Family matters: a genealogical inquiry into the familial component of longevity

Berg, N.M.A. van den

FAMILY MATTERS

A genealogical inquiry into the familial component of longevity

PROEFSCHRIFT

ter verkrijging van
de graad van Doctor aan de Universiteit Leiden,
op gezag van de Rector Magnificus Prof. Dr. C. J. J. M. Stolker,
volgens besluit van het College voor Promoties
te verdedigen op woensdag 22 januari 2020,
klokke 15:00 uur

door
Niels Martinus Anthonius **van den Berg**
geboren te Haps
in 1987

Promotoren:	Prof. dr. P. Slagboom
	Prof. dr. A. A. P. O Janssens (Radboud Universiteit)
Co-promotor:	Dr. M. Beekman
Leden promotiecommissie:	Prof. dr. D. I. Boomsma (Vrije Universiteit Amsterdam)
	Prof. dr. H. Putter
	Prof. dr. J. Kok (Radboud Universiteit)

Family matters
A genealogical inquiry into the familial component of longevity
N. M. A. van den Berg, MSc

Layout design: Kimberley Teunissen
Cover design: Kimberley Teunissen
Printed by: Proefschriftmaken.nl
Published by Global Academic Press
Imprint bibliography: Global Academic Press
ISBN/EAN: 978-94-6380-684-8

CONTENTS

CHAPTER 1

INTRODUCTION

Progress in genealogic and genetic longevity

During the past 200 years human life expectancy at birth significantly increased in industrialized countries, with record male and female life expectancy increasing from 43/45 years in 1840 to 79/85 years in 2015[1]. At first, the life expectancy increased because of a better understanding of how to care for newborns, resulting in a strong decline in infant and childhood mortality as well as mortality of women giving birth[1–4]. After 1900, the availability of public sanitation, the understanding of hygiene and effective medical healthcare increased significantly and as a result, the group of middle aged (~50 years) and older people (~80 years) also started to live longer[1,5]. Currently, the group of older people represent the fastest growing segment of the population[6]. However, even though people started to live longer, the time they spend in good physical and cognitive health did not increase at an equal rate to the life expectancy, causing many years with chronic disabilities[7–11].

The increase in life expectancy was far too rapid to be caused by genetic change[12] and this is reflected in the low heritability (12-25%)[13–16] of age at death (lifespan) in the population at large. However, the capacity to outlive one's birth cohort members across the entire life course into extremely old age (longevity) clusters strongly within families[17–21] and is known as the familial component of longevity. This familial clustering is illustrated by genealogical studies which showed that parents, siblings[17–19,21–24], and children[19,25–29] of long-lived persons lived longer than first degree relatives of non-long-lived persons or population controls. Members of such long-lived families show a holistic nature of healthy aging, which is illustrated by their life-long decreased chance of dying compared to the general population[21,30]. In addition, they have a lower risk of coronary artery disease, cancer, hypertension and type 2 diabetes[31–33], and if affected, the onset of disease is at later ages and with shorter duration[7,8,34–36]. This phenomenon of late disease onset during a short period of time is known as the "compression of morbidity". Moreover, members of long-lived families enjoy better immune and metabolic health in middle and old age than other individuals of the same age in the general population[8,9,33,37–40]. Thus, understanding the genetic factors influencing longevity may provide novel insights into the biological mechanisms that promote health and minimize disease risk[13,41].

The identification of longevity loci, however, has been challenging and only a handful of genetic variants have been shown to associate with longevity across multiple independent studies[40–47]. The most consistent evidence has been obtained for variants in the APOE and FOXO3A genes[41–46,48], in either genome-wide association studies (GWAS) or candidate gene studies. The limited success of genetic longevity studies in identifying causal variants can be attributed to a number of issues. First, the genetic and environmental heterogeneity[30] of study populations is often large, which makes it difficult to identify specific longevity-associated genes. Second, the focus has been on the discovery of common genetic variants by GWAS approaches whereas rare variants, that can be discovered in whole genome sequencing data, have expectedly bigger functional effects on the trait and may thus be most identifiable. Third, and most relevant of all: the lack of a strict definition for human longevity[13,30], as illustrated by the large variation of longevity definitions[7,13,28,29,32,33,35,40–44,15,45–54,17,19,22,24–27]. This results in a mix of sporadically long-lived cases with those descending from a long-lived family.

Most genealogical studies define longevity as age at death, reflecting a continuous/quantitative trait[17,22,24,26,27,29,33,52] while only a few studies focus on individuals who died at a late age, for example 80, 90, or even 100 years. This is the opposite situation for genetic studies, as only a few studies define longevity as age at death or include mortality information from relatives of study participants[40,47] and, most genetic studies focus on single individuals who are extremely old[43–45,48,53,54]. In other words, they apply an age threshold to define long-lived cases, which may dilute heritable with sporadic cases. The epidemiological transition of the past 200 years, by which the lifespan increased worldwide, has resulted in many long-lived singletons for which heritable factors contributed only partly to their long lifespan. Research into these long-lived singletons showed that a wide range of individual factors, such as socio-economic status, familial resources, genes and, the living environment, associate with lifespan and longevity[55–63,55,57,61,64]. In fact, many of these factors are known to cluster in families[65–70]. However, research into the driving factors of the familial component of longevity is scarce and thus it remains to be explored which factors contribute to the longevity of members of long-lived families.

Currently, many questions regarding familial longevity remain; 1. how can we identify heritable cases for inclusion in genetic longevity studies as many individuals have become long-lived due to factors driving the epidemiological transition?, 2. to what extent is longevity passed on to subsequent generations?, 3. do men and women equally transmit the longevity trait? and, 4. how many family members should be long-lived to represent a familial longevity trait passed on from one generation to the next?. Moreover, it is unknown to what extent the familial component of longevity can be explained by genetic and/or non-genetic (social) factors, connecting to the classic "nature and nurture" dilemma. This thesis focuses on multiple sources of genealogical data to define and explore the heritable longevity trait in large scale family data, taking into account genetic as well as social factors that influence the familial component of longevity so that the results can be used for future studies.

Genealogical data to investigate the familial component of longevity

Such questions, about the familial component and the definition of longevity, can be investigated using individual-level historical genealogical data. Such genealogical data can be obtained through different data sources: 1. civil certificates (birth, marriage, and death certificates), 2. population registers, 3. parish registers, 4. census records, and 5. genealogy websites. *Table 1* explains the characteristics of these different data sources and compares their advantages and disadvantages. The table also shows the types of data sources underlying the databases used for this thesis. These datasets include the Leiden Longevity Study (LLS), the Historical Sample of the Netherlands (HSN), the HSN case/control study, the Utah Population Database (UPDB), and the LINKing System for historical family reconstruction (LINKS). The data sources can be used for life course and family reconstruction. Life course reconstruction refers to the identification process of vital events that happened during the life course of an individual, such as birth, marriage, divorce, moving from one place to another, changes in occupation, and death. Family reconstruction refers to the

identification process of relatives, such as parents, siblings, aunts and uncles, and grandparents. In the next paragraphs a short description will be provided of the different data sources and the databases used in in for this thesis.

Table 1: Overview of data sources that can be used for life course and family reconstruction

Source documents	Life course and family reconstruction method and information	Pros	Cons	Dataset
Civil certificates; birth, marriage, and death certificates	Linking certificates for single individuals and between individuals to reconstruct individual life courses and families	Very high quality information Very good representation of childhood mortality	Potential of linking errors Fragmented life course information	LINKS UPDB
Population registers	Continuous observation of individuals' life courses and information on individuals' relatives are on the cards of individuals	Very high quality information Continuous life course information	Partial coverage of childhood mortality	HSN LLS
Parish registers	Church related continuous observation of individuals' life courses and information on individuals' relatives	Pedigrees can date back more than three centuries	Usually only focus on a small geographical area	UPDB
Census records	Interval (usually 10 years) information about individuals and household members	Available in most countries. Can date back more than two centuries	Long intervals between subsequent censuses, especially for family reconstructions	UPDB HSN
Self-reported genealogies and genealogy websites	Individuals reconstruct their own families. The extent of the individual life course reconstruction depends on the source material backing up the genealogies.	Extensive pedigrees dating back very long Large numbers of pedigrees and individuals	No or limited verified demographic information such as births, deaths, marriages, and profession	UPDB LLS

LLS = Leiden Longevity Study LINKS = LINKing System for historical family reconstruction and HSN = Historical Sample of the Netherlands. Life course reconstruction refers to the reconstruction of a person's life course by, for example following a person from birth to death and observing what life events this person experiences during his/her life. Family reconstruction refers to the reconstruction of family ties, such as identifying a person's parents, siblings, children, grandparents, etc. The "dataset" column refers to the dataset that is used for the analysis during the period of this thesis.

Population registers provide a continuous observation of a person's life course[71–73]. This means that whenever a vital event, such as moving from one place to the other or the birth of a child, happens, this event is registered in the population register. Next to the registration of vital events, population registers also contain identifying information on a person's parents, siblings, and children. As such, population registers are useful to obtain life course information and family ties of a person. Of course, to obtain life course information of the identified relatives, the population register for that specific person should be obtained[73]. In the Netherlands, these registers are officially maintained by the Dutch government and as such, the life course and family information is very accurate. Data from the Dutch population register is available in the Historical Sample of the Netherlands (HSN) and the HSN is supplemented with civil certificate data[72–74].

Civil certificates provide interval life course observations[73], with the main observation moments at birth, marriage and death. Birth, marriage and death certificates all contain reference to the parental names of the person to which the certificate belongs. These names can be used to identify the civil certificates of all relatives of a person. Marriage certificates are most practical to reconstruct families because they contain the parental names of both spouses. The civil certificates contain high quality life course information and using available parental information, family reconstruction is relatively easy. However, complex, common, and ambiguous first and/or last names provide difficulties in identifying the civil certificates of individuals or their parents and may even result in identifying and connecting the wrong certificates. In the Netherlands, family and life course reconstruction based on the civil certificates has been automated using a name linking algorithm[73] and resulted in the LINKing System for historical family reconstruction (LINKS) data[75–77].

Parish registers are often similar to the Dutch population registers regarding the life course and family member information they contain. The main difference is that they are initiated and maintained by a parish church and hence, usually cover a small geographical area. In contrast to the coverage of a small geographical area, census records usually cover entire countries. Census records can provide a snapshot of a person's life course and the relatives of a person, based on the moment the census was conducted. Census records are often used to supplement life course and family reconstruction information based on other sources and moreover, with the increasing availability of data, not only census records can be combined with other sources, but all different data sources can be combined. The HSN for example includes, to a small extent, census and civil certificate information to supplement the population register information. The Utah Population database (UPDB) was initially based on records of the Mormon Church in Utah (US) and now covers the entire Utah population. The UPDB is the most extensive historical genealogical database in the world at this moment[78,79], *https://uofuhealth.utah.edu* and combines information from parish registers, civil certificates, and census information to establish high quality life course and family reconstructions. Additionally, the data are extended with information from medical records, and driver license records[78,79].

Genealogical websites provide an interesting source of family reconstructions as they may cover persons well before the introduction of the civil registry and population registers[15,16]. The pedigrees contained on genealogical websites are usually constructed by hobby-genealogists who are interested in their own family history. For (historical)

genealogical research, the pros are that the pedigrees are structured in a standardized international format (GEDCOM) and that the pedigrees may date back for many years (sometimes to before 1700). A downside of this is that there are no governmental sources before ~1800 to verify demographic and mortality information of pedigree members so that the error rate increases before 1800. Similar to the misreporting of demographic and mortality information, it is also more difficult to verify family relationships in those early years. In addition, the quality of the life course and family reconstructions may vary strongly between families as one hobby-genealogist may try to verify mortality information and family ties with the corresponding civil certificates and population registers, but another may not do this. It is also important to note that there may be a self-selection bias in such data towards non-extinct families. Living people map their own family threes and thus, extinct families may be underrepresented, which may or may not cause issues, depending on the type of research that is done with the data.

Aim and outline of this thesis

Understanding the genetic factors influencing longevity may provide novel insights into the mechanisms that promote health and minimize disease risk[13,41]. Hence, the aim of this thesis is to study the familial component of longevity by first establishing a standardized definition of longevity and subsequently investigating the intergenerational transmission of longevity. Ultimately we aim to establish a genetically enriched group (cases) and a group which represents the general population (controls) that could be included in novel genetic longevity studies. In our analyses we take into account social factors, such as socio-economic status, fertility measurements, familial resources, and living environment, which potentially contribute to the familial component of longevity. We also take the survival of spouses into account and explore the association between longevity and family size.

Data

For our investigations we use different datasets. These include the data available in the Leiden Longevity Study (LLS)[19,21], which is based on self-reported genealogies, supplemented with Dutch population registers provided by the Central Bureau of Genealogy (CBG) and Dutch Personal Records Database (PRD). We further used the LINKS and HSN data, provided by the International Institute for Social History (IISH), and the UPDB provided by the University of Utah.

The LLS was initiated in 2002 to study the genetic determinants of human longevity. Men and women could participate if they were alive and aged ≥89 and ≥91 respectively and had at least one sibling fulfilling the same criteria. From these participants, relatives were identified, including their parents, spouses, siblings, and children. The LLS consists of 421 families and covers 3 generations with a birth cohort range between 1850 and 2019. The LINKS Zeeland (from here: LINKS) data indexing began in 1995. LINKS contains around 700,00 birth, 300,00 marriage, and 600,000 death certificates providing information for around 2,000,000 persons and covering a maximum number of 7 generations. LINKS covers birth cohorts between 1750 and 1920. The UPDB data construction began in the mid-1970s with genealogy records from the archives at the Utah Family History Library and was initially based on the founding members of

the Utah population, their descendants, and then subsequently all individuals living in Utah. The UPDB contains information on more than 11 million individuals and covers a maximum of 17 generations. The UPDB covers birth cohorts between 1750 and 2019. The HSN 2010.01 release is based on a sample of birth certificates and contains complete life course information for 37,137 Dutch individuals (index persons (IPs)) born between 1849 and 1923 (32–34). These 37,137 persons were subsequently identified in the Dutch population registers and followed in the registers throughout their entire life course. The HSN 2010.01 covers a maximum number of 3 generations and covers birth cohorts between 1850 and 1922. The HSN was split into a case (persons who died ≥ 80 years) and a control (persons who died between 40 and 59 years) design for the birth cohorts 1860 - 1875 and for these groups, the third generation was extended with mortality information. Subsequently, two generations were added to the data. The final database covers 57,337 persons from 1,326 five-generational families with birth cohorts ranging between 1850 and 1995. *Table 2* provides an overview of the data sources underlying the datasets that are used for this thesis and *Figure 1* illustrates the birth cohorts included in the data analyses of this thesis.

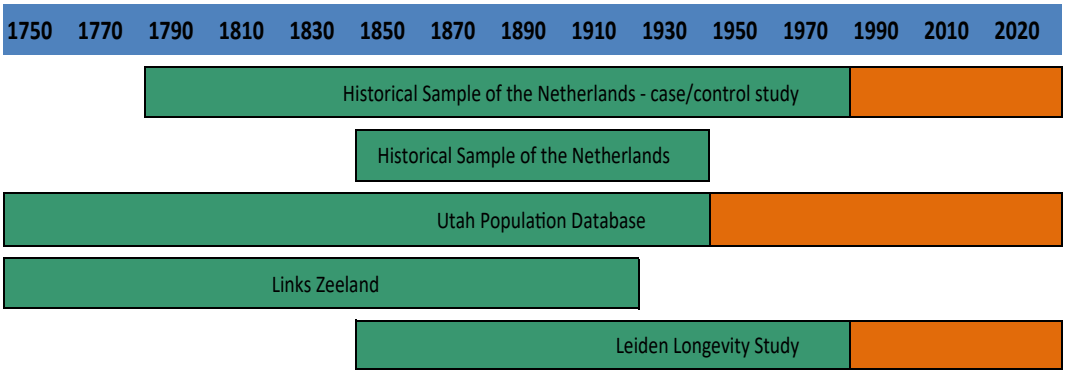


Figure 1: Overview of the databases and study periods used for this thesis.

The green colors in the time-line illustrate the included birth cohorts of all databases used during this thesis. The orange colors illustrate the birth cohorts that were not included in the analyses.

Table 2: Overview of databases that have been used for in this thesis

Dataset	Short description	Type	Birth cohorts	Continuous expanding	Source material
LLS	The LLS was initiated in 2002 and consists of 421 families, covering 3 generations.	Specific population database	First birth cohort is 1850 and the last birth cohort is 2017	No	Population registers and self-reported pedigrees
LINKS Zeeland	The LINKS data indexing began in 1995. The data currently covers over 25 million Dutch vital event records. LINKS Zeeland contains information for around 2,000,000 persons and covering a maximum number of 7 generations.	General population database	Around 1750 and around 1920	Yes	Birth, marriage, and Death certificates
UPDB	The UPDB data construction began in the mid-1970s. The UPDB contains information on more than 11 million individuals and covers a maximum of 17 generations.	General population database	Around 1750 up to 2019	Yes	Church registers, birth, marriage, and death registers, census records, and genealogy website information
HSN	The HSN was initiated around 20 years ago and contains complete life course information for 37,137 Dutch individuals The HSN 2010.01 release covers a maximum number of 3 generations.	General population database	First birth is 1850 and last birth is 1922	Yes	Population registers, birth, marriage, and death registers
HSN case/ control	The HSN case/control study was initiated in 2014 and the final database covers 57,337 persons from 1,326 five-generational families.	General population database	Around 1850 up to 1995	No	Population registers, birth, marriage, and death registers

LLS = Leiden Longevity Study, LINKS = LINKing System for historical family reconstruction and HSN = Historical Sample of the Netherlands. The "dataset" column refers to the dataset that is used for the analysis during the period of this thesis. Specific population refers to a database which reflects a specific study population, such as the LLLS which contains a directly or indirectly selected group of people, general population refers to a database which reflects a broad population, such as the HSN, which reflects the general Dutch population.

Outline

In *chapter 2* we review the relevant literature investigating the familial component of longevity. We focus on heritability studies, studies investigating the transmission of lifespan and longevity as well as lifespan and longevity inheritance patterns. We further discuss important environmental/social factors that affect individual lifespan and longevity or potentially affect the transmission of lifespan and longevity between parents and offspring. In other words, we discuss the factors that associate with individual longevity and the familial component of longevity. We emphasize the difference between lifespan and longevity traits. Lifespan generally refers to the age at death of a person whereas longevity refers to survival into extreme ages, such as 80, 90, 100 years, or an extreme survival percentile such as belonging to the top 5% birth cohort specific survivors. Finally, we discuss a strategy to study familial longevity and to identify a definition of longevity that may best represent the heritable longevity trait.

The LINKS database was recently constructed and before using these data we wanted to validate the quality of the life course and family reconstructions by comparing the LINKS data to the well-established HSN data. Hence, in *Chapter 3* we test the quality of the LINKS and HSN data by comparing life course information and family reconstruction of ~400 persons born in Zeeland who could be identified in both the HSN and LINKS data. We focus on overlap and differences in demographic information, such age at death and age at first childbirth, and family information, such as the number of identified siblings and children. Finally, we expect that some of the differences in demographic and family indicators, such as the number of children, between the two databases can be explained by how migration is represented in the two databases. In LINKS, migration during the life course to another province or country is not included. Hence, we test for migration differences as an explaining factor for discrepancies between demographic and family indicators in the HSN and LINKS.

After confirming the quality of the life span and family reconstructions in the LINKS data, we initiated a collaboration with prof. Ken Smith, who provided access to the UPDB, to investigate the definition and subsequently, the familial component of longevity. Thus, in *Chapter 4* we use three-generational demographic and mortality data from two large datasets, UPDB (US) and LINKS (Netherlands). We focus on 20,360 families who are unselected for mortality. The data contains 20,360 index persons, their parents (N=40,72), siblings (N=108,122), spouses (N=22,018), and children (N=123,599), comprising 314,819 individuals in total. We use these data to investigate which survival percentile best isolates the heritable component of longevity and we subsequently determine the importance of long-lived family members for case selection so that those insights can be used in genetic studies to identify novel longevity loci. In the analyses we include social and environmental factors, such as socio-economic status, religious denomination, number of children, birth order, and birth cohort, that may explain the intergenerational transmission of longevity. Moreover, we explore the survival of spouses marrying into longevity enriched families as an indicator for shared resources, lifestyles, and potentially socio-economic status during middle and late-life as explaining factors for the familial component of longevity.

