study's conclusions to modern populations.

Furthermore, the study's design could introduce selection bias. The participants who were still alive at the last wave of interviews were more likely to possess specific characteristics or genetic profiles that influenced their survival. This potential bias may skew the results, as individuals with particular genetic predispositions or socioeconomic advantages might be overrepresented in the sample. Consequently, the observed associations between demographic and genetic factors and survival may not accurately reflect the broader population.

Additionally, the genetic component of the study relied on polygenic scores constructed from genome-wide data. While polygenic scores are valuable tools for assessing genetic influences, they are subject to potential biases in the selection of SNPs and their weighting. The process of selecting SNPs and assigning weights to them can introduce subjectivity and lead to uncertainties in the final polygenic scores. Furthermore, the field of genetic research is rapidly evolving, and future studies may benefit from more comprehensive genomic data. As new genetic insights emerge, the understanding of the genetic factors influencing longevity may evolve and refine the interpretations of this study.

Implications for Public Health and Personalized Healthcare

The findings of this study have significant implications for public health and personalized health, with a particular focus on precision medicine. Understanding the complex interplay of demographic and genetic factors in longevity can inform public health interventions and healthcare practices.

For public health, the recognition that educational attainment plays a crucial role in longevity highlights the importance of improving access to education and health knowledge. These findings suggest that interventions targeting education and health awareness may contribute to increased lifespan. The impact of education on health and longevity has been previously highlighted. Efforts to enhance public health can include policies and programs that promote education and health literacy. Access to quality education and the dissemination of health information can empower individuals to make informed decisions about their health and well-being. For example, health campaigns can focus on raising awareness about the importance of education and its potential impact on longevity, encouraging individuals to pursue educational opportunities that may contribute to a longer and healthier life.

In the field of personalized health, the study underscores the potential of genetic information in predicting and understanding individual longevity. Precision medicine, a rapidly advancing field, can utilize genetic information to tailor healthcare interventions to an individual's unique genetic makeup. This approach allows healthcare providers to develop personalized prevention and treatment plans that consider an individual's genetic predispositions and potential health risks. While the genetic factors influencing longevity are complex and not fully elucidated, this area holds promise for tailoring healthcare and lifestyle recommendations based on an individual's genetic profile. For instance, individuals with genetic variants associated with longevity may receive personalized guidance on lifestyle choices and preventive measures to maximize their lifespan. Moreover, precision medicine can extend to lifestyle recommendations based on genetic insights. Individuals may receive personalized guidance on diet, exercise, and other lifestyle factors that align with their genetic profile, optimizing their chances of a longer, healthier life. By harnessing genetic information, healthcare can transition from a one-size-fits-all model to a more tailored approach that addresses the specific needs and risks of each individual.

The results of this study shed light on the intricate relationship between demographic and genetic factors and their impact on human longevity. They align with previous research on the Fertility-Longevity Trade-Off and provide further insights into

its non-linear nature. The differences observed between men and women underscore the need for gender-specific research in this field.

As the implications of these findings for public health and personalized health are explored, it becomes evident that the complex interplay of factors influencing longevity requires further investigation. Continued research in this area, with attention to the limitations and potential sources of bias, can contribute to a deeper understanding of the determinants of human lifespan and inform strategies for improving overall health and well-being. While the research is still evolving, these findings offer valuable insights into the multi-faceted nature of human longevity, with a particular emphasis on the potential of precision medicine to improve health outcomes.

Chapter 6 - Conclusion

This study embarked on an in-depth exploration of human longevity, delving into the complex interplay between factors such as the number of children, educational attainment, marital status, and genetics. It aimed to illuminate the intricate nature of human lifespan, examining the underlying trade-offs and synergies. Central to the research was a re-evaluation of the Fertility-Longevity Trade-Off through the lens of evolutionary theories, revealing a non-linear relationship characterized by a U-shaped association between fertility and longevity. This finding adds depth to the traditional understanding of the trade-off, highlighting the intricate balance between reproduction and individual survival.

Gender differences emerged as a significant aspect, with women showing a lower mortality risk at specific levels of fertility, contrasting with the more ambiguous results for men. This emphasizes the need for further investigation into the biological and socio-cultural factors that differentiate the trade-off in men and women.

The study also highlighted the role of education in influencing longevity. Higher education levels were consistently linked to lower mortality risks, supporting previous studies that underlined the positive impacts of education on health and economic well-being. This finding reaffirms the importance of policies that promote education and health literacy as pathways to enhance lifespan.

In terms of marital status, the research uncovered surprising correlations. Contrary to the expectation that being in a relationship correlates with higher life expectancy, it was found that widowed individuals often had longer lifespans. This unexpected correlation calls for a deeper examination of the resilience and social dynamics of widowed individuals.

The genetic aspect of the study provided a deeper understanding of the factors influencing human longevity. The analysis of polygenic scores highlighted specific genetic factors associated with lower mortality risks, particularly in women, while the

genetic influences on male longevity presented a less clear picture, signalling an area ripe for gender-specific genetic research.

The contributions of this research extend the current understanding of human longevity. By challenging established notions and providing fresh insights into the Fertility-Longevity Trade-Off, the role of education, marital status, and genetics, the study paves the way for a more detailed comprehension of the factors that influence lifespan.

Future Research Directions

Looking forward, the study opens several avenues for further research. These include in-depth genetic investigations to understand the specific genes and variants influencing lifespan, with a focus on gender-specific genetic pathways. Exploring the genetic predispositions in relation to lifestyle factors like diet, exercise, and stress management could also provide valuable insights into how these interactions affect longevity.

The unexpected findings related to marital status warrant a deeper exploration into the psychosocial and support mechanisms for widowed individuals. Studies could examine how social support networks, mental health, and lifestyle changes post-widowhood contribute to longevity. Additionally, investigating the impact of different forms of social relationships, beyond marital status, could provide a more holistic view of the social determinants of longevity.

Furthermore, the integration of genetic data into healthcare practices could lead to more personalized approaches to health and longevity. Research could explore the development of targeted healthcare interventions based on individual genetic profiles, including preventive strategies and treatments for age-related diseases.

Environmental factors and their interaction with genetic predispositions also present an interesting area for future research. Understanding how environmental

exposures, such as pollution and climate change, interact with genetics to influence longevity could lead to comprehensive strategies for promoting health and longevity in different environmental contexts.

This research sets a foundation for future studies to build upon, enhancing our understanding of the complex tapestry of factors that determine human longevity. The pursuit of this knowledge has the potential to significantly impact public health strategies and improve quality of life as populations age.

References

Arias, E., & Xu, J. Q. (2020). United States life tables, 2018. National Vital Statistics Reports, 69(12). Hyattsville, MD: National Center for Health Statistics.

Arias, E. (2002). United States life tables, 2000. National Vital Statistics Reports, 51(3). Hyattsville, MD: National Center for Health Statistics.

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. https://doi.org/10.1046/j.1532-5415.2002.49283.x

Health and Retirement Study. (2023). About the Health and Retirement Study. Retrieved from https://hrs.isr.umich.edu/about/

Kirkwood, T. B. (1977). Evolution of ageing. Nature, 270(5635), 301-304. https://doi.org/10.1038/270301a0

Williams, G. C. (1957). Pleiotropy, natural selection, and the evolution of senescence. Evolution, 11(4), 396-411.

Kuningas, M., Altmäe, S., Uitterlinden, A. G., Hofman, A., van Duijn, C. M., & Tiemeier, H. (2011). The relationship between fertility and lifespan in humans. Age, 33(4), 615-622. https://doi.org/10.1007/s11357-010-9202-4

Gavrilova, N. S., Gavrilov, L. A., Semyonova, V. G., & Evdokushkina, G. N. (2004). Does exceptional human longevity come with a high cost of infertility? Annals of the New York Academy of Sciences, 1019, 513-517. https://doi.org/10.1196/annals.1297.095

Gagnon, A., Smith, K. R., Tremblay, M., Vézina, H., Paré, P. P., & Desjardins, B. (2009). Is there a trade-off between fertility and longevity? A comparative study of women from three large historical databases accounting for mortality selection.