We apply the novel survival percentile threshold based longevity definition in *Chapter 5*, where we focus on 2 generations from the Leiden Longevity Study, containing 944 long-lived siblings (participants), their parents (N=842), siblings (N=2302), and spouses (N=809) from 421 LLS families. We define longevity as belonging to the top 1% survivors of their birth cohort to investigate 1. a potential sex-specific inheritance pattern of longevity, 2. a potential survival advantage of long-lived sibships as compared to long-lived singletons and 3. whether the parents of these siblings had a life-long sustained survival advantage. Similar to *Chapter 4*, we include the spouses of the 944 LLS participants to explore mid

and late-life shared environmental/social factors contributing to the familial component of longevity.

Following-up on the longevity definition as established in chapter 4, we focus on establishing the proportion of ancestral blood relatives that should be long-lived (at least belonging to the top 10% survivors of their birth cohort) in order to observe a survival advantage in their descendants and subsequently define cases with the heritable longevity trait for inclusion in genetic studies. In *Chapter 6*, we therefore describe and explore the HSN case/control data, which is specifically compiled to cover 5 generations and connect extinct family members to their living descendants. Data construction started with 1326 families form the original HSN and was extended by acquiring population registers and population register information from the CBG and the PRD respectively. Moreover, we were granted permission by the Dutch government to obtain the current addresses of all living descendants. We use the data to investigate if longevity is transmitted for multiple generations and if the longevity effect dwindles over generations by comparing long-lived cases (died ≥ 80 years) and their descendants to population resembling controls (died between 40 and 59 years) and their offspring. Furthermore, we establish how many family members should be long-lived in order to avoid phenocopies and subsequently investigate how often long-lived parents from a long-lived family pass on their longevity to their children compared to long-lived parents from general population families. Equal to LINKS and the UPDB analyses, here we include spouses over multiple generations to explore mid and late-life environmental effects. Finally, we utilize the insights to identify a novel case and control group for future genetic and social studies into the heritable longevity trait.

References

1 Oeppen, J. & Vaupel, J. W. Broken limits to life expectancy. *Science* **296**, 1029–1031 (2002).

2 Fogel, R. W. Aging, *Health and Public Policy.* **30**, (2004).

3 Fogel, R. W. & Costa, D. L. A Theory of Technophysio Evolution, With Some Implications for Forecasting Population, Health Care Costs, and Pension Costs. *Demography* **34**, 49 (1997).

4 Wachter, K. W. & Finch, C. E. *Between Zeus and the Salmon: The Biodemography of Longevity. Population (English Edition)* (Washington, DC: The National Academies Press, 1997).

5 Vaupel, J. W. Biodemographic Trajectories of Longevity. *Science* (80-.). **280**, 855–860 (1998).

6 Kontis, V. *et al.* Future life expectancy in 35 industrialised countries: projections with a Bayesian model ensemble. *Lancet* **389**, 1323–1335 (2017).

7 Christensen, K., McGue, M., Petersen, I., Jeune, B. & Vaupel, J. W. Exceptional longevity does not result in excessive levels of disability. *Proc. Natl. Acad. Sci.* **105**, 13274–13279 (2008).

8 Christensen, K. *et al.* Physical and cognitive functioning of people older than 90 years: a comparison of two Danish cohorts born 10 years apart. *Lancet* **382**, 1507–1513 (2013).

9 Zeng, Y., Feng, Q., Hesketh, T., Christensen, K. & Vaupel, J. W. Survival, disabilities in activities of daily living, and physical and cognitive functioning among the oldest-old in China: a cohort study. *Lancet* **389**, 1619–1629 (2017).

10 Pignolo, R. J. Exceptional Human Longevity. *Mayo Clin Proc* **94**, 1–15 (2018).

11 Crimmins, E. M. Lifespan and Healthspan: Past, Present, and Promise. *Gerontologist* **55**, 901–911 (2015).

12 Burger, O., Baudisch, A. & Vaupel, J. W. Human mortality improvement in evolutionary context. *Proc. Natl. Acad. Sci.* **109**, 18210–18214 (2012).

13 van den Berg, N., Beekman, M., Smith, K. R., Janssens, A. & Slagboom, P. E. Historical demography and longevity genetics: Back to the future. *Ageing Res. Rev.* **38**, 28–39 (2017).

14 Herskind, A. M. *et al.* The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Hum. Genet.* **97**, 319–323 (1996).

15 Kaplanis, J. *et al.* Quantitative analysis of population-scale family trees with millions of relatives. *Science* **360**, 171–175 (2018).

16 Ruby, J. G. *et al.* Estimates of the Heritability of Human Longevity Are Substantially Inflated due to Assortative Mating. *Genetics* **210**, 1109–1124 (2018).

17 Perls, T. T. *et al.* Life-long sustained mortality advantage of siblings of centenarians. *Proc. Natl. Acad. Sci.* **99**, 8442–8447 (2002).

18 Pedersen, J. K. et al. The Survival of Spouses Marrying Into Longevity-Enriched Families. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **72**, 109–114 (2017).

19 Schoenmaker, M. *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur. J. Hum. Genet.* **14**, 79–84 (2006).

20. Ljungquist, B., Berg, S., Lanke, J., McClearn, G. E. & Pedersen, N. L. The Effect of Genetic Factors for Longevity: A Comparison of Identical and Fraternal Twins in the Swedish Twin Registry. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **53A**, 441–446 (1998).

21 Berg, N. van den *et al.* Longevity Around the Turn of the 20th Century: Life-Long Sustained Survival Advantage for Parents of Today's Nonagenarians. *Journals Gerontol. Ser. A* **73**, 1295–1302 (2018).

22 Willcox, B. J., Willcox, D. C., He, Q., Curb, J. D. & Suzuki, M. Siblings of Okinawan centenarians share lifelong mortality advantages. *J. Gerontol. A. Biol. Sci. Med. Sci.* **61**, 345–54 (2006).

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

24 Sebastiani, P., Nussbaum, L., Andersen, S. L., Black, M. J. & Perls, T. T. Increasing Sibling Relative Risk of Survival to Older and Older Ages and the Importance of Precise Definitions of "Aging," "Life Span," and "Longevity". *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **71**, 340–346 (2016).

25 Kerber, R. A., Brien, E. O., Smith, K. R. & Cawthon, R. M. Familial Excess Longevity in Utah Genealogies. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **56**, 130–139 (2001).

26 Gudmundsson, H., Gudbjartsson, D. F. & Kong, A. Inheritance of human longevity in Iceland. *Eur. J. Hum. Genet. EJHG* **8**, 743–749 (2000).

27 Houde, L., Tremblay, M. & Vézina, H. Intergenerational and Genealogical Approaches for the Study of Longevity in the Saguenay-Lac-St-Jean Population. *Hum. Nat.* **19**, 70–86 (2008).

28 Kemkes-Grottenthaler, A. Parental effects on offspring longevity—evidence from 17th to 19th century reproductive histories. *Ann. Hum. Biol.* **31**, 139–158 (2004).

29 Deluty, J. A., Atzmon, G., Crandall, J., Barzilai, N. & Milman, S. The influence of gender on inheritance of exceptional longevity. *Aging (Albany. NY).* **7**, 412–418 (2015).

30 van den Berg, N. *et al.* Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nat. Commun.* **10**, 35 (2019).

31 Westendorp, R. G. J. et al. Nonagenarian Siblings and Their Offspring Display Lower Risk of Mortality and Morbidity than Sporadic Nonagenarians: The Leiden Longevity Study. *J. Am. Geriatr. Soc.* **57**, 1634–1637 (2009).

32 Terry, D. F. *et al.* Lower All-Cause, Cardiovascular, and Cancer Mortality in Centenarians' Offspring. *J. Am. Geriatr. Soc.* **52**, 2074–2076 (2004).

33 Dutta, A. et al. Longer lived parents: Protective associations with cancer incidence and overall mortality. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **68**, 1409–1418 (2013).

34 Andersen, S. L., Sebastiani, P., Dworkis, D. a., Feldman, L. & Perls, T. T. Health Span Approximates Life Span Among Many Supercentenarians: Compression of Morbidity at the Approximate Limit of Life Span. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **67A**, 395–405 (2012).

35 Ash, A. S. *et al.* Are Members of Long-Lived Families Healthier Than Their Equally Long-Lived Peers? Evidence From the Long Life Family Study. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **70**, 971–976 (2015).

36 Evert, J., Lawler, E., Bogan, H. & Perls, T. Morbidity Profiles of Centenarians: Survivors, Delayers, and Escapers. *Journals Gerontol. - Med. Sci.* **58**, 232–237 (2003).

37 Vaarhorst, A. A. M. *et al.* Lipid metabolism in long-lived families: the Leiden Longevity Study. *Age (Omaha).* **33**, 219–227 (2011).

38 Wijsman, C. A. *et al.* Familial longevity is marked by enhanced insulin sensitivity. *Aging Cell* **10**, 114–121 (2011).

39 Rozing, M. P. *et al.* Low serum free triiodothyronine levels mark familial longevity: The leiden longevity study. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **65A**, 365–368 (2010).

40 Joshi, P. K. *et al.* Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nat. Commun.* **8**, 1–13 (2017).

41 Eline Slagboom, P., van den Berg, N. & Deelen, J. Phenome and genome based studies into human ageing and longevity: An overview. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1864**, 2742–2751 (2018).

42 Shadyab, A. H. & LaCroix, A. Z. Genetic factors associated with longevity: A review of recent findings. *Ageing Res. Rev.* **19**, 1–7 (2015).

43 Deelen, J. *et al.* Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* **23**, 4420–4432 (2014).

44 Broer, L. *et al.* GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **70**, 110–118 (2015).

45 Willcox, B. J. *et al.* FOXO3A genotype is strongly associated with human longevity. *Proc. Natl. Acad. Sci.* **105**, 13987–13992 (2008).

46 Flachsbart, F. *et al.* Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc. Natl. Acad. Sci.* 106, 2700–2705 (2009).

47 Pilling, L. C. *et al.* Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany. NY).* **9**, 2504–2520 (2017).

48 Sebastiani, P. *et al.* Four Genome-Wide Association Studies Identify New Extreme Longevity Variants. *J. Gerontol. A. Biol. Sci. Med. Sci.* **72**, 1453–1464 (2017).

49 Sebastiani, P. *et al.* A Family Longevity Selection Score: Ranking Sibships by Their Longevity, Size, and Availability for Study. *Am. J. Epidemiol.* **170**, 1555–1562 (2009).

50 Christensen, K., Johnson, T. E. & Vaupel, J. W. The quest for genetic determinants of human longevity: challenges and insights. Nat. Rev. Genet. **7**, 436–48 (2006).

51 Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).

52 Jarry, V., Gagnon, A. & Bourbeau, R. Survival advantage of siblings and spouses of centenarians in 20th-century Quebec. *Can. Stud. Popul.* **39**, 67 (2013).

53 Flachsbart, F. *et al.* Immunochip analysis identifies association of the RAD50/IL13 region with human longevity. *Aging Cell* **15**, 585–588 (2016).

54 Zeng, Y. *et al.* Novel loci and pathways significantly associated with longevity. *Sci. Rep.* **6**, 21243 (2016).

55 Sebastiani, P. *et al.* Families Enriched for Exceptional Longevity also have Increased Health-Span: Findings from the Long Life Family Study. *Front. Public Heal.* **1**, 1–9 (2013).

56 Ailshire, J. A., Beltrán-Sánchez, H. & Crimmins, E. M. Social Characteristics and Health Status of Exceptionally Long-Lived Americans in the Health and Retirement Study. *J. Am. Geriatr. Soc.* **59**, 2241–2248 (2011).

57 Ash, A. S. *et al.* Are Members of Long-Lived Families Healthier Than Their Equally Long-Lived Peers? Evidence From the Long Life Family Study. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **70**, 971–976 (2015).

58 Jopp, D. S., Boerner, K., Ribeiro, O. & Rott, C. Life at Age 100: An International Research Agenda for Centenarian Studies. *J. Aging Soc. Policy* **28**, 133–147 (2016).

59 Boerner, K., Jopp, D. S., Park, M.-K. S. & Rott, C. Whom Do Centenarians Rely on for Support? Findings From the Second Heidelberg Centenarian Study. *J. Aging Soc. Policy* **28**, 165–186 (2016).

60 Seeman, T. E. *et al.* Behavioral and Psychosocial Predictors of Physical Performance: MacArthur Studies of Successful Aging. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **50A**, M177–M183 (1995).

61 Andersen, S. L. *et al.* Personality Factors in the Long Life Family Study. *Journals Gerontol. Ser. B* **68**, 739–749 (2013).

62 Yang, Y. C. *et al.* Social relationships and physiological determinants of longevity across the human life span. *Proc. Natl. Acad. Sci.* **113**, 578–583 (2016).

63 Hensley, B. *et al.* Life Events and Personality Predicting Loneliness Among Centenarians: Findings From the Georgia Centenarian Study. *J. Psychol.* **146**, 173–188 (2012).

64 Yashin, A. I. *et al.* "Predicting" parental longevity from offspring endophenotypes: Data from the Long Life Family Study (LLFS). *Mech. Ageing Dev.* **131**, 215–222 (2010).

65 Heath, A. C. *et al.* Education policy and the heritability of educational attainment. *Nature* **314**, 734–736 (1985).

66 Spera, C., Wentzel, K. R. & Matto, H. C. Parental Aspirations for Their Children's Educational Attainment: Relations to Ethnicity, Parental Education, Children's Academic Performance, and Parental Perceptions of School Climate. *J. Youth Adolesc.* **38**, 1140–1152 (2009).

67 Eccles, J. S. Influences of parents' education on their children's educational attainments: the role of parent and child perceptions. *London Rev. Educ.* **3**, 191–204 (2005).

68 Edvardsson, K. *et al.* Giving offspring a healthy start: parents' experiences of health promotion and lifestyle change during pregnancy and early parenthood. *BMC Public Health* **11**, 936 (2011).

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

70 Thibault, H., Contrand, B., Saubusse, E., Baine, M. & Maurice-Tison, S. Risk factors for overweight and obesity in French adolescents: Physical activity, sedentary behavior and parental characteristics. *Nutrition* **26**, 192–200 (2010).

71 Mandemakers, K. Building life courses datasets from population registers by the historical sample of the Netherlands. in *History and Computing* **14**, 87–107 (2002).

72 Mandemakers, K. *https://socialhistory.org/en/hsn/hsn-releases. HSN 2010.01 release* (2010).

73 van den Berg, N. *et al.* Families in Comparison: An individual-level comparison of life course and family reconstructions between population and vital event registers. *SocArXiv* (2018).

74 Mandemakers, K. Historical sample of the Netherlands. In P. K. Hall, R. McCaa, & G. Thorvaldsen (Eds.). in *Handbook of International Historical Microdata for Population Research* 149–177 (2000).

75 Mandemakers, K. & Laan, F. LINKS dataset Genes Germs and Resources. WieWasWie Zeeland. Civil Certificates. (2017).

76 *https://socialhistory.org/en/hsn/linking-system-historical-family-reconstruction-links*

77 Mandemakers, K. *LINKing System for historical family reconstruction.* (2007).

78 *https://healthcare.utah.edu/huntsmancancerinstitute/research/updb/data/family-records.php*

79 Bean, L. L., May, D. L. & Skolnick, M. The Mormon Historical Demography Project. *Hist. Methods A J. Quant. Interdiscip. Hist.* **11**, 45–53 (1978).

CHAPTER 2

HISTORICAL DEMOGRAPHY AND LONGEVITY GENETICS:

BACK TO THE FUTURE

Niels van den Berg[1]
Marian Beekman[1]
Ken Robert Smith[2]
Angelique A.P.O. Janssens[3]
P. Eline Slagboom[1]

[1] Department of Biomedical Data Sciences, section of Molecular Epidemiology,
Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands
[2] Department of Family and Consumer Studies; Population Sciences, Huntsman Cancer Institute,
University of Utah, 225 S. 1400 E. Rm 228 Salt Lake City, United States of America
[3] Radboud Group for Historical Demography and Family History,
Radboud University, Erasmusplein 1, 6525 HT Nijmegen, the Netherlands

Aging Research Reviews (September 2017)
https://doi.org/10.1016/j.arr.2017.06.005

Abstract

Research into the genetic component of human longevity can provide important insights in mechanisms that may protect against age-related diseases and multi-morbidity. Thus far only a limited number of robust longevity loci have been detected in either candidate or genome wide association studies. One of the issues in these genetic studies is the definition of the trait being either lifespan, including any age at death or longevity, i.e. survival above a diverse series of thresholds. Likewise heritability and segregation research have conflated lifespan with longevity. The heritability of lifespan estimated across most studies has been rather low. Environmental factors have not been sufficiently investigated and the total amount of genetic variance contributing to longevity has not been estimated in sufficiently well-defined and powered studies. Up to now, genetic longevity studies lack the required insights into the nature and size of the genetic component and the optimal strategies for meta-analysis and subject selection for Next Generation Sequencing efforts. Historical demographic data containing deep genealogical information may help in estimating the best definition and heritability for longevity, its transmission patterns in multi-generational datasets and may allow relevant additive and modifying environmental factors such as socio-economic status, geographical background, exposure to environmental effects, birth order, and number of children to be included. In this light historical demographic data may be very useful for identifying lineages in human populations that are worth investigating further by geneticists.

Introduction

During the past 200 years human life expectancy at birth significantly increased in western societies, with record female life expectancy increasing from 45 years in 1840 to 85 years in 2015[1]. Around 1950, even the oldest old (age 85 or older) started to show a pattern of extended life expectancy and today they are the fastest growing segment of older people[1]. This means that populations not only survive to higher ages than in the past, they also have a lower mortality rate, during their young and middle years[2]. Remarkably, the survival of a select few persons stands out of an otherwise aging population[3]. These persons were extremely long-lived and, most of all, showed little to no signs of age-related disease, allowing them to have extremely long and healthy lives[4–7]. Research into first-degree relatives of these long-lived individuals showed that they also had extremely long and healthy lives compared to relatives of individuals with more normative ages at death[8,9]. Hence, the familial component, including both genetic and environmental contributions, seemed to play a key role in gaining more knowledge about factors involved in healthy aging and in the capability to survive into extreme old ages (often called longevity)..

In the literature, the familial component of human longevity has been investigated using survival to extreme age and age at death as phenotypes of survival (see *Table 1*). The former actually refers to longevity whereas the latter refers to individual or population based lifespan. Both definitions are often used in the context of longevity research which is confusing and incorrect. Another complication is that most studies exclude infant and child mortality by applying a lower limit age threshold when considering the lifespan of a population or group of individuals. Unfortunately, there is no consensus on the age threshold for longevity studies. As a result of both the inconsistent use of terminology and different lower and upper limit age thresholds, the comparison of longevity studies is generally problematic[10]. We will refer to longevity as survival into extreme old ages whereas lifespan refers to age at death related measures (see *Table 1* and *Figure 1*).

Progress in longevity research is also hampered by the fact that longevity is likely dependent on an interplay between combinations of multiple genes and environmental factors[11–15] which makes it difficult to separate environmental from genetic influences. In fact, environmental influences likely moderate genetic effects on longevity[16–18]. Hence, in this review we describe how historical genealogical data can be used to study familial longevity by including family history information to identify longevous families with a high potential for genetic analysis, such as Next Generation Sequencing (NGS). We start by discussing the state of the art of genealogical heritability and segregation studies in the context of lifespan and longevity. Next we discuss the influence of environmental factors in longevity research, and finally we propose how historical genealogical and demographic data, and the results of genealogical studies can be included in genetic longevity research.

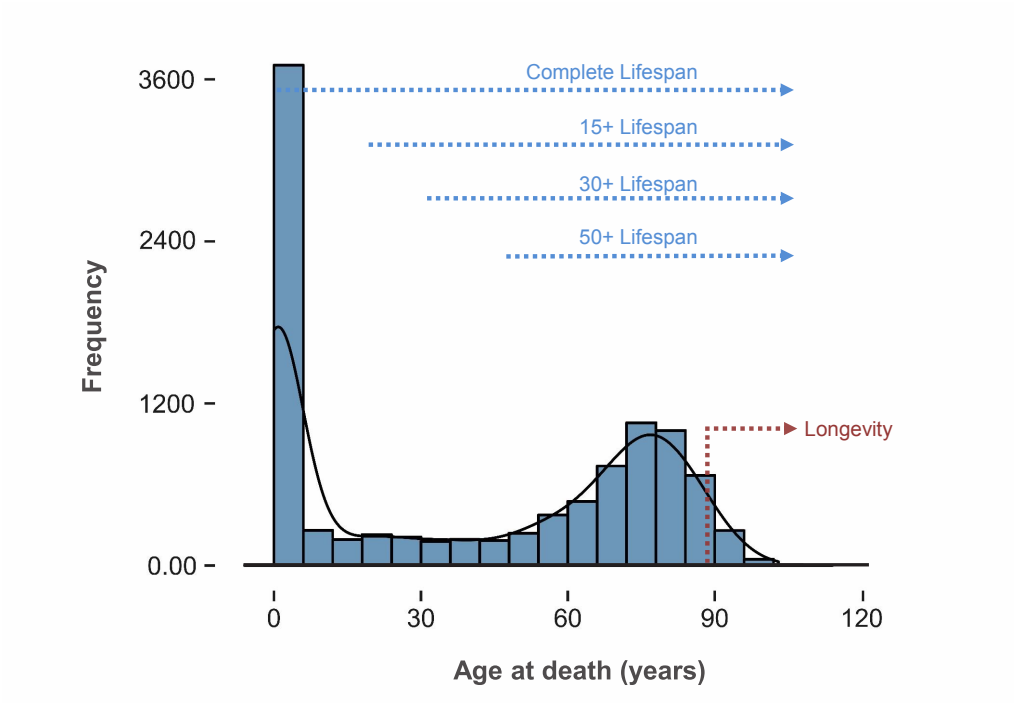
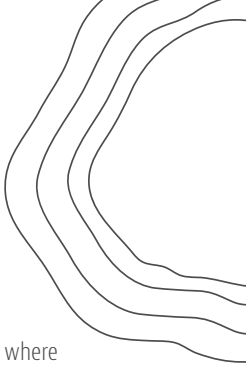


Figure 1: Difference between lifespan and longevity.

Figure is based on data from the Historical Sample of the Netherlands (1860 – 1875). This figure illustrates the distribution of "age at death" in the form of a histogram combined with a density plot. The bars in the histogram represent the number of individuals who died at the age depicted at the x-axis. The line is a density line representing the same concept as the bars. The x-axis represents age at death groups for HSN research persons born between 1860 and 1875. The y-axis represents the number of individuals who died in the different age at death groups. The distribution depicted in this figure is used to illustrate the difference between lifespan and longevity on an individual level in terms of the place of an individual within the distribution.



Heritability of longevity has not been established yet

The broad sense heritability (H^2) of a trait can be considered as the upper limit for genetic studies, where heritability coefficients can be seen as a progress indicator, indicating whether after identification of a first gene set for a trait, additional genes may still be determined. Heritability coefficients are differentially interpreted, depending on the type of data used for analysis. When estimated in genealogical data, heritability coefficients provide an estimation of the familial influence on a trait in which the combined effects of genes and shared environment within families are difficult to separate. As a consequence, heritability estimates depend on the environmental context[16,17]. Twin studies are more suitable than other genealogical studies to provide a first estimate of the influence of genetic, shared, and non-shared environmental influences on a trait. In practice, studies often report the narrow sense heritability (h^2), which is solely based on additive effects (see *Table 3* for a summary of key quantitative genetics concepts).

Research into siblings of centenarians showed that persons with a centenarian sibling have a four to eight times higher chance of becoming a centenarian as compared to persons with a sibling who died at a normative age[8]. A study into parent – offspring relations focused on parents belonging to the top 1 percent of their birth cohort and shows that these parents have a recurrence risk of 2.31 to have children who also belong to the oldest 1 percent of their birth cohorts[19]. Similarly, long-lived parents (>95th percentile) have a greater chance of having offspring who also live up to the 95th percentile or above[20]. Consistent with these findings, it has been shown that siblings of long-lived sib-pairs (men 89+ and women 91+), their parents, and their offspring live significantly longer than members of their own birth cohorts[21].

Table 1: Phenotypes of survival

Longevity and lifespan can be measured in multiple ways, mainly depending on specific research questions, time frame, and available information. Below the most common measures are set forth within the dichotomous framework of lifespan and longevity.

	Lifespan	**Longevity**
1. Age at death	X	-
2. Age at death > threshold	X	X
3. Cohort specific top X percentile	-	X

"X" means present and "–" means not present

age (at death): The most basic definition of lifespan refers to an individual's age at last observation or if known, the age at death. The age at death refers to the complete lifespan of an individual (Lutz et al., 2013). An advantage of using this definition is that it is easy to construct this measure and there is only little data loss.

age (at death) > certain threshold: This definition refers to the age at last observation or at death after surviving passed a specific age threshold. The advantage of this definition is that certain age specific biases can be controlled for by excluding individuals below the age threshold. Early life effects are often accounted for by using a lower limit age threshold of >15 or >30 years, whereas later life effects are often accounted for by using >90, or >100 years as an upper age threshold. Whether this measure represents lifespan or longevity depends on the height of the upper age threshold and its operationalization.

cohort specific top x percentile: This definition refers to the x percent most long-lived individuals depending on the cohort specific age at death distribution. The main advantage of this measure is that it can be used to eliminate the effects of secular trends.

Spouses of nonagenarian siblings did not show a survival advantage in the study of Schoenmaker et.al. (2006). Pedersen et al., however, did observe a survival advantage for spouses of long-lived siblings when comparing them to a birth cohort and sex matched control group. The authors attribute this survival advantage to assortative mating in their population[9]. An earlier Quebec study also reported a survival advantage of spouses[23] and a study of Southern Italy found male nonagenarians to outlive their spouses, whereas this was not the case for female nonagenarians[24]. Clearly, biological, environmental, and cultural factors influence survival to advanced ages in longevity families. These genealogical studies did not provide a quantification of the effects in terms of heritability estimates.

Several genealogical studies have attempted to estimate the heritability of lifespan and longevity (see supplemental data for a description of genealogical data). These studies can be divided into two categories based on the type of data they used; (1) twin data and (2) pedigree data. Unlike animal studies in a lab setting, the effects of the environment on longevity in human studies cannot be controlled. In twins at least the variation in early environment is minimized as compared to other family based studies. In all cases, heritability estimates and the effect of specific gene variants on lifespan and longevity depends on the populations studied and their past and present environmental conditions.

Twin studies

Twin studies have shown that genetic influences account for 1-27% of lifespan variation in the population (the overall heritability (h2 and H2) is between 0.01-0.27)[25–29]. In these studies minimum age thresholds were used, ranging from 15 to 37 years. Overall, twin studies rigorously differ, besides the variability in age thresholds, in their methodology, sample selection, and design. For example, a number of studies are unable to correctly establish twin zygosity[30]. Other studies result in inaccurate and overestimated heritability coefficients because they suffer from small sample sizes, censoring and truncation problems[26,31]. Taking these issues into account, we consider the twin studies of McGue et al., Herskind et al., and Ljungquist et al.[27–29], as the most robust (see *Table 2*).

McGue and colleagues estimated a heritability of 0.22 in a Nordic European twin sample of cohorts born between 1800 and 1950. They have found a minor and non-significant difference between men (H^2 = 0.23) and women (H^2 = 0.21) for lifespan, using an age threshold of 15 years. They have used structural equation modelling techniques to compare the fit of different models and concluded that there was significant evidence for non-additive effects and in particular for intra-locus interactions (dominance). Based on this dominance model a broad sense heritability coefficient of 0.22 was estimated which was larger than the heritability component for the additive model (h^2 = 0.13)[28]. These results have been replicated in the more recent study of Herskind et al. who came to the same conclusion, although the differences between the additive and the dominance model were more modest[29]. In addition, only one study distinguishes between twins reared together and twins reared apart, acknowledging the relevant environmental effects[32], which may limit the findings resulting from twin research[27]. The study has shown that the narrow sense heritability of lifespan beyond the age of 37 is 0.01 for men and 0.15 for women. However, these estimates are limited owing to low sample sizes for twins reared together (n_men = 82 and n_women = 97 pairs). Overall, the heritability of lifespan seems to be low and likely below 0.23.

Table 2: Overview of twin and genealogical heritability studies for lifespan and "longevity"

Twin studies

Study/population	Age	Method	Total n	h^2	Men n	h^2	Women n	h^2	Time span born	Phenotype	Ref.
Swedish twin registry | Sweden	37+	ICCs and SEM	358		164	0.01	194	0.15	1868 - 1925	Age at death (lifespan) and IMRs	27
Danish twin registry | Denmark	15+	ICCs based on ANOVA and SEM	1200	0.22| 0.33	652	0.23| 0.36	548	0.21| 0.32	1870 - 1880	Age at death in years and percentiles (lifespan)	28
Danish twin | Denmark	15+	ICCs based on ANOVA and SEM	5744		2816	0.26	2928	0.23	1870 - 1900	Age at death (lifespan)	29
GenomEUtwin | multiple countries	15+	Corrected ICCs	9334		4598	0.22	4736	0.32	1870 - 1910	Age at death (lifespan)	25
NAS-NRC twin registry | U.S.	19+	Corrected ICCs	31848		31848	0.54			1946 - 1978	Age at death (lifespan)	26

Pedigree studies

Study/population	Age	Method	Total n	h^2	Men n	h^2	Women n	h^2	Time span born	Phenotype	Ref.
MICROS study | Italy	50+	Variance components analysis	8277	0.15	4299	0.16	3978	0.18	1658 - 1907	Age at death (lifespan)	33
Genealogica sursilliana | Finland	15+	REML mixed-model	2614		1226	0.18	1388	0.17	1745 - 1903	Age at death (lifespan)	34
UPDB | US	30+	ANOVA with ML	14618	0.18	7601	0.14	7017	0.22	1850 – 1913	Excess longevity (lifespan)	35
UPDB | US	65+	Correlations (Rao model)	78994	0.15					1870 - 1907	Excess longevity (lifespan)	19
Royal & noble families | Europe	30+	Multiple linear regression	12150		8409	0.18	3741	0.20	?	Age at death (lifespan)	36
Villiage genealogies | Germany	0+	Correlation analysis for	9979		5315	0.20	4664	0.19	1650 - 1925	Age at death (lifespan)	37
OOA | US	30+	Variance components analysis	1655	0.25					1727 - 1890	Age at death (lifespan)	38
Valserine Valley | France	55+	ANOVA	1102	0.27					1745 - 1849	Age at death (longevity)	39
Village of Arthez d'Asson | Canada	20+	Correlations (Tau model)	2446	0.17					1686 - 1899	Age at death (lifespan)	40
OOA | US	30+	Variance components analysis	1655	0.25					1749 - 1890	Age at death (lifespan)	41

Heritability in all twin studies is based on differences between mono and dizygotic twins and heritability in all genealogical studies are based on parent offspring correlations. In twin studies we reported the broad sense heritabilities whereas narrow sense heritabilities are reported for the pedigree studies. For twin studies additional heritabilities are provided in the text. The Danish twin registry study from Denmark shows different heritability estimates. The top value in each column refers to "age at death in years" and the bottom value in each column refers to "age at death in percentiles". methods: 1.ICCs, 2.SEM, 3.ICCs based on ANOVA, 4.Variance components analysis, 5.REML mixed-model, 6.ANOVA with ML, 7.Correlations (Rao model), 8.Multiple linear regression, 9.Correlation analysis, 10.ANOVA, 11.Correlations (TAO model). List of abbreviations: intraclass correlation (ICC), analysis of variance (ANOVA), structural equation modelling (SEM), restricted maximum likelihood estimator (REML), Maximum likelihood (ML), Integrated Mortality Risk (IMR), age at death in years (AAD yrs), age at death in percentiles (AAD pct), excess longevity = EL. For an overview of the RAO and TAO model see the papers of Rao, et. al. and Cloninger et. al.[42,43]. All phenotypes are lifespan based, except Cournil et al. (2000) which is based on longevity. studies are prioritized, with the top rows showing the highest quality studies. twin studies: based on sample size, censoring, truncation, rearing, zygosity, and study design. pedigree studies: based on sample size, generalizability, and study design.

Table 3: Short introduction into quantitative genetics

A phenotypic trait can follow a Mendelian and a non-Mendelian inheritance pattern. If a Mendelian inheritance pattern is followed, the trait originates from the effect of only one gene and it is considered to have a discrete variance. However, most traits do not originate from only one gene and thus follow a non-Mendelian inheritance pattern. Such traits have a continuous variance and examples are: Intelligence and longevity.

Quantitative genetics focuses on mapping this continuous variance, distinguishing between additive and non-additive variance. Non-additive variance may be the result of gene interactions among gene effects either within (dominance) or between (epistasis) gene loci. Non-additive effects can be determined by establishing (dis) concordance between twins. Because of this, additive genetic variance always refers to genes directly transmitted from parents to their progeny.

A phenotypic trait is not only influenced by genetic effects but also by environmental factors. For example: monozygotic twins share 100 percent of their genes but as they age, they will phenotypically differentiate because of an accumulation of personal experiences and exposure to environmental factors.

In quantitative genetics the sum of variance for a phenotypic trait is designated as follows:

$$\sigma_p^2 = \sigma_g^2 + \sigma_e^2$$

Where σ_p^2 is the total phenotypic variance, σ_g^2 is the genetic variance and σ_e^2 is the environmental variance. The phenotypic variance can also be further broken down:

$$\sigma_p^2 = \sigma_a^2 + \sigma_d^2 + \sigma_i^2 + \sigma_e^2$$

Where σ_p^2 again is the total phenotypic variance, σ_a^2 is the additive variance, σ_d^2 is the variance due to dominant effects, σ_i^2 is epistatic variance, and σ_e^2 is environmental variance. The heritability of a phenotypic trait in a population (H^2) represents the amount of phenotypic variance explained by genetic differences. In its broadest meaning the heritability is given by:

$$H^2 = \frac{\sigma_g^2}{\sigma_p^2}$$

Where H^2 is the broad sense heritability, σ_g^2 is the is the genetic variance and σ_p^2 is the total variance. The heritability coefficient varies between 0 and 1 because the numerator in the fraction is smaller than the denominator and both are positive values. Selection can only affect additive genetic variance because dominant and epistatic components are broken by processes of recombination and independent segregation. Hence, a more strict definition of heritability is often used:

$$h^2 = \frac{\sigma_a^2}{\sigma_p^2}$$

Where h^2 is the narrow sense heritability, σ_a^2 represent the additive genetic effects, and σ_p^2 is the total variance.

In an attempt to investigate the heritability of surviving to advanced ages, Ljungquist et al. have estimated the heritability at different age cut-off values[27]. In this analysis the narrow sense heritability increased with age up to 0.28 in 80+ men and 0.23 in 85+ women which may be considered extreme ages (authors denote this as 'longevity') for the investigated birth cohorts (1886 – 1925). However, sample sizes at these extreme ages were small, and negative statistically insignificant heritability coefficients were estimated in the analysis for men at the age of 85 and for women at the age of 90, indicating statistical power problems. Moreover, it remains elusive whether the increase in heritability with age is statistically significant as this is not illustrated in the study[27]. Hence, compelling results have been obtained with regard to the heritability of lifespan, though the extreme heterogeneity in heritability estimates between studies may indicate that heritability estimates are strongly influenced by study size and environmental factors. The heritability of longevity has however not been robustly estimated as yet. Consequently, future heritability studies should make a more robust assessment of the role of the environment and of longevity as a trait.

Pedigree studies

Pedigree studies overall suffer from comparable issues as twin studies regarding small sample sizes, methodology, sample selection, and design issues. In pedigree studies the heritability of lifespan generally does not exceed 0.27[19,33–41] with the larger studies estimating the heritability to be below 0.20. Pedigree studies base their estimates on parent - offspring correlations (also indicating that they tend to report narrow sense heritabilities), which impedes the estimation of heritability coefficients that are less influenced by environmental effects. Hence, pedigree research is often conducted in extremely homogenous populations, such as the village of Arthez d'Asson, as environmental influences are relatively constant in such populations[36,40,41]. However, studies conducted in such homogenous populations have limited generalizability. An important benefit of pedigree studies over twin studies is the possibility of having access to a much larger sample size, especially for the older members of a population. Taking all these aspects into account, we consider four studies as the most accurate and robust[19,33–35]. These will be discussed in more detail.

Two studies from Utah have shown an estimated heritability for lifespan of 0.15 and 0.18 for persons above 65 and 30 respectively[19,35]. Another inquiry evinces that the heritability of lifespan for persons above 15 years is 0.18 and if stratified by sex it is 0.19 for men and 0.17 for women, although this difference is not statistically significant[34]. Another, elaborate study, conducted in three semi-isolated populations in Italy shows that the heritability of lifespan is 0.16 for men and 0.18 for women at the age of 50 and beyond. The joint heritability is estimated to be 0.15 and all estimates are corrected for confounding environmental effects. Moreover, this research illustrates that the heritability of lifespan above 50 years is constant during the 17th and 18th centuries and across different populations. This same study imposes different age thresholds for lifespan and concludes that the heritability increases with age at death to a maximum of 0.35. However, the heritability drops below 0.35 at the highest age thresholds and this is likely a function of small sample sizes. Furthermore, the study does not provide statistics of the increase in heritability estimates and besides that, ages at death were transformed into standardized scores which are difficult to interpret in relation to actual ages at death[33]. The heritability of lifespan seems comparable in pedigree and twin studies; it does not exceed 0.27. In pedigree studies, the heritability of longevity has been under-investigated and consequently, comparable to twin studies, the heritability of longevity has not been robustly estimated in pedigree studies. Moreover, pedigree studies also show a large variation in

heritabilities (0.15 – 0.27), which may be attributed to study size, selection criteria, and variation in environmental factors.

Longevity

The heritability of lifespan has been well documented by means of twin studies and pedigree studies, and it can be concluded that the heritability of lifespan is between 0.01 and 0.27 in the population at large. The large variation in the heritability estimates indicates a prominent role for differential environmental influences on the estimates. Studies showing that siblings of centenarians and longevous sib-pairs have a high probability to also become a centenarian or longevous, respectively, and studies, which show that longevous parents have a high probability to bear longevous offspring, provide indications that the heritability of longevity may be higher than that of lifespan[8,19–21]. However, the heritability of longevity has only been investigated once in a twin study design, though of limited sample size[27]. In addition, the heritability of longevity has been investigated more often in pedigree studies but the studies raise several questions about their design, sample size, and generalizability. Establishing the heritability of longevity is necessary for case definitions in genetic studies focused on gene mapping[21]. Hence, researchers' attention should shift from lifespan to longevity and the heritability of longevity should be estimated in an appropriate design with a sufficiently large sample. Both the heritability of lifespan and longevity should be investigated in different environments to investigate environmental influences.

Historical genealogical data in inheritance pattern research

Inheritance patterns of any complex trait generally provide insight into the contributions attributable to shared genes. Longevity is expected to be a complex trait with a complicated inheritance pattern, resulting from interactions between the environment and many genes[11–15]. Such effects may be additive or non-additive where one gene may be rate limiting over the action of another, or enhance or multiply the effect of another gene. A traits' genetic inheritance pattern can be investigated using historical genealogical data. In the context of survival to extreme ages the inheritance pattern has often been investigated by estimating correlations between the lifespan of parents and children[19,40,44–47] and, stratifying these correlations by sex[19,37,40,45–47]. The inheritance pattern of survival to extreme ages had also been investigated with survival analysis, logistic regression, and analysis of variance instead of basic correlations[20,39,48–55].