American Journal of Human Biology, 21(4), 533-540. https://doi.org/10.1002/ajhb.20893

Kaptijn, R., Thomese, F., Liefbroer, A. C., Van Poppel, F., Van Bodegom, D., & Westendorp, R. G. (2015). The Trade-Off between Female Fertility and Longevity during the Epidemiological Transition in the Netherlands. PLoS One, 10(12), e0144353. https://doi.org/10.1371/journal.pone.0144353

Hsu, C. H., Posegga, O., Fischbach, K., & Engelhardt, H. (2021). Examining the trade-offs between human fertility and longevity over three centuries using crowdsourced genealogy data. PLoS One, 16(8), e0255528. https://doi.org/10.1371/journal.pone.0255528

Ehrlich, S. (2015). Effect of fertility and infertility on longevity. Fertility and Sterility, 103(5), 1129-1135. https://doi.org/10.1016/j.fertnstert.2015.03.021

Most, J., Dervis, S., Haman, F., Adamo, K. B., & Redman, L. M. (2019). Energy Intake Requirements in Pregnancy. Nutrients, 11(8), 1812. https://doi.org/10.3390/nu11081812

Bongaarts, J., & Watkins, S. C. (1996). Social Interactions and Contemporary Fertility Transitions. Population and Development Review, 22(4), 639-682. https://doi.org/10.2307/2137804

Bongaarts, J., & Hodgson, D. (2022). Socio-Economic Determinants of Fertility. In Fertility Transition in the Developing World (SpringerBriefs in Population Studies). Springer, Cham. https://doi.org/10.1007/978-3-031-11840-1_4

Christenson, B. A., & Johnson, N. E. (1995). Educational inequality in adult mortality: an assessment with death certificate data from Michigan. Demography, 32(2), 215-229.

Kenkel, D. S. (1991). Health Behavior, Health Knowledge, and Schooling. Journal of Political Economy, 99(2), 287-305. https://doi.org/10.1086/261751

Cutler, D. M., & Lleras-Muney, A. (2010). Understanding differences in health behaviors by education. Journal of Health Economics, 29(1), 1-28. https://doi.org/10.1016/j.jhealeco.2009.10.003

Seeman, T. E., Crimmins, E., Huang, M. H., Singer, B., Bucur, A., Gruenewald, T., Berkman, L. F., & Reuben, D. B. (2004). Cumulative biological risk and socio-economic differences in mortality: MacArthur studies of successful aging. Social Science & Medicine, 58(10), 1985-1997. https://doi.org/10.1016/S0277-9536(03)00402-7

Shadyab, A. H., Gass, M. L. S., Stefanick, M. L., Waring, M. E., Macera, C. A., Gallo, L. C., Shaffer, R. A., Jain, S., & LaCroix, A. Z. (2017). Maternal Age at Childbirth and Parity as Predictors of Longevity Among Women in the United States: The Women's Health Initiative. American Journal of Public Health, 107, 113-119. https://doi.org/10.2105/AJPH.2016.303503

Balter, A., Bjerre, D., & Kallestrup-Lamb, M. (2023). The effect of marital status on life expectancy: Is cohabitation as protective as marriage? Journal of Demographic Economics, 89(3), 373-394. https://doi.org/10.1017/dem.2023.10

Moon, J. R., Glymour, M. M., Vable, A. M., Liu, S. Y., Subramanian, S. V. (2014). Short- and long-term associations between widowhood and mortality in the United States: longitudinal analyses. Journal of Public Health, 36(3), 382-389. https://doi.org/10.1093/pubmed/fdt101

Jia, H., & Lubetkin, E. I. (2020). Life expectancy and active life expectancy by marital status among older U.S. adults: Results from the U.S. Medicare Health Outcome Survey (HOS). SSM Population Health, 12, 100642. https://doi.org/10.1016/j.ssmph.2020.100642