Table 4: Inheritance pattern studies using genealogical data

Study	Age	Method	Time-span	Sample size	Phenotype	Parent-offspring	Father-son	Father-daughter	Mother-son	Mother-daughter	Siblings / cousins
51	90+ versus 50-75	Linear and logistic regression	1850 -1884	1152	Lifespan and longevity	+	X	X	+	+	-
37	50+	Correlation	1650 - 1927	9979	Lifespan	+	X	X	+	+	-
48	Parents >90 Offspring >20	Survival analysis	1922 - 1930	8869	Lifespan	+	-	-	+	+	-
47	0+	Correlations	1860 - 1909	12317	Lifespan	-	+	X	+	+	-
44	0+	Correlation	1652 - 1850	32897	Lifespan	+	-	-	-	-	+
52	Parents >50 versus <50 Offspring 55+	ANOVA	1700 - 1976	2172	Lifespan	+	X	X	X	X	+
45	30+	Correlation	1800 - 1880	?	Lifespan	+	+	X	+	+	-
19	65+	Correlations	1870 - 1907	78994	Lifespan	+	-	-		-	+
56	30+	Regression	1800 - 1880	17136	Lifespan	-	+	+	+	+	-
53	Parents ? (>0) Offspring > 75	Logistic regression	700 - 1875	1132	Lifespan	-	X	+	X	+	-
39	Parents >50 versus <50 Offspring >55	ANOVA	1745 - 1849	3733	Lifespan	+	X	+	X	X	-
20	95th percentile	Logistic regression	1870 - 1900	1531	Longevity	+	-	-	-	-	+
40	>0, >20, and >50	Correlation	1686 - 1899	11102	Lifespan	-	+	+	+	+	+
49	50+	Cox regression	1850 - 1910	874	Lifespan	+	X	X	X	+	+
46	0+	Correlations	1820 - 1899	200	Lifespan	+	X	X	X	+	+
54	Parents >70 and > 80 Offspring <80 and >100	Logistic regression	1818 - 1898	8585	Longevity	-	+	X	X	+	-
55	Offspring 100+ Parents ?	Logistic regression And linear regression	?	1018	Longevity and lifespan	-	+	X	+	+	-
50	>0	Cox regression	<1919	759150	Lifespan	-	+	+	-	-	-
57	100+	Descriptives	1889 - 2003	207	Longevity	+	X	X	+	+	-
58	Parents 60+ Offspring <75 versus >90	T-tests	1887	193	Lifespan	-	+	+	X	X	-

-"+" ⟶ relation is found ; "-" ⟶ relation is not tested ; " X" ⟶ relation is not found. studies are prioritized with the top rows showing the highest quality studies. The prioritization is based on sample size, study design and population, control variables, and statistical analyses. phenotypes: 1=lifespan, 2=longevity. Methods: 1=linear regression, 2=logistic regression, 3=correlation , 4=survival analysis, 5=ANOVA, 6=descriptives, 7=T-tests.

Apart from the variety of analytical methods used in the literature, inheritance pattern research is very heterogeneous with regard to study designs. Most research has a cross-sectional nature using either a multiple cohort or a case – control design, in which a group of old persons is compared to a control group, over two generations[19,20,37,39,40,44–58]. Moreover, most studies focus on lifespan instead of longevity[20,53–55,57] and they often involve entire populations which are either extremely homogenous[44] or heterogeneous[19], depending on the research question. Homogenous populations suffer from generalizability problems whereas heterogeneous populations are difficult to analyze because of a larger amount of environmental variance and founder effects. Furthermore, a minimum of two generations should be available to conduct analyses (parents and their offspring). In practice a more than two generational approach has almost never been applied.

Patterns of inheritance

The main results of a range of pedigree studies are shown in *Table 4*. First, many studies have found evidence for a father – son inheritance pattern[19,40,45,47,50,54,55,58] although an equal number of studies has not found this evidence[37,39,49,51–53,57]. The same pattern can be observed for mother – son inheritance and the least evidence seems to point in the direction of a father – daughter inheritance pattern. Most evidence points in the direction of a mother – daughter pattern of inheritance with twelve confirming studies[37,40,45–51,53,55–57] and only three disconfirming studies[39,52,58].

Most of the evidence is not compelling because of persistent challenge of establishing a genetic inheritance pattern which is uninfluenced by environmental factors (e.g. Socio-economic status, mothers age at birth, and the physical environment)[33,44,46]. Furthermore, secular trends, caused by the increased average lifespan through improved nutrition, hygiene, and medical treatment, are important factors when comparing the lifespan of parents and their offspring[33,38,45,46]. Many attempts to control for secular trends and environmental factors by applying specific statistical techniques, including control variables, focusing on homogenous populations, and excluding infant and child mortality from the sample have not led to consistent results[19,33,37,44,46,52]. A few studies attempted a different approach by focusing on longevity instead of lifespan[20,54,55,57]. However, these studies did not examine more than two generations and focused on extremely homogenous populations. As a consequence the generalizability of their results may be limited. Thus, results of inheritance pattern studies have been largely inconsistent and strong differences exist between studies. A few studies stand out given their sample size and design, population, control variables, and statistical analyses and will be examined further here[37,47,51,55].

Kemkes-Grottenthaler (2004) found a significant correlation between maternal lifespan (lower limit age threshold 50 years) and the lifespan of sons and daughters by studying genealogies of two historical homogenous German villages during 1412 – 1912. Correlations between paternal lifespan and the lifespan of sons and daughters has also been estimated but no significant relationship was found[37]. Similarly, the study of Parman (2010) provided evidence for a correlation between maternal lifespan and the lifespan of sons and daughters. In contrast to Kemke-Grottenthaler (2004) the Parman (2010) study focused on the heterogenous setting of North Carolina during 1860 – 1909. The study included more than 12,000 individuals[47]. Deluty, Atzmon, Crandall, Barzilai, and Milman (2015) focused on a combination of lifespan and longevity instead of only lifespan. The authors defined longevity as being 100 years or above and focused

on 291 centenarians and their parents in the contemporary homogenous society of Ashkenazi Jews in the United States. They concluded that mothers of longevous men and women had significantly longer lifespans as compared to mothers of non longevous individuals. An attenuated but similar pattern could be seen for fathers although the lifespan of fathers did not differ between longevous and non longevous daughters. In addition, logistic regression models indicated that the odds of having longevous offspring increased for every 10 years of life achieved for mothers whereas this is not the case for fathers[55]. Lastly, a study focused on the homogenous population of Saguenay-Lac-St-Jean. The study investigated the familial transmission of longevity in a group of 576 individuals aged over 90 years as compared to an equally sized control group aged between 50 and 75 years in the time frame between 1950 and 1974. It was concluded that the probability of having longevous offspring (both boys and girls) was elevated with an increase in mothers' lifespan and that this is not the case for fathers[51]. Overall, there are several indications for maternal transmission of lifespan with some preference to daughters over transmission to sons. Furthermore we conclude that studies of the inheritance of longevity over multiple generation remain limited.

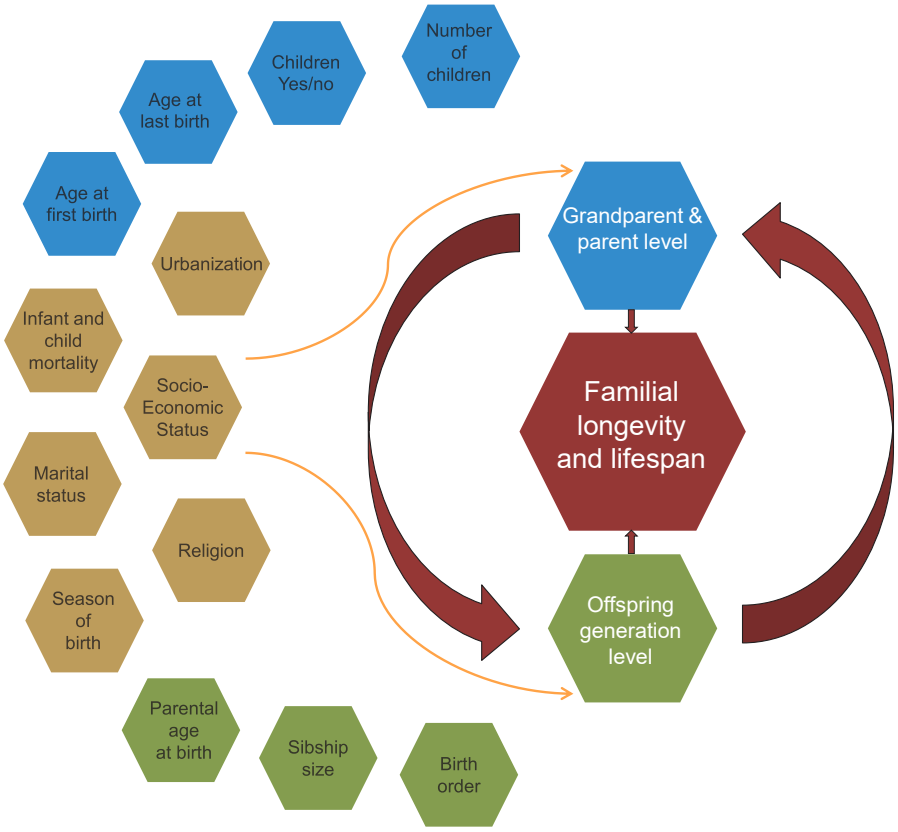


Figure 2: Covariates and confounders of lifespan and longevity.

The figure consist of two half circles. The outer rim mainly represents within family factors whereas the inner rim represents outside family factors. The blue color factors have a direct influence on the parental level (also depicted in blue). the green color factors have a direct influence on the offspring level (also depicted in green). The light-brown color predictors have an influence on both the parental and offspring levels which is illustrated by the arrows. The red arrows around the parental and offspring level indicate intergenerational effects of the factors illustrated in the left of the figure. The combining factor is the familial level of lifespan and longevity, which is depicted in the same red color as the arrows.

Up to now, the inheritance pattern of longevity has hardly been studied, with only two of the discussed studies estimating the (sex specific) effect of parental lifespan on the probability of having longevous offspring[51,55]. This is however not an optimal design to determine the inheritance pattern of longevity. Furthermore, longevity, is likely a polygenic trait, influenced by many environmental factors with small effects[11–15]. Hence, lifespan and longevity inheritance patterns may be influenced by environmental factors[52]. *Table 4* illustrates the large heterogeneity in inheritance pattern outcomes between studies. The study of Matthijs and colleagues (2002) is an example of how inheritance patterns may be influenced by the environment. The authors show that the inheritance pattern of lifespan in Moerzeke (Flanders) is different compared to a couple of Jura villages (France) using exactly the same methodological approach in a comparable time period (1700 - 1900)[39,52]. Ideally, multiple generations of families with a strong family history of extreme survival should be studied, which may reveal the interaction with environmental factors and may contribute to clarifiy the inheritance patterns by which longevity is transmitted.

Environmental influences in longevity research

Environmental factors such as socio-economic status, sibship size, parental age at birth, and geographical origin, reflecting exposure to epidemics, famines and war, are important variables within lifespan research[59–61]. This is because environmental factors can covary with and modify the lifespan of parents and children[18,39,52]. These factors can also confound the statistical relation between parents and their offspring, with respect to survival[19]. Longevity is derived from lifespan *(see Table 1)* and thus it can be expected that the same environmental factors which influence the results of genealogic research into lifespan also affect longevity research. In fact, some evidence for this exists but genealogical longevity research is scarce and sample sizes are generally small[54,62–67]. For example, one study found that environmental factors such as birth order and age at last birth slightly affected the relationship between parents and offspring longevity, defined as belonging to the oldest 5% of a person's birth cohort[54]. Hence, it is important to take environmental factors into account when inquiring into longevity, and because of this, we will outline the most important ones (see *Figure 2* for an overview).

Reproductive factors

Reproductive aging factors play a vital role in lifespan and longevity research[33,63–65,67–74]. The influence of reproductive factors on lifespan and longevity can take place on two levels; the level of the grandparents/parents and that of their offspring/subsequent generations. On the grandparents/parents level the following parameters will be described: Having children yes or no, parental age at first and last birth, and number of children. On the offspring/subsequent generations level, parental age at birth and birth order will be described.

Level of the grandparents/parents

The disposable soma theory suggest a biological trade-off between energy investment in reproduction and somatic maintenance[63,66,68,75–80]. This trade-off implies that an increase in the number of children causes a decrease in maternal lifespan. Such evolutionary trade-off has indeed been found in the study of laboratory animal models[81]. Just as the disposable soma theory, the maternal depletion theory suggests a trade-off between the number of children and maternal lifespan, although the theoretical mechanism is somewhat different[72,82]. In the maternal depletion theory the trade-off between number of children and maternal lifespan is explained by the emotional and physical investment of upbringing, and not necessarily a biological trade-off[72]. Hence, the maternal depletion theory also explains a paternal trade-off between the number of children and lifespan. In contrast to the maternal depletion and disposable soma theory, it is theorized that an increase in age at last birth is associated with an increase in maternal lifespan. One mechanism for this effect is that age at last birth may be a marker for general health and aging. Healthy aging persons may be predisposed to have slow aging tissues, which may subsequently cause the ability to reproduce late in life[83].

The theories that are described above have been extensively tested with genealogical data in natural fertility populations of various sample sizes, ranging from less than 100 to more than 10,000. One study found that on average women with children lived longer than women without children[84], but another study has not found evidence for this effect[85]. A few studies show that women without children reach older ages than women with children[71,86]. Similarly, many studies have

shown an increase in maternal lifespan if the number of children decreased[63,68,70,73,76,84,86–91], while only three studies found no effect at all[72,74,92] and two studies found the opposite effect[93,94]. For men, however, the relation between number of children and lifespan was inconsistent[63,66,68,74–76,87]. When it comes to mothers' age at last child, research has shown that an increase in age at last child is associated with an increase in maternal lifespan[75,80,83,89,94–96] and only two researchers found no link[68,76]. On the one hand, studies showed that if maternal age at birth increases, maternal lifespan equally increases[68,71,76,97]. Evidence for such an effect has also been provided for fathers[66,75]. On the other hand, studies have also shown negative effects for the relation between maternal age at birth and maternal lifespan6[3,98] or no relation at all[64,72,92]. All these reproductive effects on lifespan have typically been investigated for lifespan beyond 50 years in order to control for early deaths caused by childbearing.

For longevity, Tabatabaie (2011) showed that an increase in number of children correlates with a decrease in the odds of becoming 100+[63]. Tabatabaie et al. (2011) and sun et al. (2015) also showed that the odds of becoming longevous, defined as 100+, increase as the age at last child increases[63,65].

Level of the offspring generations

The human mutation rate of DNA base substitutions is high and increases with chronological age[99]. As a result, deleterious mutations in germ cells may cause a decrease in the lifespan of offspring as the parental age at conception increases[100]. The resource theory explains that being among the first children in the birth order may be beneficial for a persons' lifespan. Persons among the first in the birth order tend to receive more attention from their parents and do not have to share resources with multiple siblings[69].

A minority of studies focused on reproductive factors on the offspring level. The effects of birth order and paternal age at birth are relatively consistent. One small study showed that old fathers have daughters who die young as compared to young fathers. However, this study did not provide evidence for sons[100]. Furthermore, a large study of over 14,000 persons provided evidence that first born children live longer than those who are born later, regardless of their sex[69]. Lastly, all research on the offspring level focused on lifespan and no study looked into the effects on longevity.

Additional factors

Besides reproductive factors, other factors are also of significant importance for lifespan and longevity research. Persons with a high socio-economic status (SES) have a longer lifespan and a higher probability to become longevous than persons with a low SES[62,101–103]. This can mainly be attributed to the fact that high SES persons have better access to clean drinking water, high quality health care, and nutrition[13,95,104,105]. Furthermore, the season in which individuals are born has been shown to be an important measure. The effect of seasonality may be attributed to seasonal periods which encompass more danger for infections than others. Studies showed that the best months of birth are September until November[33,60,101]. Religion and marital status also influence lifespan. Religion is associated with a healthy lifestyle, causing religious persons to live a longer and healthier life than non-religious persons. Married persons also have a healthier lifestyle, explaining why they live longer and healthier than non-married persons[62,101]. Another important factor is the degree of urbanization, as urbanized areas have higher population densities than rural areas, making such areas

more susceptible to the distribution of diseases[101,106]. Infant mortality has also shown to associate with the survival of mothers, as infant mortality can be considered a proxy for maternal health[107,108]. Furthermore, both infant and child mortality can be considered proxies for children's early life circumstances[35]. Hence, not only reproductive factors are important for lifespan and longevity research, but also SES, marital status, religion, the degree of urbanization, and infant/child mortality.

Gene-environmental interactions in longevity research

Genetic effects in longevity studies are always influenced by different environmental factors[11–15] which can have additive and modifying influences. An example of additive influence is demonstrated if an individual is longevous because of a high SES and parental care, because the person was a first born. In this example, the modifying role of environmental factors is that the effect of some hypothetical gene set associated with longevity is more likely to become expressed in the phenotype when you are a firstborn child as compared to not being the firstborn child. The modifying nature of environmental factors can at best be addressed once genetic loci associating with longevity have been identified. In contrast, the additive nature of environmental factors can be used to screen potentially interesting persons for genetic analysis (this will be explained in more detail in chapter 5). In this regard genealogical data may provide new opportunities to record environmental effects which would otherwise have remained unknown in genetic studies.

Conclusions and future perspectives

In this review we focused on summarizing genealogical studies, and the beneficial role of genealogical studies and data for genetic longevity research. We conclude that lifespan and longevity research is very heterogeneous with regard to study designs, analytical methods, study populations, and results. This heterogeneity is problematic for comparative research. In addition, many studies have misused and misinterpreted the term longevity as it often refers to the complete lifespan of an individual or the lifespan beyond a certain lower limit threshold. As a result, many studies have investigated the heritability of lifespan instead of longevity. Irrespective of the twin or pedigree study design, the heritability estimates of lifespan are between 0.01 and 0.27 in the population at large, with an average of 0.25 (see *Table 2*). Inheritance pattern research has likely found evidence consistent with maternal transmission of both lifespan and longevity. This pattern has been identified on the basis of two generational analyses, which is a relatively weak design to identify a pattern of inheritance. When taking all inheritance pattern studies into account there is a large heterogeneity between the study results (see *Table 4*). As a next step, we suggest research into lifespan and longevity to take a standard minimum number of environmental factors into account (see *Figure 2*). Moreover, environmental factors in historical, demographic, and genealogical multi-generational data can be used to gain insight in the individual and family history of potentially interesting individuals for discovery genetics, such as by NGS. Selection of the most informative families and cases for these studies increases the probability to find longevity genes and one may gain insight in the differential role of the environment for specific gene variants. In conclusion, many studies have been conducted using different methodologies and focusing on different definitions of longevity. Hence, much knowledge has been gained from genealogical studies with regard to lifespan, though little is known about longevity and environmental influences (either additive or modifying) have been largely neglected in genetic lifespan and longevity research. Because of this, we propose a new approach and recommend integrating insights from genealogical studies into genetic studies to answer the still unsolved aspects of longevity.

Defining longevity in terms of the family over multiple generations

Lifespan and longevity are two distinct and incompatible concepts. In this paper, we defined longevity as survival into extreme old ages across an upper limit and lifespan as age at death behind a lower limit threshold such as 15, 30, or 50 years (see *Table 1*). However, in the literature, specific definitions aimed at quantifying the concept of "oldest old" are often used, where people need to have reached a certain age threshold (for example 75, 80, 85 years of age). Which persons actually are the oldest old in terms of absolute ages differs per time-period and population. By the use of absolute ages, comparisons across studies and populations become problematic and secular trends may cause extreme biases. Therefore, it has been suggested to define the age threshold for longevity at the oldest five percent of a population[10], allowing comparisons over time (including secular trends) and between populations. Limitations of taking a population percentage as age threshold for longevity come forward in certain specific study designs. When for example the oldest five percent is only 60 years of age such selection criteria will not refer to longevity. The percentiles should therefore always be weighted towards the life-tables of representative cohorts of an entire population. To sum up, lifespan and longevity are different concepts, which are preferably defined in terms of weighted percentiles instead of absolute ages.

Using the oldest five percent of a birth cohort seems appealing, but evidence that this oldest five percent of singletons indeed represents the best impact of genetic influence on longevity is not available. There is no evidence that using the oldest five percent creates new opportunities to distinguish genetic effects from environmental effects. Belonging to the oldest five percent may still harbor phenocopies caused by healthy lifestyles and excellent health care[109]. We propose to first investigate what the best definition of longevity is. This can be done by studying families over multiple generations in genealogic and demographic databases. Such data will allow the testing of different longevity definitions that reveal the most prominent familiarity in excess survival. One may compute whether an optimal familial clustering of longevity is observed for the 95th or 99th percentile surviving singletons, or for example, first degree relative-pairs (siblings, parents/offspring etc.).

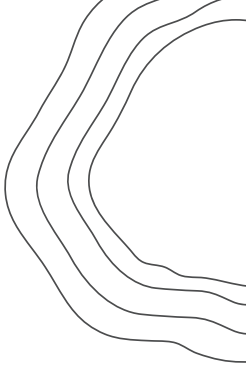
In addition to the definition problem of longevity hampering genetic research in detecting common genetic variants, research should focus on rare genetic variants in long-lived families; longevity is probably dependent on many genes with relatively small effects[11–15]. Some attempts have been made to identify rare variants contributing to human longevity by whole genome/exome sequencing of extremely long surviving individuals with as yet little robust findings[110,111]. In search for rare variants, we propose to select (in NGS efforts) families based on multiple generations of long-lived (top survival percentiles) descendants/first degree relatives, distant long-lived relatives, and to include environmental factors (such as birth order, SES, physical environment, the presence of a war or famine, and cause of death). Of course worldwide (joined) efforts will be needed for these analyses since the environment may modulate genetic effects, confining their detection.

Conflict of interest

The authors have nothing to disclose.

Funding

This work was supported by the Netherlands Organization for Scientific Research [360-53-180].



References

1 Oeppen, J. & Vaupel, J. W. Broken limits to life expectancy. *Science* **296**, 1029–1031 (2002).

2 Wachter, K. W. & Finch, C. E. *Between Zeus and the Salmon: The Biodemography of Longevity. Population (English Edition)* (Washington, DC: The National Academies Press, 1997).

3 Christensen, K. & Vaupel, J. W. Determinants of longevity: genetic, environmental and medical factors. *J. Intern. Med.* **240**, 333–341 (1996).

4 Andersen, S. L., Sebastiani, P., Dworkis, D. a., Feldman, L. & Perls, T. T. Health Span Approximates Life Span Among Many Supercentenarians: Compression of Morbidity at the Approximate Limit of Life Span. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **67A**, 395–405 (2012).

5 Ash, A. S. *et al.* Are Members of Long-Lived Families Healthier Than Their Equally Long-Lived Peers? Evidence From the Long Life Family Study. *Journals Gerontol. Ser. A Biol. Sci. Med.* Sci. **70**, 971–976 (2015).

6 Christensen, K., McGue, M., Petersen, I., Jeune, B. & Vaupel, J. W. Exceptional longevity does not result in excessive levels of disability. *Proc. Natl. Acad. Sci.* **105**, 13274–13279 (2008).

7 Evert, J., Lawler, E., Bogan, H. & Perls, T. Morbidity Profiles of Centenarians: Survivors, Delayers, and Escapers. *Journals Gerontol. - Med. Sci.* **58**, 232–237 (2003).

8 Perls, T. T. *et al.* Life-long sustained mortality advantage of siblings of centenarians. *Proc. Natl. Acad. Sci.* **99**, 8442–8447 (2002).

9 Pedersen, J. K. *et al.* The Survival of Spouses Marrying Into Longevity-Enriched Families. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **72**, 109–114 (2017).

10 Sebastiani, P., Nussbaum, L., Andersen, S. L., Black, M. J. & Perls, T. T. Increasing Sibling Relative Risk of Survival to Older and Older Ages and the Importance of Precise Definitions of "Aging," "Life Span," and "Longevity". *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **71**, 340–346 (2016).

11 Deelen, J., Beekman, M., Capri, M., Franceschi, C. & Slagboom, P. E. Identifying the genomic determinants of aging and longevity in human population studies: progress and challenges. *Bioessays* **35**, 386–96 (2013).

12 Shadyab, A. H. & LaCroix, A. Z. Genetic factors associated with longevity: A review of recent findings. *Ageing Res. Rev.* **19**, 1–7 (2015).

13 Christensen, K., Johnson, T. E. & Vaupel, J. W. The quest for genetic determinants of human longevity: challenges and insights. *Nat. Rev. Genet.* **7**, 436–48 (2006).

14 Finch, C. E. & Tanzi, R. E. Genetics of Aging. *Science (80-.).* **278**, 407–411 (1997).

15 Kirkwood, T. B. L., Cordell, H. J. & Finch, C. E. Speed-bumps ahead for the genetics of later-life diseases. *Trends Genet.* **27**, 387–388 (2011).

16 Lewontin, R. C. Annotation: the analysis of variance and the analysis of causes. *Am. J. Hum. Genet.* **26**, 400–11 (1974).

17 Rose, S. P. R. Commentary: Heritability estimates - Long past their sell-by date. *Int. J. Epidemiol.* **35**, 525–527 (2006).

18 Montesanto, A. et al. Demographic, genetic and phenotypic characteristics of centenarians in Italy: Focus on gender differences. *Mech. Ageing Dev.* (2017). doi:10.1016/j.mad.2017.04.008

19 Kerber, R. A., Brien, E. O., Smith, K. R. & Cawthon, R. M. Familial Excess Longevity in Utah Genealogies. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **56**, 130–139 (2001).

20 Gudmundsson, H., Gudbjartsson, D. F. & Kong, A. Inheritance of human longevity in Iceland. *Eur. J. Hum. Genet. EJHG* **8**, 743–749 (2000).

21 Schoenmaker, M. et al. Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur. J. Hum. Genet.* **14**, 79–84 (2006).

22 *Vienna Yearbook of Population Research 2013.* (Verlag der ¨ Osterreichischen Akademie derWissenschaften, 2013).

23 Jarry, V., Gagnon, A. & Bourbeau, R. Survival advantage of siblings and spouses of centenarians in 20th-century Quebec. *Can. Stud. Popul.* **39**, 67 (2013).

24 Montesanto, A. *et al.* The genetic component of human longevity: analysis of the survival advantage of parents and siblings of Italian nonagenarians. *Eur. J. Hum. Genet.* **19**, 882–886 (2011).

25 Skytthe, A. *et al.* Longevity Studies in GenomEUtwin. *Twin Res.* **6**, 448–454 (2003).

26 Hrubec, Z. & Neel, J. V. Familial factors in early deaths: Twins followed 30 years to ages 51?61 in 1978. *Hum. Genet.* **59**, 39–46 (1981).

27 Ljungquist, B., Berg, S., Lanke, J., McClearn, G. E. & Pedersen, N. L. The Effect of Genetic Factors for Longevity: A Comparison of Identical and Fraternal Twins in the Swedish Twin Registry. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **53A**, 441–446 (1998).

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

29 Herskind, A. M. et al. The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Hum. Genet.* **97**, 319–323 (1996).

30 Wyshak, G. Biodemography and Social Biology Fertility and longevity in twins , sibs , and parents of twins Parents of Twins. *Biodemography Soc. Biol.* **25**, 315–330 (2010).

31 Jarvik, L. F., Falek, A. & Kallmann, F. J. Survival Trends in a Senescent Twin Population. *Am. J. Hum. Genet.* **12**, 170–179 (1960).

32 Falconer, D. S. & Mackay, T. F. *Introduction to quantitative genetics.* (Pearson Education Limited, 1995).

33 Gögele, M. *et al.* Heritability analysis of life span in a semi-isolated population followed across four centuries reveals the presence of pleiotropy between life span and reproduction. *J. Gerontol. A. Biol. Sci. Med. Sci.* **66**, 26–37 (2011).

34 Pettay, J. E., Kruuk, L. E. B., Jokela, J. & Lummaa, V. Heritability and genetic constraints of life-history trait evolution in preindustrial humans. *PNAS* **102**, 2838–2843 (2005).

35 Hanson, H. A., Smith, K. R. & Hasstedt, *S. Heritability of longevity and the role of early and mid-life environments.* **022095**, (2012).

36 Gavrilova, N. *et al.* Evolution, mutations, and human longevity European royal and noble families. *Hum. Biol.* **70**, 799–804 (1998).

37 Kemkes-Grottenthaler, A. Parental effects on offspring longevity—evidence from 17th to 19th century reproductive histories. *Ann. Hum. Biol.* **31**, 139–158 (2004).

38 Mitchell, B. D. et al. Heritability of Life Span in the Old Order Amish. *Am. J. Med. Genet.* **352**, 346–352 (2001).

39 Cournil, A., Legay, J. M. & Schächter, F. Evidence of sex-linked effects on the inheritance of human longevity: a population-based study in the Valserine valley (French Jura), 18-20th centuries. *Proc. Biol. Sci.* **267**, 1021–5 (2000).

40 Bocquet-Appel, J. P. & Jakobi, L. Familial transmission of longevity. *Ann. Hum. Biol.* **17**, 81–95 (1990).

41 Sorkin, J. e*t al.* Exploring the genetics of longevity in the Old Order Amish. *Mech. Ageing Dev.* **126**, 347–50 (2005).

42 Cloninger, C. R., Rice, J. & Reich, T. Multifactorial Inheritance with Cultural Transmission and Assortative Mating II . A General Model of Combined Polygenic and Cultural Inheritance. *Hum. G* **31**, 176–198 (1979).

43 Rao, D. C., Morton, E. N. & Yee, S. Analysis of Family Resemblance. II. A Linear Model for FAmilial Correlation. *Hum. Genet.* **26**, 331–359 (1974).

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

45 Gavrilov, L. a., Gavrilova, N. S., Olshansky, S. J. & Carnes, B. a. Genealogical data and the biodemography of human longevity. *Biodemography Soc. Biol.* **49**, 160–173 (2002).

46 Philippe, P. & Opitz, J. M. Familial correlations of longevity: An isolate-based study. *Am. J. Med. Genet.* **2**, 121–129 (1978).

47 Parman, J. Gender *and Intergenerational Mobility: Using Health Outcomes to Compare Intergenerational Mobility Across Gender and Over Time*. (2010).

48 Abbott, M. H., Abbey, H., Bolting, D. R. & Murphy, E. A. The Familial Component in Longevity-A Study of Offspring of Nonagenarians: 111 . Intrafamilial Studies. *Am. J. Med. Genet.* **2**, 105–120 (1978).

49 Salaris, L., Tedesco, N. & Poulain, M. Familial transmission of human longevity : a population-based study in an inland village of Sardinia (Italy), 1850 – 2010. *Vienna Yearb. Popul. Res.* **11**, 325–349 (2013).

50 Lach, S., Riton, Y. & Simhon, A. *Longevity across generations*. (2006).

51 Houde, L., Tremblay, M. & Vézina, H. Intergenerational and Genealogical Approaches for the Study of Longevity in the Saguenay-Lac-St-Jean Population. *Hum. Nat.* **19**, 70–86 (2008).

52 Matthijs, K., Putte, B. & Vlietinck, R. The Inheritance of Longevity in a Flemish Village. *Eur. J. Popul. Eur. Démographie* **18**, 59–81 (2002).

53 Westendorp, R. G. J. & Kirkwood, T. B. L. Maternal and Paternal Lines of Familial Longevity. *Popul. An English Sel.* **13**, 223–235 (2001).

54 Danzhen You, Danan Gu & Zeng Yi. Familial Transmission of Human Longevity Among the Oldest-Old in China. *J. Appl. Gerontol.* **29**, 308–332 (2010).

55 Deluty, J. A., Atzmon, G., Crandall, J., Barzilai, N. & Milman, S. The influence of gender on inheritance of exceptional longevity. *Aging (Albany. NY).* **7**, 412–418 (2015).

56 Gavrilov, L. A. & Gavrilova, N. S. Biodemographic Study of Familial Determinants of Human Longevity. *Popul. An English Sel.* **13**, 197–221 (2001).

57 Caselli, G. *et al.* Family clustering in Sardinian longevity: A genealogical approach. *Exp. Gerontol.* **41**, 727–736 (2006).

58 Hallett, M., Human, R. a T. & Physiology, C. familial longevity: the older you are the older your father may have been. *J. Am. Geriatr. Soc.* **53**, 357–358 (2005).

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

60 Gavrilov, L. a. & Gavrilova, N. S. New Developments in the Biodemography of Aging and Longevity. *Gerontology* **61**, 364–371 (2014).

61 Le Bourg, É. Does reproduction decrease longevity in human beings? *Ageing Res. Rev.* **6**, 141–149 (2007).

62 Palmore, E. Predictors of successful aging. *Gerontologist* **19**, 427–431 (1979).

63 Tabatabaie, V. *et al.* Exceptional longevity is associated with decreased reproduction. *Agign* **3**, 202–205 (2011).

64 Robine, J., Cournil, A., Henon, N. & Allard, M. Have centenarians had younger parents than the others? *Exp. Gerontol.* **38**, 361–365 (2003).

65 Sun, F. et al. Extended maternal age at birth of last child and women's longevity in the Long Life Family Study. *Menopause* **22**, 26–31 (2015).

66 Helle, S., Käär, P. & Jokela, J. Human longevity and early reproduction in pre-industrial Sami populations. *J. Evol. Biol.* **15**, 803–807 (2002).

67 Jarry, V., Gagnon, A. & Bourbeau, R. *Parental Age at Birth and Longevity of Offspring in Centenarian Families: The role of biology, social interaction and culture.* (2012).

68 Doblhammer, G. & Oeppen, J. Reproduction and longevity among the British peerage: the effect of frailty and health selection. *Proc. Biol. Sci.* **270**, 1541–1547 (2003).

69 Modin, B. Birth order and mortality : a life-long follow-up of 14,200 boys and girls born in early 20th century Sweden. *Soc. Sci. Med.* **54**, 1051–1064 (2002).

70 Westendorp, R. G. J. & Kirkwood, T. B. L. Human longevity at the cost of reproductive success. *Nature* **396**, 743–746 (1998).

71 Lycett, J. E., Dunbar, R. I. M. & Voland, E. Longevity and the costs of reproduction in a historical human population. *Proc. R. Soc. London. Ser. B Biol. Sci.* **267**, 31–35 (2000).

72 Alter, G., Dribe, M. & Poppel, F. Van. Widowhood, Family Size, and Post-Reproductive Mortality: A Comparative Analysis of Three Populations in Nineteenth-Century Europe. *Demography* **44**, 785–806 (2007).

73 Harrell, C. J., Smith, K. R. & Mineau, G. P. Are Girls Good and Boys Bad for Parental Longevity? *Hum. Nat.* **19**, 56–69 (2008).

74 Hurt, L. S. *et al.* Long-term effects of reproductive history on all-cause mortality among adults in rural Bangladesh. *Stud. Fam. Plann.* **35**, 189–196 (2004).

75 McArdle, P. F. *et al.* Does having children extend life span? A genealogical study of parity and longevity in the Amish. *J. Gerontol. A. Biol. Sci. Med. Sci.* **61**, 190–195 (2006).

76 Mueller, U. Does late reproduction extend the life span? Findings from European royalty. *Popul. Dev. Rev.* **30**, 449–466 (2004).

77 Kirkwood, T. B. & Rose, M. R. Evolution of senescence: late survival sacrificed for reproduction. *Philos. Trans. R. Soc. Lond. B. Biol. Sci.* **332**, 15–24 (1991).

78 Bowen, R. L. & Atwood, C. S. Living and dying for sex: A theory of aging based on the modulation of cell cycle signaling by reproductive hormones. *Gerontology* **50**, 265–290 (2004).

79 Atwood, C. S. & Bowen, R. L. The reproductive-cell cycle theory of aging: An update. *Exp. Gerontol.* **46**, 100–107 (2011).

80 Perls, T. T., Alpert, L. & Fretts, R. C. Middle-aged mothers live longer. *Nature* **389**, 133 (1997).

81 Robins, C. & Conneely, K. N. Testing evolutionary models of senescence: traditional approaches and future directions. *Hum. Genet.* **133**, 1451–1465 (2014).

82 Merchant, K. & Martorell, R. Frequent reproductive cycling: Does it lead to nutritional depletion of mothers? *Prog. Food Nutr. Sci.* **12**, 339–369 (1988).

83 Crawford, S. L. What explains the link between reproductive events and women's longevity? *Menopause* **22**, 6–8 (2015).

84 Lund, E., Arnesen, E. & Borgan, J. K. Pattern of childbearing and mortality in married women--a national prospective study from Norway. *J. Epidemiol. Community Health* **44**, 237–240 (1990).

85 Gavrilova, N. S., Gavrilov, L. A., Semyonova, V. G. & Evdokushkina, G. N. Does Exceptional Human Longevity Come with a High Cost of Infertility. *Ann. N. Y. Acad. Sci.* **1019**, 513–517 (2004).

86 Friedlander, N. The relation of lifetime reproduction to survivorship in women and men: A prospective study. *Am. J. Hum. Biol.* **8**, 771–783 (1996).

87 Dribe, M. Long-term effects of childbearing on mortality: evidence from pre-industrial Sweden. *Popul. Stud. (NY).* **58**, 297–310 (2004).

88 Helle, S., Lummaa, V. & Jokela, J. Sons reduced maternal longevity in preindustrial humans. *Science* **296**, 1085 (2002).

89 Doblhammer, G. Reproductive history and mortality later in life: a comparative study of England and Wales and Austria. *Popul. Stud. (NY).* **54**, 169–176 (2000).

90 Green, a, Beral, V. & Moser, K. Mortality in women in relation to their childbearing history. *BMJ* **297**, 391–395 (1988).

91 Kvåle, G., Heuch, I. & Nilssen, S. Parity in relation to mortality and cancer incidence: a prospective study of Norwegian women. *Int. J. Epidemiol.* **23**, 691–699 (1994).

92 Beise, J. & Voland, E. Effect of producing sons on maternal longevity in premodern populations. *Science* **298**, 317; author reply 317 (2002).

93 Philippe, P. & Yelle, L. Effect of family size on mother's longevity. *Ann. Hum. Biol.* **3**, 431–439 (1976).

94 Müller, H., Chiou, J., Carey, J. R. & Wang, J. Fertility and Life Span : Late Children Enhance Female Longevity. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **57**, 202–206 (2002).

95 Smith, K. R., Mineau, G. P. & Bean, L. L. Fertility and post-reproductive longevity. *Soc. Biol.* **49**, 185–205 (2002).

96 Helle, S., Lummaa, V. & Jokela, J. Are reproductive and somatic senescence coupled in humans? Late, but not early, reproduction correlated with longevity in historical Sami women. *Proc. Biol. Sci.* **272**, 29–37 (2005).

97 Yi, Z. & Vaupel, J. Association of late childbearing with healthy longevity among the oldest-old in China. *Popul. Stud. (NY).* **58**, 37–53 (2004).

98 Le bourg, E., Thon, B., Légaré, J., Desjardins, B. & Charbonneau, H. Reproductive Life of French-canadians in the 17-18th Centuries: A Search for a Trade-Off Between early Fecundity and Longevity. *Exp. Gerontol.* **28**, 217–232 (1993).

99 Crow, J. F. How much do we know about spontaneous human mutation rates? *Environ. Mol. Mutagen.* **21**, 122–9 (1993).

100 Gavrilov, L. A. & Gavrilova, N. S. human longevity and parental age at conception. *Sex Longev. Sex. Gender, Reprod. Parent.* **7**–31 (2000).

101 Gavrilov, L. A. & Gavrilova, N. S. Predictors of Exceptional Longevity: Effects of Early-Life and Midlife Conditions, and Familial Longevity. *North Am. Actuar. J.* **19**, 174–186 (2015).