Balawender, K., & Orkisz, S. (2020). The impact of selected modifiable lifestyle factors on male fertility in the modern world. Central European Journal of Urology, 73(4), 563-568. https://doi.org/10.5173/ceju.2020.1975

Mills, M. C., Tropf, F. C., Brazel, D. M., van Zuydam, N., Vaez, A., eQTLGen Consortium, BIOS Consortium, Human Reproductive Behaviour Consortium, Pers, T. H., Snieder, H., Perry, J. R. B., Ong, K. K., den Hoed, M., Barban, N., Day, F. R. (2021). Identification of 371 genetic variants for age at first sex and birth linked to externalising behaviour. Nature Human Behaviour, 5(12), 1717-1730. https://doi.org/10.1038/s41562-021-01135-3

Reznick, D. N. (2005). The genetic basis of aging: an evolutionary biologist's perspective. Science Aging Knowledge Environment, 2005(11), pe7. https://doi.org/10.1126/sageke.2005.11.pe7

Tesi, N., van der Lee, S. J., Hulsman, M., Jansen, I. E., Stringa, N., van Schoor, N. M., Scheltens, P., van der Flier, W. M., Huisman, M., Reinders, M. J. T., & Holstege, H. (2021). Polygenic Risk Score of Longevity Predicts Longer Survival Across an Age Continuum. The Journals of Gerontology: Series A, 76(5), 750-759. https://doi.org/10.1093/gerona/glaa289

Tazearslan, C., Cho, M., & Suh, Y. (2012). Discovery of functional gene variants associated with human longevity: Opportunities and challenges. The Journals of Gerontology: Series A, 67(4), 376-383. https://doi.org/10.1093/gerona/glr200

Sgrò, C. M., & Partridge, L. (1999). A delayed wave of death from reproduction in Drosophila. Science, 286(5449), 2521-2524. https://doi.org/10.1126/science.286.5449.2521

Medawar, P. B. (1952). An Unsolved Problem of Biology. H. K. Lewis.

Stead, E. R., & Bjedov, I. (2021). Balancing DNA repair to prevent ageing and cancer. Experimental Cell Research, 405(2), 112679. https://doi.org/10.1016/j.yexcr.2021.112679

Bin-Jumah, M. N., Nadeem, M. S., Gilani, S. J., Al-Abbasi, F. A., Ullah, I., Alzarea, S. I., Ghoneim, M. M., Alshehri, S., Uddin, A., Murtaza, B. N., & Kazmi, I.

(2022). Genes and Longevity of Lifespan. International Journal of Molecular Sciences, 23(3), 1499. https://doi.org/10.3390/ijms23031499

Horvath, S. (2013). DNA methylation age of human tissues and cell types. Genome Biology, 14(10), R115. https://doi.org/10.1186/gb-2013-14-10-r115

Harman, D. (1956). Aging: A theory based on free radical and radiation chemistry. Journal of Gerontology, 11(3), 298-300. https://doi.org/10.1093/geronj/11.3.298

Speakman, J. R. (2005). Body size, energy metabolism, and lifespan. Journal of Experimental Biology, 208(Pt 9), 1717-1730. https://doi.org/10.1242/jeb.01556

Corbo, R. M., Pinto, A., & Scacchi, R. (2013). Gender-specific association between FSHR and PPARG common variants and human longevity. Rejuvenation Research, 16(1), 21-27. https://doi.org/10.1089/rej.2012.1365

Atzmon, G., Barzilai, N., Hollowell, J. G., Surks, M. I., & Gabriely, I. (2009). Extreme longevity is associated with increased serum thyrotropin. Journal of Clinical Endocrinology & Metabolism, 94(4), 1251-1254. https://doi.org/10.1210/jc.2008-2325

Sendak, R. A., Sampath, T. K., & McPherson, J. M. (2007). Newly reported roles of thyroid-stimulating hormone and follicle-stimulating hormone in bone remodeling. International Orthopaedics, 31(6), 753-757. https://doi.org/10.1007/s00264-007-0417-7