102 Cutler, D. M., Lleras-Muney, A. & Vogl, T. Socioeconomic Status and Health: *Dimensions and Mechanisms. Journal of Chemical Information and Modeling (2008).* doi:10.1017/CBO9781107415324.004

103 Jasilionis, D. & Shkolnikov, V. M. Longevity and Education: A Demographic Perspective. *Gerontology* (2015). doi:10.1159/000438901

104 Meij, J. J. et al. Quality-quantity trade-off of human offspring under adverse environmental conditions. *J. Evol. Biol.* **22**, 1014–1023 (2009).

105 Temby, O. F. & Smith, K. R. the Association Between Adult Mortality Risk and Family History of Longevity: the Moderating Effects of Socioeconomic Status. *J. Biosoc. Sci.* **46**, 703–716 (2014).

106 Mazan, R. & Gagnon, A. Familial and Environmental Influences on Longevity in Historical Quebec. *Popul. (english Ed.* **62**, 271–291 (2007).

107 Gagnon, A. *et al.* Is there a trade-off between fertility and longetivity? A comparative study of women from three large historical databases accounting for mortality selection. *Am. J. Hum. Biol.* **21**, 533–540 (2009).

108 Hanson, H. A., Smith, K. R. & Zimmer, Z. Reproductive History and Later-Life Comorbidity Trajectories: A Medicare-Linked Cohort Study From the Utah Population Database. *Demography* **52**, 2021–2049 (2015).

109 Barzilai, N. *et al.* The place of genetics in ageing research. *Nat. Rev. Genet.* **13**, 589–594 (2012).

110 Akker, E. *et al.* Uncompromised 10-year survival of oldest old carrying somatic mutations in DNMT3A and TET2. *Blood* **127**, 1509–1512 (2016).

111. Gierman, H. J. *et al.* Whole-genome sequencing of the world's oldest people. *PLoS One* **9**, 1–10 (2014).

CHAPTER 3

FAMILIES IN COMPARISON:

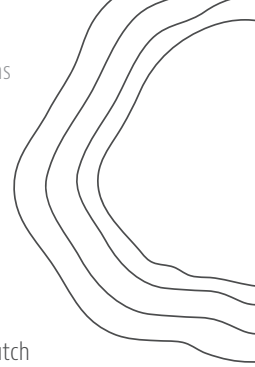
AN INDIVIDUAL-LEVEL COMPARISON OF LIFE COURSE AND FAMILY RECONSTRUCTIONS BETWEEN POPULATION AND VITAL EVENT REGISTERS

Niels van den Berg[1,2*]
Ingrid K. van Dijk[2*]
Rick J. Mourits[2*]
Angelique A.P.O. Janssens[2]
P. Eline Slagboom[1,3]
Kees Mandemakers[4]

[1] Department of Biomedical Data Sciences, section of Molecular Epidemiology,
Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands
[2] Radboud Group for Historical Demography and Family History,
Radboud University, Erasmusplein 1, 6525 HT Nijmegen, the Netherlands
[3] Max Planck Institute for Biology of Ageing,
Joseph-Stelzmann-Str. 9B, D-50931 Cologne, Germany
[4] International Institute of Social History,
Cruquiusweg 31, 1019 AT Amsterdam

*. Shared first authorship

Population Studies (March 2018)



Abstract

It remains unknown how different types of sources affect the reconstruction of life courses and families in large-scale databases, increasingly common in demographic research. Here, we compare the family and life-course reconstructions for 495 individuals, simultaneously present in two well-known Dutch datasets. LINKS-Zeeland is based on a province's full-population vital-event registration data (passive registration). The HSN is based on a national sample of birth certificates, with follow-up of individuals in population registers (active registration). We compare indicators of fertility, marriage, mortality, and occupational status and conclude that reconstructions in the HSN and LINKS reflect each other well, although LINKS provides more complete information on siblings and parents, whereas the HSN provides more complete life-course information. We conclude that life-course and family reconstructions based on linked, passive registration of individuals constitute a reliable alternative to such reconstructions based on active registration, if case selection is carefully considered.

Introduction

Demographic research is increasingly conducted on large-scale longitudinal datasets. Underlying these databases are sources such as population registers, parish registers, registrations of vital events, censuses, and genealogical databases. Names, ages, birthplaces, and other personal characteristics in these sources are used to link life-course events, e.g. marriage or migration, to individuals (life-course reconstruction), and individuals to each other into family networks (family reconstruction) . Characteristics of underlying data sources may affect the completeness and quality of life-course and family reconstructions in databases[1–3]. This is particularly the case for comparisons between databases derived from active registration where individuals are followed continuously over time, registering specific events as they happen, and databases produced from passive registration where individuals are only observed when specific events, such as a birth or marriage, are registered and the separate documents are linked together[4,5]; see *Supplementary Table 1* for an overview of active and passive registration). Well-known examples of databases based on active registration include the Roteman database for Stockholm in Sweden and the Scanian Demographic Database (SEDD). Databases based on passive registration include the Utah Population Database, the English Family Reconstitutions, and Knodels German village family reconstitutions.

Databases based on passive registration can more easily miss a vital event, such as the birth of a child. Migration movements are not registered, making it unclear whether, where, and when an individual experienced vital events in another region. As households or individuals are followed actively during their lives in active registration, observations generally contain relatively complete information on individuals and their families. Out-migration is commonly observed, so that when individuals leave the municipality or the region of residence, they can easily be traced to their new place of residence. Thus, both differences in source material as well as strategies for following individuals across data sources are likely to be crucial for the quality of reconstructed lives and families in historical databases. The extent to which they result in differently reconstructed life courses and families remains unexplored in the literature, however, due to a lack of data which enable cross-checks of the same life courses and families using different sources, with the exception of Wisselgren et al. (2014)[6].

In this paper, we show a comparison of life-course and family reconstructions for the same individuals in demographic datasets derived from two different, independent data sources: one based on the Dutch population registers, reflecting active registration, and one based on Dutch vital-event registers, reflecting passive registration. Our purpose is twofold: to investigate to what extent life courses and family reconstructions are represented similarly in databases which are based on active versus passive registration, and to determine the suitability of the types of data for different research questions, including questions on life spans and mortality, marriage behaviour, and fertility. The results are of interest to researchers working with individual level longitudinal demographic data of either sort.

[1] Combined, life-course and family reconstructions form the basis of the practice known as family reconstitution. Family reconstitution is the process of reconstructing historical data on family membership and the events occuring to these family members during the course of their lives[47].

An overview of the literature

Earlier research focusing on the quality of individual-level, large-scale longitudinal demographic databases has used a variety of approaches which consider the characteristics of the source material and the logic of the construction of the database[7–10]. Other studies have used approaches based on internal consistency of databases and comparisons to external data sources such as mortality statistics in life tables. Schellekens & Van Poppel (2016) compared population-register-data to national statistics and reported that in the Historical Sample of the Netherlands (HSN) cohort life expectancy at age 30 may be overestimated for men, but not for women[11]. Adams et al. (2002) concluded that observations on migration in vital event registrations reflected migration information in population registers well[12].

One of the main drawbacks of data based on passive registrations is that analysis is usually restricted to the residentially stable part of the population and excludes those without an age at out-migration or age at death, raising issues of representativeness[4,5,13,14]. Importantly, Ruggles (1992) observed that migration causes underestimation of population-level demographic indicators, e.g. age at marriage, age at first and last birth, and number of children[15]. After migration, migrants are right-censored and demographic events are no longer observed, causing an underestimation of the number of events as well as the mean age at the corresponding events, and this is all the more problematic when the date of migration is not recorded so that only the last recorded observation can be used. If the last observation is not a death, a potential source of bias is introduced, because individuals are still at risk of experiencing events after their last observation in the population. Statistical inferences have been developed to estimate dates of last observation when censoring occurred[4,5,16]. However, there may be true differences between the migrating and non-migrating part of the population[13,17].

While approaches based on external data sources are useful instruments to judge the quality of databases, they only provide insight into deviations at the aggregate level, such as differences in mortality rates. Whether individual life courses and families are reconstructed accurately remains an open question. Some efforts have been made in this direction, as linkage success and percentages of correct matches across sources - such as parish records and census material - have been used as an indicator of data quality[6,18,19]. In addition, several studies explored the success of linking strategies by comparisons between databases (see, for instance, Wisselgren et al. (2014) for comparisons between Swedish censuses and parish registers and Massey (2017) for historical US data)[6,20]. Ruggles et al. (2018) emphasize that most studies focus on missed links (type II errors), so that false links (type I errors) are given too little attention[21]. Both errors may introduce bias into the life-course and family reconstructions. However, missed and false links not only affect whether individuals are included in demographic databases, but also whether the correct children, spouses and parents are linked to them. By paying proper attention to false links, life-course transitions may be more accurately incorporated in databases. False matches and failed matches mostly occur in sources based on passive registration where individuals are not continuously followed over time. However, direct comparisons with sources based on active registration may reveal areas where passive registration may provide more complete data.

Data

In the Netherlands, a unique opportunity has opened up to compare the individual life-course and family reconstructions in two different types of datasets. The first is based on a sample of birth certificates (HSN) and contains active registration on households originating from the nationwide population registers. The second is based on the civil registry of Zeeland (LINKS) and contains linked civil certificates of birth, marriage and death. Individuals born in Zeeland who were included in the HSN can be identified in LINKS through an identifying combination of the municipality, year and sequence number provided on each civil certificate.

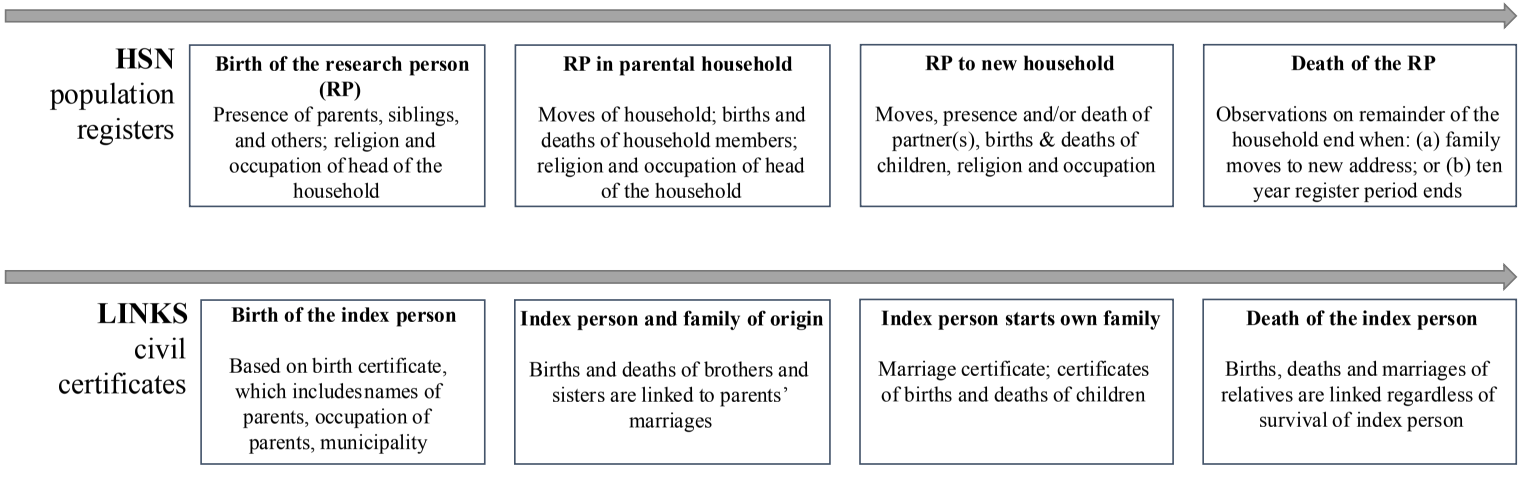
The civil registry and LINKS

The civil registry

The Dutch civil registry is one of the oldest in the world, and has covered the entire country from 1812 onward. Birth, marriage, and death certificates were kept in separate books, made in duplicate, controlled by local judiciaries, and stored at separate locations (see Vulsma, 1988[22]). The Dutch civil registry of birth, marriages and deaths is a good source for life-course and family reconstructions. All certificates contain the date of the event, the date of the registration (birth and death certificates), the place of registration, the name and age of the person reporting the event, and the names and places of residence of the witnesses. The birth certificates contain - if known - the name of the father as well as the name of the mother, and the name and sex of the child. The marriage certificates contain the age, occupation, civil status before the marriage and place of residence of both spouses, the names of the bride's and groom's parents, and - if they were alive - their age, place of residence and occupation. For death certificates, one of the two informers - one informer after 1935 - reporting the death is often a spouse or parent; they report the name, occupation, age, place of residence of the person reporting and the person reporting and the deceased person. The civil registers of births, marriages and deaths become public with delays of 100, 75, and 50 years, respectively (Burgerlijk Wetboek [Dutch civil code], article 1:17A).

The LINKing System for historical family reconstruction (LINKS)

LINKS (LINKing System for historical family reconstruction) is based on digitised certificates from the civil registries, as indexed by the WieWasWie project, to reconstruct families. The Zeeland 2017.01 release of the database contains around 700 thousand birth certificates, 200 thousand marriage certificates, and 650 thousand death certificates. Multigenerational families were built using linked marriage certificates, reconstructing life courses and families (see *Figure 1*). Of the births detailed in LINKS, 81% were linked to the marriage of their parents. In total, the dataset contains almost 2 million persons covering a maximum of seven generations. Individual life-course reconstructions resulted from linking civil birth, marriage and death certificates: 68% of all birth certificates and 66% of all marriage certificates were linked to a death certificate[19]. The scope of the database is large regarding intergenerational networks of family members[23], but the successful reconstruction of life courses and families depends on the linkage of passively registered data sources. In addition, LINKS does not contain information on addresses, co-residence of individuals, migration movements, and religion.



Description

In the HSN the RP's parental household and own household are included, but only rarely the family in law. In LINKS all relatives may be followed, independent from residency. In this study, we focus on parents, siblings, spouses and offspring of RPs.

- RP's parental household
- RP's own household
- RP's family in law

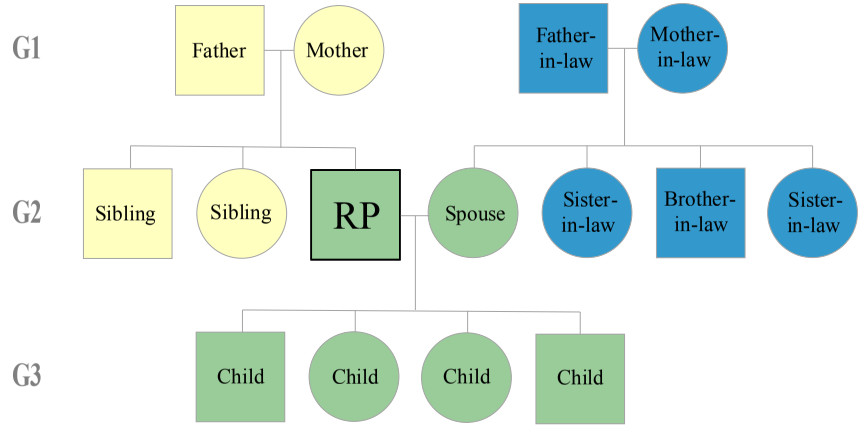


Figure 1: Data structure for HSN and LINKS.

Both LINKS and HSN can be used for life course (top two rows) and pedigree reconstruction. The top row shows the information based on the life course reconstruction in the HSN. The second row shows the information based on the family reconstruction in LINKS. The pedigree shows in general the available persons in LINKS and HSN. The green/light grey and yellow/white color represent the selection for the paper which resembles the pedigree structure that can be derived from HSN. Color descriptions first state the color version color and subsequently state the black and white version color (color/black and white).

Indexes of civil birth, marriage and death certificates were linked together, using combinations of at least two pairs of names of individuals, spouses and parents, combined with time-constraints based on age. Variations in the spelling of names, name changes, and the non-uniqueness of many names renders family reconstruction a complicated task. To prevent missed matches due to spelling variations, all first and last names were corrected for minor known variations in spelling. All name combinations for at least two persons, the individual and one or two of his or her parents and possibly a spouse or child, were matched. In the data release used here, certificates were only linked within the province of Zeeland, so that certificates of individuals who out-migrated from Zeeland to another province in the Netherlands or abroad were missed. This concerned a sizable part of the population, for example, those who migrated to Belgium and Rotterdam[24].

Population registers and the HSN

Population registers

Population registers were introduced in the Netherlands from 1850. The population registers were maintained by each municipality in large books, organized by streets or neighbourhoods. This makes it possible to follow households - and the persons in them - over time. For each household, the registers contain information on the family name, given names, sex, marital status, birth dates, death dates, birth places, address, professions, and religious denominations. For married couples, the head of the household was the male spouse. After his death, his widow would become the head of the household until her death, until remarriage, or until she moved into a household with an existing male head[25,26]. Relationships between the members of the household are included from the perspective of the head of the household, allowing the reconstruction of relationships between other household members. Movements into and out of the household as well as births and deaths were actively tracked. The books containing the population registers were replaced every ten years and updated with a coinciding nationwide census. This active registration allows the follow-up of households for longer periods of time.

In the period of research of this paper, two important changes in the population registries were implemented. In 1920 – and earlier in the large cities – the population registration was no longer ordered by street or neighbourhood, but by individual household in a card system with separate documents. From 1939 onward, the registration was no longer focused on households, but on individuals by means of personal cards. Later, in 1994, this personal card system was completely digitized. Nowadays, the system is known as the Personal Records Database (Dutch: Basisregistratie Personen, BRP) and is maintained on the national level. One year after a person's death a summary of personal and family information becomes available for scientific and genealogical research (CBG, 2019) and for specific research purposes a request can be made to Dutch government to directly access the BRP.

The Historical Sample of the Netherlands (HSN)

In the HSN, the life courses of a representative group of individuals in the Netherlands are followed. The HSN enables research on detailed life courses of individuals from the 19th century for The Netherlands[9,27]. The HSN is based on a sample of birth certificates of all individuals living in the Netherlands, stratified by cohorts of ten years for the period between 1811 and 1922 and according to regional levels of population density. The sample consists of 0.75% of the births for the period

1812 – 1872 and 0.5% of the births for the period 1873 – 1922. In total, the sample consists of about 85,500 individuals[28]. Up to now, about 40,000 of these 85,500 persons have been followed in the population registers throughout their life course. In the HSN these persons are referred to as "Research Persons" (RPs). The population-register information in the HSN is supplemented with information from the Dutch birth and marriage certificates.

In the HSN release 2010.01, entries in population registers and on personal cards of 37,137 RPs were made available[9]. For some regions, including Zeeland, the HSN starts already in 1850, when the population registers were introduced. The database includes information about the RPs' household, including co-residents, occupation, and religion. Households were, in principle, only followed as long as the RP was present in that household. Siblings and other kin were eventually lost from observation when the RP moved out of the household or died, after a follow-up to the end of the 10-year population register period. For the period after the implementation of family cards for individual households, the remaining family members were followed for up to 40 or 50 years.

Structural differences between the HSN and LINKS

Because of the sampling procedure and independent sources of information, structural differences exist between the databases in terms of the life course and family reconstructions (see *Figure 1 and Table 1*). In the HSN, siblings and parent information was only available to the extent to which family members cohabited with RPs. Therefore, questions with topics such as intergenerational and horizontal kin relations – for instance, sibling similarities in mortality – cannot be answered. Second, in LINKS individuals were observed when vital events occurred to them, their spouses, or their children. Consequently, the HSN is primarily focused on life course reconstruction and less on family reconstructions, whereas the opposite applies to LINKS, in the sense that observations on life events are used to trace family members. We will explore to what extent events of fertility marriage, migration, mortality, and occupational careers were observed and differed between the HSN and LINKS.

First, in contrast to the HSN, LINKS does not encompass unmarried cohabitation or extramarital children, which may lead to an underestimation of the number of children or siblings. Second, the lack of continuous follow-up of individuals in the civil registry makes it necessary to link certificates. The automatic record linking procedure might occasionally miss matches between vital event certificates. Moreover, certificates were only linked within the geographic area of a province, so persons were lost if they migrated to another province or country. Thus, mortality in early life was most likely measured quite accurately, but certificates of deaths and marriages occurring later in life are more likely to be unavailable. Finally, key indicators such as occupation and place of residence were only observed in concordance with vital events of individuals, their spouses or their children.

Table 1: Expected availability of demographic indicators in the HSN and LINKS

	HSN		LINKS	
	Availability on data sources	**Reason**	**Availability on data sources**	**Reason**
Parents Marriages	*Incomplete*	Not included if parents were not in household; marriage date of parents often not known	*Incomplete*	Not available for RP's who moved out of Zeeland
RP's Sibship size	*Incomplete*	Not included if siblings died before follow-up of the RP, or were born after RP moved out of the household	*Incomplete*	Not available for RP's who moved out of Zeeland
Marriages of RP	*Incomplete*	Marriages incompletely registered in population registers	*Incomplete*	Not available for RP's who moved out of Zeeland
Fertility	*Incomplete*	Offspring not included if they died before registration; no stillbirths recorded	*Incomplete*	Not available for RP's who moved out of Zeeland
Family relations	*Not always clear*	Relations within household need to be logically reconstructed for the period 1850-1862; family relations in 3rd or 4th degree may be unclear in subsequent registers	Clear	-
Occupation	Complete	Updated regularly	*Incomplete*	Not available for RPs who moved out of Zeeland; only known when a vital event was registered; measured relatively early in the life course
Later-life mortality	Complete	-	*Incomplete*	Not available for RP's who moved out of Zeeland
Extramarital fertility	Complete	Premarital fertility included; RPs who lived together but were not married	*No information*	No information on extramarital fertility
Migration	Complete	Continuous follow-up of migration in the Netherlands	*Incomplete*	Only known when a vital event was registered; persons are followed through Zeeland only
Children Child mortality	*Incomplete*	No information on offspring outside the RP's household	*Incomplete*	Not available for RP's who moved out of Zeeland

Migration patterns and occupational careers can be reconstructed from an individual's civil certificates, as well as from their children's civil certificates. Death certificates contain occupational information if the deceased person had an occupation at the time of death. Hence, more observations on occupation and place of residence were available for RP's who married or had children. Moreover, most of these vital events occur relatively early in life, so that later changes in place of residence and occupation could easily be missed. For unmarried individuals, only vital events in the family of origin and one's death certificate will be observed.

In the HSN, there were no systematic observations of events before the sampled RP was followed. Observations on RPs does not always start at birth, leading to gaps in life course and family reconstructions. The implication is that siblings who reside elsewhere or died young may not be included in the register in which the RP first appears. As a result, the count of all known siblings reflects the count of surviving siblings – the net fertility – rather than the count of all siblings ever born – the total fertility. At the same time, RPs children were identified very accurately in the HSN because RPs were, in principle followed for their entire life course. This is illustrated by[29], who showed for Tilburg (1849-99) that 99.8% of the children found in the birth registers were identified in the population registers. At the same time, stillbirths and children who died very soon after birth were usually not included in the birth or population registers, but only in the death registers (hereinafter, "lifeless reported infants"). These characteristics limit opportunities for research on events early in the life course – such as exposure to sibling mortality or the length of birth intervals – and research on intergenerational relations in longevity, mortality, and fertility.

Data construction and approach

For the comparison between the HSN and LINKS, we used persons identified in both databases who were born between 1863 and 1872 . Drawing on data from LINKS 2017.01[30] and the HSN 2010.01 population register release[31–33], the 495 Zeeland-born individuals included in the HSN were traced in LINKS via unique identifiers of the birth certificates. We analysed differences in life-course and family reconstructions of RPs in the estimation of key demographic and socioeconomic indicators. We test whether the characteristics of the databases might have lead to an underestimation in the number of links. Demographic linking strategies tend to go for precision (few false matches) at the expense of recall (few missed matches)[6,34]. Moreover, biases in the registration procedure leads to omissions in the data. Therefore, different observations between the HSN and LINKS are most likely indicative of false negatives, i.e. missed observations.

An overview of all available information in both datasets and expected completeness is provided in *Table 1*. For our analyses, we used the following indicators: sex, start and end dates of observation (HSN) or first and last observation (LINKS), birth year, and death year. We counted the number of siblings and children known, and the birth order of the RP in their family of origin. With regard to the number of siblings and children, stillbirths and infants lifeless upon civil registration were excluded, as they were unavailable in the HSN. In addition, we measured ages at first and last childbirth. Furthermore, we noted whether RPs married or not and had children or not; calculated their age at first marriage and at death; traced whether they migrated within Zeeland, outside Zeeland (HSN), or never; and tested their socioeconomic position on consistency between both datasets using HISCLASS, a social class scheme to classify historical professions[35,36].

Results

Table 2 presents the number of RPs for whom parents, siblings and children could be identified. Because entire households of individuals are actively registered in the source material underlying the HSN and observations are available for the entire country, information on parents, spouses, and children is more often available in the HSN than in LINKS, which is based on linked civil certificates from Zeeland. In the HSN, 96% of the RPs had available parent information, for a total of 932 parents. In LINKS, 82% of the RPs had available parent information (814 parents). In the HSN 1,060 children were identified (for 40% of the RPs), whereas in LINKS 810 children were identified (for 31% of the RPs). However, fewer siblings are known in the HSN than in LINKS (1,447 and 2,804 siblings, for 72 and 83% of RPs, respectively), as these were only observed if they lived together with the RP in a household. Fewer spouses were known in the HSN than in LINKS, because marriages were registered in the civil records in the first place, and may not always have been registered correctly in the population registers. A total of 233 spouses were found in the HSN (28% of the RPs), while 188 spouses were identified in LINKS (36% of the RPs). Hence, active registration increases the number of RPs with known family relations, but might be related to missed events that occurred outside of an RPs household or in other registers. This difference between events within and outside the household does not exist for passively registered sources. For both datasets, the number of individuals without spouses and children appears to be high. However, many individuals in Zeeland did not reach reproductive ages, as infant and child mortality in Zeeland was very high, reaching up to 50% in some municipalities and years[37,38].

Table 2: Available family ties in the HSN and LINKS for the selected 495 RPs from the 1862-1871 Zeeland cohort

Relatives	Sample size	RPs with known relatives (%)
HSN		
Parents	*932*	475 (96)
RPs	495	
Siblings	1447	336 (72)
Spouses	233	138 (28)*
Children	1060	196 (40)
LINKS		
Parents	814	407 (82)
RPs	495	
Siblings	2804	413 (83)
Spouses	188	177 (36)
Children	810	151 (31)

The 233 spouses are identified using the population registers. These numbers are used in the paper itself. When adding marriage certificates to the population registers, we identified 237 spouses and 277 unmarried RPs. Combined, the population registers and marriage certificates identify 324 spouses and 270 unmarried RPs. The RPs with known relatives refer to the number of RPs with for example known parents (N=475). Spouses are based on the number of marriages. Hence, one RP could have had multiple spouses.

Comparisons between the databases were conducted in two ways. First, we compared all individuals for whom relevant observations can be expected in both databases separately, with the purpose of exploring all life course and family reconstructions *(Table 3A)*. Because the mean scores in this table were based on different RPs, these means must be interpreted for each dataset separately. Second, we analysed only the subsets for which we could reconstruct life courses in an identical way, hence, we selected individuals for whom a relevant observation may be expected in both databases *(Table 3B)*. Both tables show key demographic information for all RPs for whom it is possible to know whether they experienced the demographic event. Cases without information on the relevant selection criteria were not included. Differences between the HSN and LINKS in demographic indicators in *Table 3B* indicate differences in the reconstructions of life courses and families between the HSN and LINKS, whereas differences in these indicators in *Table 3A* may also be caused by differences between the subsets of individuals for whom information was available.

Table 3: Number of available cases and mean score on demographic indicators after selection of best cases in the HSN and LINKS separately (3A) and simultaneously (3B)

	Date selection	Indicators	Table 3A				Table 3B		
			HSN N	LINKS n	HSN N (mean)	LINKS N (mean)	HSN-LINKS N	HSN N (mean)	LINKS N (mean)
Siblings	Lived until age 30	Number of RPs with identified siblings (mean number of siblings)	221	-	221 (3.9)	-	186	186 (3.8)	-
	Lived until age 30	Birth order	221	-	221 (1.8)	-	186	-	186 (6.6)
	Parents known	Number of RPs with identified siblings (mean number of siblings)	-	407	-	407 (6.7)	186	186 (1.8)	-
	Parents known	Birth order	-	407	-	407 (4.2)	186	-	186 (4.0)
Marriages	Lived until age 30	Ever-married*	221	146	122 (55.2)a	124 (84.9)	138	62 (44.9)a	118 (85.5)
	Lived until age 30	Never-married*	221	146	99 (44.8)b	22 (15.1)	138	776 (55.1)b	20 (14.5)
	Ever-married	Age at first marriage	137	178	137 (28.4)c	178 (26.3)	97	97 (27.7)c	97 (26.8)c*
	Ever-married	No children identified*	122	178	13 (9.5)d	26 (14.6)	97	8 (12.9)d	15 (15.5)d*
Children	Children identified	Age at first childbirth	196	152	196 (27.0)	152 (26.5)	146	146 (26.8)	146 (26.6)
	Children identified	Age at last childbirth	196	152	196 (37.4)	152 (36.6)	146	146 (37.4)	146 (36.8)
	Children identified	Number of RPs with identified children (mean number of children)	196	152	196 (5.4)	152 (5.2)	146	146 (5.4)	146 (5.4)
Migration behavior	Lived until age 18	Never moved*	236	157	62 (26.3)	90 (57.3)	149	56 (37.6)	86 (57.7)
	Lived until age 18	Moved within Zeeland*	236	157	78 (33.1)	67 (42.7)	149	71 (47.7)	63 (42.3)
	Lived until age 18	Moved out of Zeeland*	236	157	95 (40.3)	-	149	21 (14.1)	-
Age at death	All	All ages at death	409	313	409 (40.8)	313 (34.7)	306	306 (33.6)	306 (33.9)
	Lived until age 18	Ages at death after 18	236	157	236 (69.4)	157 (67.4)	149	149 (67.1)	149 (67.1)
	Lived until age 50	Ages at death after 50	204	134	204 (75.1)	134 (73.6)	126	126 (73.6)	126 (73.6)

ndicators in Table 3A do not concern the same RPs, as best cases are selected separately in the HSN and LINKS. Identical RPs are selected for both the HSN and LINKS in *Table 3B*. All indicators with a * are categorical. Hence, N (mean) represents the number and the proportion of RPs belonging to that sub-group. Marriages are based on the population registers in the HSN. Using marriage certificates provides the following numbers in *Table 3A*: a. N= 194 mean=87.8, b. N=27 mean=12.2, c. N=137 mean=28.4, d. N=30 mean=13.3. Reading example: row 7 shows that 122 persons lived beyond the age of 30 years and married at some point in their lives in the HSN and that this was the case for 124 in LINKS. These respectively reflect 55.2% of the 221 persons who ever married in the HSN and 84.9% of the 146 persons who ever married in LINKS. Using the HSN marriage registers results in the following numbers is *Table 3B*: a. N=120 mean=87.0, b. N=18 mean=13.0, c. N=177 mean=27.7, c* N=177 mean=26.3, d. N=23 mean=13.0, d* N=26 mean=14.7. Reading example: row 7 shows that 138 persons lived beyond the age of 30 years in both the HSN and LINKS. Out of those 138 persons, 62 persons married at some point in their lives in the HSN and 118 in LINKS. These respectively reflect 44.9% and 85.5% of the in total 138 persons identified in both the HSN and LINKS. For siblings and children the N refers to the number of RPs with identified siblings. The mean refers to the mean number of siblings or children.

Comparisons of demographic indicators in the HSN and LINKS

Table 3A shows that the mean number of siblings and birth order were lower in the HSN (3.9 and 1.8) than in LINKS (6.7 and 4.2). These results were similar for the 186 identical cases. These differences are mainly a consequence of the research design of the HSN, in which siblings are only observed if they are part of the RPs' household. Therefore, information on siblings who died young or who did not live in the household is often missing, leading to an underestimation of sibship size in the parental household of the RP.

Within LINKS, marriages were available for 84.9% of the RPs of 30 years and older, whereas in the population register release of the HSN, marriages were available for 55.2% of the selected RPs. *Table 3B* shows that for the 138 common RPs, marital information was available for 85.5% in LINKS and for 44.9% in the HSN, which indicates that marriages were often not included in the population registers. Without selections, the mean age at marriage in the civil certificates and the population registers were 26.3 and 28.4, respectively. The higher mean age at marriage in the HSN is partly caused by right-censored observations in LINKS. Out-migration is known to cause underestimation of the number of events as well as the age at which demographic events occur[15]. Nevertheless, the number of known marriages is higher in LINKS than the HSN after we selected only individuals who married in Zeeland. The mean age at first marriage in LINKS was, at 26.8 years, higher than in the LINKS only selection as shown in *Table 3A*. In the HSN the age at first marriage in *Table 3B* was lower than in *Table 3A*, at 27.7 years. The higher age at first marriage in the HSN *(Table 3B)*, may be related to left truncation in the HSN, as not all RPs were followed for their entire life course, so that second marriages were counted as a first marriage, resulting in an overestimation of the mean age at first marriage. After combining the HSN population registers with the HSN marriage certificates, we observed that marriages were available for 87.0% of the RPs with a mean age at first marriage of 27.7 years (see the notes to *Table 3*). There is no evidence that passive registration lead to biased estimates. Differences between both datasets originate from registration procedures and censoring due to migration.

Table 3A shows that the mean number of identified children in the families of the RPs was similar between both datasets: 5.4 children for RPs in the HSN and 5.2 in LINKS. However, the number of RPs with identified children was lower in LINKS (N=152) than in the HSN (N=196). Furthermore, the mean ages at first and last birth in LINKS (26.5 and 36.6 years) were lower than in the HSN (27.0 and 37.4 years). The percentage of married couples without identified children is 14.6% in LINKS and 9.5% in the HSN. These differences are probably caused by right-censored observations in LINKS due to out-migration. *Table 3B* shows that for the 146 RPs who are included in both datasets, the mean age at first childbirth was 26.6 in LINKS and 26.8 in the HSN. This selection of common cases also showed the same mean number of children (5.4), although the mean age at last birth is lower in LINKS than in the HSN and the percentage of married couples without identified children is 15.5% in LINKS and 12.9% in the HSN. Apparently, the automatic linking procedure failed up to pick up specific certificates. Later-born children and entire families might be missing, as differences in mean age at last birth and mean number of children remain after selecting identical RPs.

The HSN and LINKS include different information on migration behaviour, as out-migration from Zeeland was not observed in LINKS. *Table 3A* shows that, according to the HSN, 95 (40.3%) of the RPs who were alive at age 18 migrated out of the province at some point in their lives. In LINKS, 157 death certificates are available for the RPs who lived at least until age

18, suggesting that these RPs either never left Zeeland or returned to Zeeland at a later age. The HSN indicated that 140 RPs (59.6%) never lived outside Zeeland, of which 62 (26.3%) never moved at all, and 78 (33.1%) only moved within Zeeland. Vital events outside the place of birth of the RP, indicating migration between municipalities within Zeeland, were identified for 67 RPs. This pattern was similar when we compared identical individuals *(Table 3B)*. In LINKS, we observed that 63 RPs (42.3%) who were observed after age 18, died in another municipality than their municipality of birth, whereas both vital events occurred in the municipality of birth for 86 RPs (57.7%). According to the HSN, 56 RPs (37.6%) remained in their municipality of birth, 71 (47.7%) moved within Zeeland and 21 (14.1%) lived outside Zeeland at some point in their lives. As about one out of seven adults who were born and died in Zeeland lived outside Zeeland at some time, assumptions about cross-provincial migration behaviour or the lack thereof should not be based on the presence of a death certificate in LINKS alone. The passive registration of individuals in the source material of LINKS means that migration movements can easily be missed.

The last rows in *Table 3A* present the number and mean age of death for all RPs for whom an age at death was known and the mean ages at death for individuals reaching 18 and 50 years. Because persons out-migrating from Zeeland are known in the HSN but not in LINKS, we expected that in the HSN more ages at death would be known and that the mean age at death would be higher. Indeed, fewer ages at death were known for RPs in LINKS than in the HSN, resulting in a lower mean age at death in LINKS (34.7 years) than in the HSN (40.8 years). The difference between the databases in the mean ages at death was smaller at higher ages. For those surviving until age 18, the mean age at death was 67.4 in LINKS and 69.4 in the HSN; after age 50, the mean ages at death are 73.6 and 75.1. An important reason for the declining difference with age is the declining likelihood with age that individuals will out-migrate (Kok, 1997). Differences between the HSN and LINKS are mitigated after identical cases were selected, which supports our assumption that selective availability of information for out-migrated individuals plays an important role. Hence, passive registration itself does not seem to cause biases in mortality estimates.

Comparing life-course and family reconstructions of RPs between the HSN and LINKS

Here, we take a closer look at deviations in individual life-course and family reconstructions. *Figure 2* shows whether estimations of outcomes in the HSN and LINKS deviate upward, downward, or are identical. Because information may be more complete for some subsets of individuals, four groups are included: individuals with a) no selections, b) known marriage certificates for parents of siblings and known death certificates for RPs, c) known migration inside Zeeland, and d) known migration outside Zeeland. Different estimations are seen as indicative of missed observations, as the matching procedure in both datasets has a low chance of producing false positives.

The HSN misses siblings that were not living in the RP's household. Without any selections on the data (Panel 2A), the number of siblings was higher for 69% of the RPs in LINKS, whereas 16% of the RPs in the HSN contained more siblings. However, LINKS also contained missed observations. The differences between the number of siblings in the HSN and LINKS were even more pronounced if a marriage certificate of the parents was known in LINKS (Panel 2B). Out-migration partly explains why family reconstructions in LINKS are better when a marriage certificate is available.

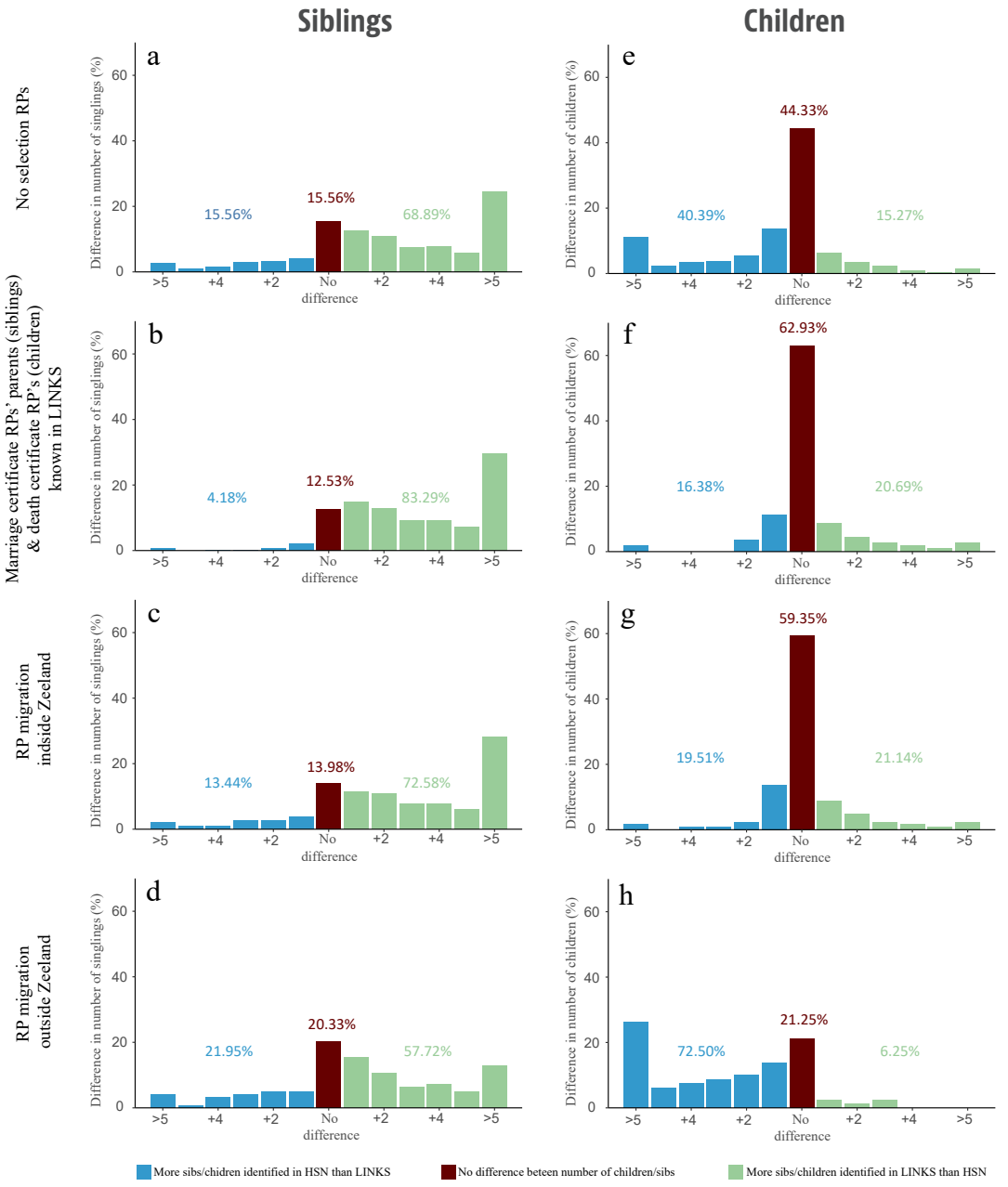


Figure 2: Matching for number of siblings and children between the HSN and LINKS.