Kenyon, C. (2011). The first long-lived mutants: Discovery of the insulin/IGF-1 pathway for ageing. Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences, 366(1561), 9-16. https://doi.org/10.1098/rstb.2010.0276

Robinson, A. C., Davidson, Y. S., Roncaroli, F., Minshull, J., Tinkler, P., Horan, M. A., Payton, A., Pendleton, N., & Mann, D. M. A. (2020). Influence of APOE Genotype on Mortality and Cognitive Impairment. Journal of Alzheimer's Disease Reports, 4(1), 281-286. https://doi.org/10.3233/ADR-200203

Raulin, A. C., Doss, S. V., Trottier, Z. A., et al. (2022). ApoE in Alzheimer's disease: Pathophysiology and therapeutic strategies. Molecular Neurodegeneration, 17, 72. https://doi.org/10.1186/s13024-022-00574-4

National Institute of Aging (NIA). (2023). Alzheimer's Disease Genetics Fact Sheet. Retrieved from https://www.nia.nih.gov/health/alzheimers-disease-genetics-fact-sheet

Jasienska, G., Ellison, P. T., Galbarczyk, A., Jasienski, M., Kalemba-Drozdz, M., Kapiszewska, M., Nenko, I., Thune, I., & Ziomkiewicz, A. (2015). Apolipoprotein E (ApoE) polymorphism is related to differences in potential fertility in women: A case of antagonistic pleiotropy? Proceedings of the Royal Society B: Biological Sciences, 282(1803), 20142395. https://doi.org/10.1098/rspb.2014.2395

Mahley, R. W. (2016). Apolipoprotein E: From cardiovascular disease to neurodegenerative disorders. Journal of Molecular Medicine, 94(7), 739-746. https://doi.org/10.1007/s00109-016-1427-y

Corbo, R. M., Ulizzi, L., Scacchi, R., Martínez-Labarga, C., & De Stefano, G. F. (2004). Apolipoprotein E polymorphism and fertility: A study in pre-industrial populations. Molecular Human Reproduction, 10(8), 617-620. https://doi.org/10.1093/molehr/gah082

Deelen, J., Evans, D. S., Arking, D. E., et al. (2019). A meta-analysis of genome-wide association studies identifies multiple longevity genes. Nature Communications, 10(1), 3669. https://doi.org/10.1038/s41467-019-11558-2

Li, J., Chen, Y., Wu, H., & Li, L. (2014). Apolipoprotein E (Apo E) gene polymorphisms and recurrent pregnancy loss: A meta-analysis. Journal of Assisted Reproduction and Genetics, 31(2), 139-148. https://doi.org/10.1007/s10815-013-0128-5

Chen, Y. C., Pohl, G., Wang, T. L., Morin, P. J., Risberg, B., Kristensen, G. B., Yu, A., Davidson, B., & Shih IeM. (2005). Apolipoprotein E is required for cell proliferation and survival in ovarian cancer. Cancer Research, 65(1), 331-337.

Bao, J. M., Song, X. L., Hong, Y. Q., Zhu, H. L., Li, C., Zhang, T., Chen, W., Zhao, S. C., & Chen, Q. (2014). Association between FOXO3A gene polymorphisms and human longevity: A meta-analysis. Asian Journal of Andrology, 16(3), 446-452. https://doi.org/10.4103/1008-682X.123673

Zhang, H., Lin, F., Zhao, J., & Wang, Z. (2020). Expression Regulation and Physiological Role of Transcription Factor FOXO3a During Ovarian Follicular Development. Frontiers in Physiology, 11, 595086. https://doi.org/10.3389/fphys.2020.595086

Kilic, U., Gok, O., Erenberk, U., Dundaroz, M. R., Torun, E., Kucukardali, Y., Elibol-Can, B., Uysal, O., Dundar, T. (2015). A remarkable age-related increase in SIRT1 protein expression against oxidative stress in elderly: SIRT1 gene variants and longevity in human. PLOS ONE, 10(3), e0117954. https://doi.org/10.1371/journal.pone.0117954

Alam, F., Shahid, M., Riffat, S., Zulkipli, I. N., Syed, F., Ashraf, M., Rehman, R. (2023). SIRT1 and antioxidants in infertile females: Exploration of the role of vitamin D. PLOS ONE, 18(7), e0287727. https://doi.org/10.1371/journal.pone.0287727

Mora Rodríguez, J. A., Porchia, L. M., Camargo, F., López-Bayghen, E. (2019). The use of insulin-like growth factor 1 improved the parameters of the seminogram in a patient with severe oligoasthenoteratozoospermia. SAGE Open Medical Case Reports, 7, 2050313X19834154. https://doi.org/10.1177/2050313X19834154

Grimberg, A. (2003). Mechanisms by which IGF-I may promote cancer. Cancer Biology & Therapy, 2(6), 630-635.

Vitale, G., Pellegrino, G., Vollery, M., Hofland, L. J. (2019). Role of IGF-1 System in the Modulation of Longevity: Controversies and New Insights From a

Centenarians' Perspective. Frontiers in Endocrinology, 10, 27. https://doi.org/10.3389/fendo.2019.00027

Chabris, C. F., Lee, J. J., Cesarini, D., Benjamin, D. J., Laibson, D. I. (2015). The Fourth Law of Behavior Genetics. Current Directions in Psychological Science, 24(4), 304-312. https://doi.org/10.1177/0963721415580430

Choi, S. W., Mak, T. S., O'Reilly, P. F. (2020). Tutorial: A guide to performing polygenic risk score analyses. Nature Protocols, 15(9), 2759-2772. https://doi.org/10.1038/s41596-020-0353-1

Mills, M. C., Barban, N., Tropf, F. C. (2020). An Introduction to Statistical Genetic Data Analysis. MIT Press.

Bush, W. S., & Moore, J. H. (2012). Chapter 11: Genome-wide association studies. PLoS Computational Biology, 8(12), e1002822. https://doi.org/10.1371/journal.pcbi.1002822

Sebastiani, P., Gurinovich, A., Bae, H., Andersen, S., Malovini, A., Atzmon, G., Villa, F., Kraja, A. T., Ben-Avraham, D., Barzilai, N., Puca, A., Perls, T. T. (2017). Four Genome-Wide Association Studies Identify New Extreme Longevity Variants. The Journals of Gerontology: Series A, 72(11), 1453–1464. https://doi.org/10.1093/gerona/glx027

Sebastiani, P., Solovieff, N., Dewan, A. T., Walsh, K. M., Puca, A., Hartley, S. W., Melista, E., Andersen, S., Dworkis, D. A., Wilk, J. B., Myers, R. H., Steinberg, M. H., Montano, M., Baldwin, C. T., Hoh, J., Perls, T. T. (2012). Genetic signatures of exceptional longevity in humans. PLoS ONE, 7(1), e29848. https://doi.org/10.1371/journal.pone.0029848

Joshi, P. K., Pirastu, N., Kentistou, K. A., et al. (2017). Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. Nature Communications, 8, 910. https://doi.org/10.1038/s41467-017-00934-5

Marioni, R. E., Ritchie, S. J., Joshi, P. K., Hagenaars, S. P., Okbay, A., Fischer, K., Adams, M. J., Hill, W. D., Davies, G., Nagy, R., Amador, C., Läll, K., Metspalu, A., Liewald, D. C., Campbell, A., Wilson, J. F., Hayward, C., Esko, T., Porteous, D. J., & Benjamin, D. J. (2016). Genetic variants linked to education predict longevity. Proceedings of the National Academy of Sciences of the United States of America, 113(47), 13366-13371. https://doi.org/10.1073/pnas.1605334113

Mathieson, I., Day, F. R., Barban, N., et al. (2023). Genome-wide analysis identifies genetic effects on reproductive success and ongoing natural selection at the FADS locus. Nature Human Behaviour, 7, 790-801. https://doi.org/10.1038/s41562-023-01528-6

Barban, N., De Cao, E., Francesconi, M. (2021). Gene-Environment Effects on Female Fertility. CESifo Working Paper No. 9337. https://doi.org/10.2139/ssrn.3938650