The figure shows the matching for siblings and children of research persons between HSN and LINKS. Colors: red/black=exact match, blue/dark grey=more siblings or children identified in HSN than in LINKS, green/light grey=more siblings or children identified in LINKS than in HSN. Color descriptions first state the color version color and subsequently state the black and white version color (color/black and white). The x-axis show how much more siblings or children have been identified in either LINKS (green) or HSN (blue). The y-axis shows the percentage of matches corresponding to the x-axis. The marginal sums are illustrated with colors corresponding to the bars and sum up to 1 (100%). The legends on the left illustrate different data selections based on HSN or LINKS if explicitly stated. Numbers (N) per panel: A=495, B=407, C= 372, D=123, E=203, F=116, G=123, H=80.

In LINKS, fewer siblings were found in 4% of all cases compared to the HSN, the same number of siblings was found in 13% of all cases, and more siblings were found in 83% of all cases. Migration within Zeeland did not affect the results (Panel 2C), whereas for RPs who migrated out of Zeeland, the number of siblings in LINKS was lower than in the HSN in 22% of all cases, identical in 20% of the cases, and higher in the remaining 58% (Panel 2D). The availability of a parental marriage certificate is an independent observation that hints at successful matches between parents and their children. In general, reconstructions of sibships can be considered complete if such an independent observation is available.

The number of children of the RPs was more similar between both datasets than the number of siblings. *Figures 2E-H* show the difference in number of children between the HSN and LINKS, which was calculated for RPs who had children identified in either or both datasets. The active registration in the source of the HSN initially returned better results than the passive registration in LINKS. Without selections on the data, the HSN provided the most accurate results. For 40% of all RPs more children were found in the HSN than in LINKS; for 44% of all RPs the same number of children was found in both datasets; and in the remaining 15%, more children were found in LINKS than in the HSN. The difference between family size in the HSN and LINKS may have been caused by interprovincial migration, as births outside Zeeland are not included in LINKS. To indicate the quality of the linking process, RPs were selected who were known to have married, had children and died in Zeeland. The availability of a Zeeland death certificate for the RP and at least one Zeeland certificate for his children indicates that the RP spent a large part of his or her life in the province, thus reducing the chances that the RP migrated to a minimum. For these RPs, the same number of children was found in the HSN and LINKS in 63% and 59% of all cases respectively (Panels 2F and 2G). Where the number differed between the HSN and LINKS, there was no clear distinction in performance between the databases: the HSN performed better in some cases, whereas LINKS performed better in the others. If RPs moved out of Zeeland, a larger number of children was found in the HSN in 73% of all RPs, the same result in both sets in 21% of all cases, and a smaller number in the other 6%. Hence, the differences in family size between the HSN and LINKS are caused by migration rather than linking quality. Thus, the availability of a death certificate in LINKS indicates that observations on childbirth are likely available as well. This shows that passive registration can approach the quality of active registration when a later observation is available, e.g. a death certificate.

Figure 3 shows comparisons in the available mortality information between both datasets in four panels. The HSN returns more observations than LINKS, whereas the quality of matches is highly similar between both datasets. Panel 3A shows that ages at death were known for 409 RPs (83%) in the HSN, whereas 313 RPs had an available age at death in LINKS (63%). The age at death overlapped in 304 cases (99%) for whom a death certificate was available in both databases. Panel 3B presents childhood mortality for the RPs, their siblings, and their children. In the HSN, childhood mortality for RPs was estimated to be 6-7% higher than in LINKS, reflecting the good coverage of RP information in the HSN. However, observations on sibling and offspring mortality are of lower quality. Childhood mortality was estimated to be twice as high for siblings and almost three times as high for children in LINKS compared to the HSN. Panel 3C shows that adult mortality estimates were influenced by migration outside Zeeland. Among individuals who stayed in their municipality of birth or who moved within Zeeland, the mean and median ages at death were similar between both datasets.

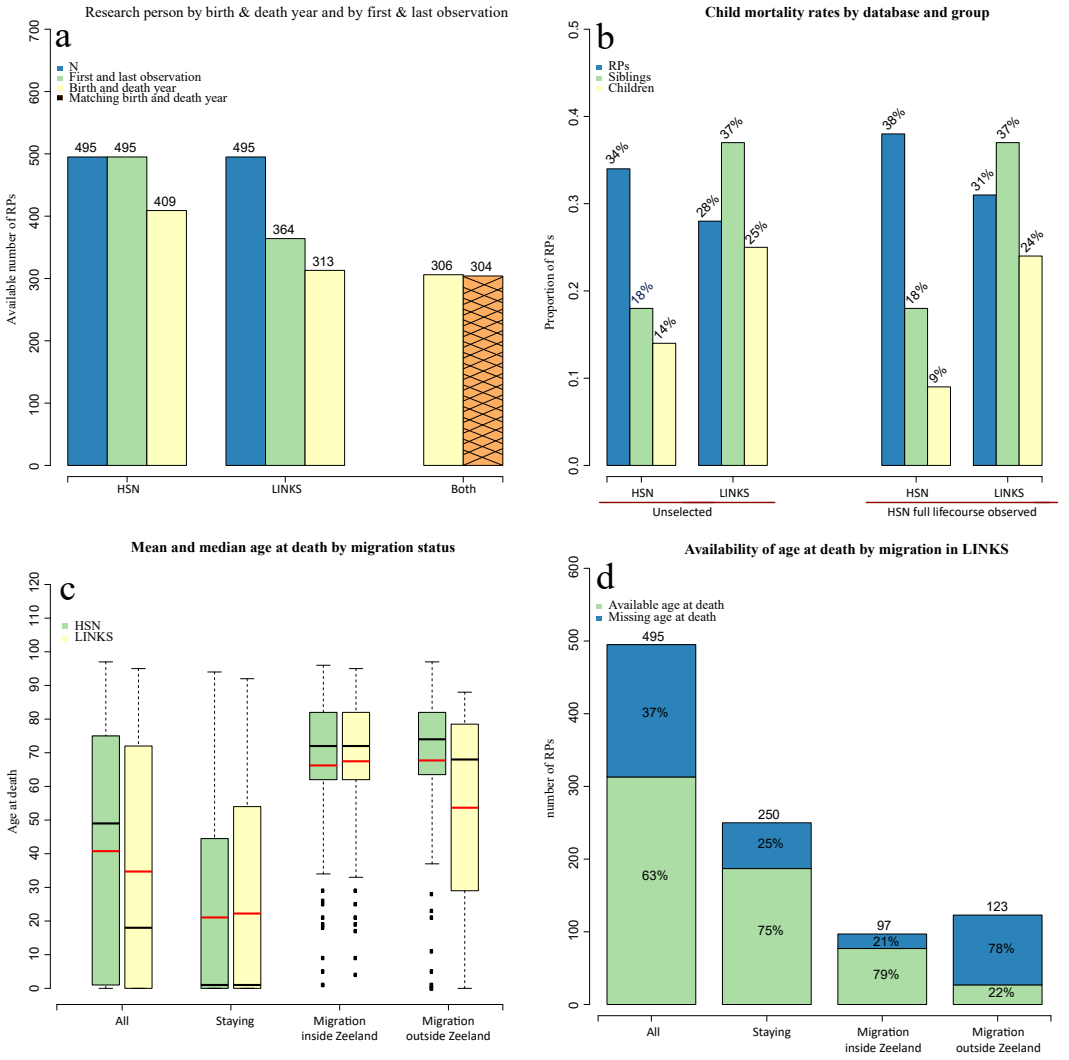


Figure 3: Comparison between HSN and LINKS for mortality data.

Panel A shows the research persons by birth and death year and by first and last observations (HSN only) in absolute numbers. Panel B shows the percentage of childhood mortality (mortality <5 years) by database (HSN and LINKS) and group (unselected and full life course). Full life course indicates that HSN RP's are observed from birth. Panel C shows the mean (red/light grey) and median (black) age at death by migration (staying, migration inside Zeeland, and migration outside Zeeland) by database (HSN and LINKS). Panel D shows the availability of age at death by migration only for the LINKS database in absolute numbers and percentages. Migration for RP's is determined based on the HSN since migration in LINKS is not available by definition of the source material.

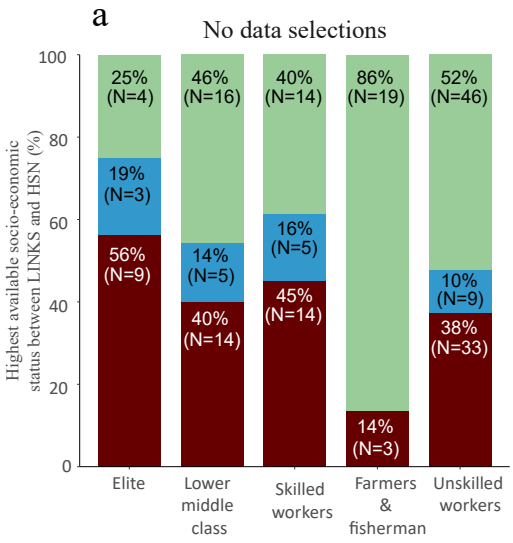
However, death certificates were not linked for 25% of the individuals who were marked as stayers in the HSN and 21% of the individuals who were identified as internal migrants. Some of these individuals might have survived the observation

period, as death certificates are not available after 1962. For other cases the death certificates were not linked due to spelling and age variations on the death certificates. There also is evidence of a "salmon bias" effect. For RPs who left Zeeland according to the HSN, the mean and median age at death is lower in LINKS than in the HSN. The date of death is known for only 22% of all RPs for whom we know, based on the HSN, that they lived outside Zeeland at some point in their life course (Panel 3D). These return migrants have a significantly lower age at death than interprovincial migrants who died outside Zeeland. Thus, passive registration returns fewer observations, but we find no proof for systematic biases related to the linking process. In addition, LINKS contains a selective group of stayers and return migrants, of which especially the latter may affect population estimates.

Figure 4 shows the differences in socioeconomic position between the HSN and LINKS based on the HISCLASS[35]. We present social class on the abbreviated HISCLASS scale with 5 categories: 1. elite, 2. lower middle class, 3. skilled workers, 4. farmers and fishermen, and 5. unskilled workers[10]. *Figure 4* further shows whether RPs in the HSN with an available HISCLASS-5 score had none, the same or a different score in the LINKS dataset. In general, the active registration in the HSN returned more cases than passive registration in LINKS. Panel 4A shows that in total 73 RPs - 33 women and 40 men - had known socioeconomic information in the HSN but not in LINKS. Conversely, 32 RPs - 25 women and 7 men - were recorded in LINKS, but not in the HSN. The share of missing values varied between 38% and 45% for unskilled workers, skilled workers, and the lower middle class, was slightly higher for the elite with 56%, and only occurred for 14% of the farmers. *Figure 4B* shows that HISCLASS scores were identical for 80% of the RPs for whom occupational information was known in both datasets. All farmers in the HSN were also identified as farmers in LINKS. However, differences in social position were found for 22% of the other RPs. Most discrepancies with the HSN occurred for the elite (43%), more than for the lower middle class (24%) and skilled workers (29%). Fewer differences with the HSN were found for the unskilled workers (16%). Underestimation of socioeconomic status generally occurred when information on occupational status was not known after marriage (Delger & Kok, 1998). These problems with censoring are probably caused by migration, rather than passive registration in the source. Geographic mobility is known to be higher for individuals with a better socioeconomic position[39], so that observations of those who reach a higher social position in society are more likely to be censored. Therefore, local datasets underestimate the social position of migrants as less occupational information is available at higher ages, and are biased towards stayers who, on average, reach a lower social standing.

Figures 4C-E show comparisons of the occupational score in the HSN with the LINKS score on the RP's death certificate, his or her marriage certificate, and the marriage and death certificates of the RP´s children. The choice for a certain certificate determined the sample size. Occupations were only recorded on death certificates if the deceased held an occupation at the time of death. As a result, occupational information on death certificates is limited and only available for 29 cases, but the HISCLASS scores were very similar between the datasets. Marriage certificates were available for 112 RPs, of whom 52 were identified as unskilled workers in the HSN. In 98% of the cases, these were also identified as unskilled labourers on their marriage certificates. However, marriage certificates are less concordant with the HSN for socially mobile individuals. Between 36% and 42% of the farmers, skilled labourers, and lower middle class had a different occupational position on their marriage certificate than in the HSN. This difference was larger for the elite (57%). The 59 RPs with marriage and death certificates of children in the LINKS dataset (Panel 4E) have a better balance between sample size

and matching quality in socioeconomic position than the comparison made in Panel 4D. Similarly, farmers show no differences at all between the HSN and LINKS datasets. For the other groups, socioeconomic positions differ from 23% to 33% of unskilled workers, skilled workers, and the lower middle class. For the elite two out of the three observations are different. Individuals with more children have more observations of their socioeconomic status, and for them HSN and LINKS reflect each other better. More generally, because in passive registration databases the number of observations depends on the number of linked events, passive registration databases reflect the active registration database better when more events were linked.



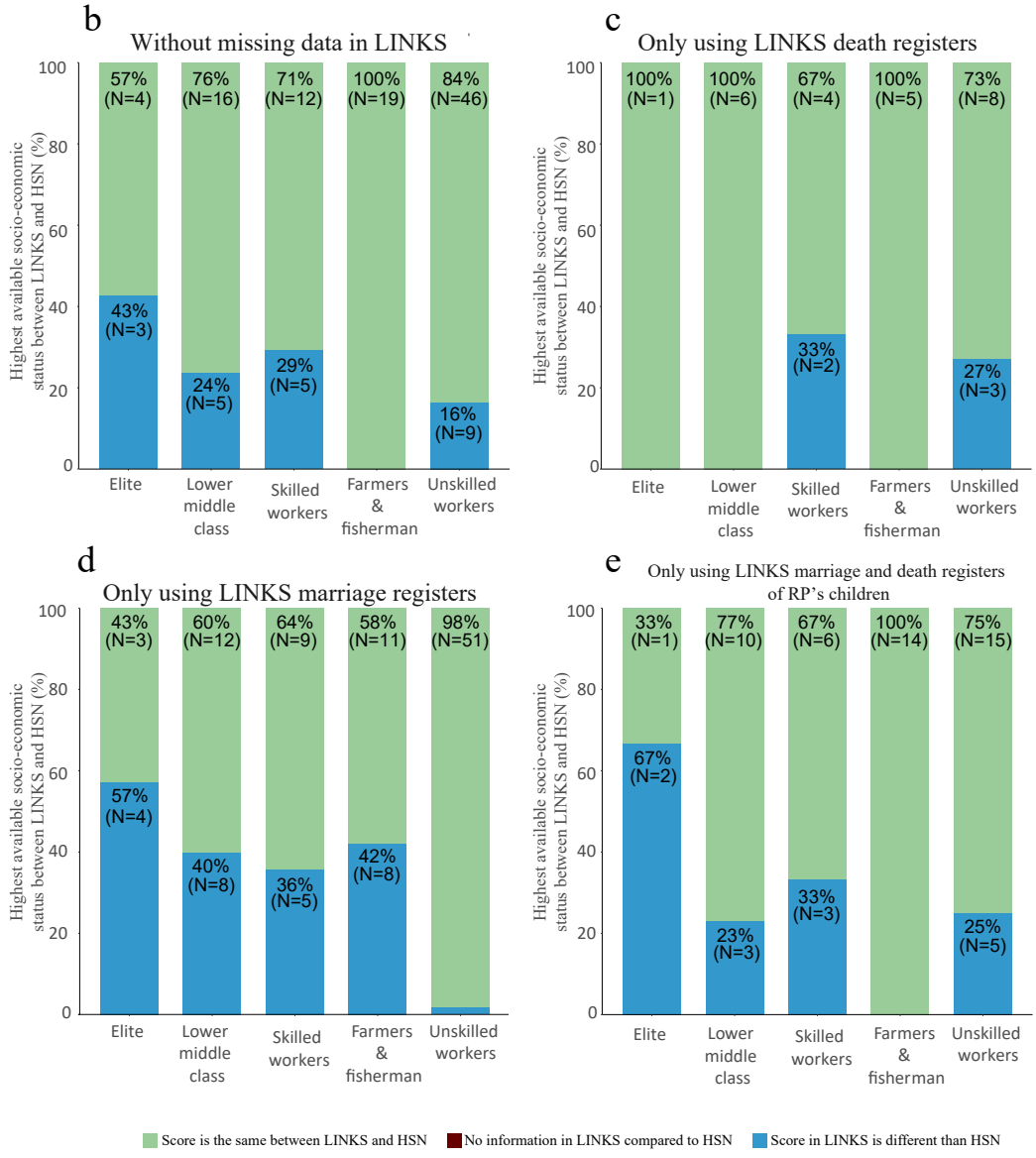


Figure 4: HISCLASS score for research persons in HSN and LINKS.

32 RP's are available in LINKS and not in HSN (7 males, 25 females. 119 RP's are available in both the HSN and LINKS. (91 males, 28 females). Panel A shows the proportion of highest available socio-economic status between LINKS and HSN without any data selection LINKS. Panel B shows the proportion of highest available socio-economic status between LINKS and HSN without the 73 missings in LINKS. Panel C shows the proportion of highest available socio-economic status between LINKS and HSN with only information of death certificates used in LINKS. Panel D shows the proportion of highest available socio-economic status between LINKS and HSN with only information of marriage certificates used in LINKS. Panel D shows the proportion of highest available socio-economic status between LINKS and HSN with only information of marriage and death certificates of RP's children used in LINKS.

Conclusion and discussion

In this paper, we compared life course and family reconstructions for 495 individuals who are available in two different types of data sources: the HSN based on active registration in the population registers, and LINKS, based on passive registration from the civil certificates. We found that differences between the HSN and LINKS were caused by censoring due to migration, rather than the nature of the administrative process which seems to induce more random missings. The established practice of selecting specific cases[1] made most differences in demographic estimates between the databases based on active and passive registration disappear, but only for demographic estimates at the individual level.

In general, the identification of children appears to be more complete when databases are based on active registration. The total number of families with children as well as the number of identified children per family is higher in the HSN than in LINKS. However, after adjustments to exclude inter-provincial migration, the number of children identified in the databases was usually identical between the databases. This finding illustrates that the identification of children using passive registration is of similar quality to active registration for non-migrants. In line with our expectations, the number of RPs with known siblings or the size of the RPs' sibling set is smaller in the HSN than in LINKS. Sibling reconstructions in LINKS are complete when a marriage certificate of the parents is available. Due to the research design of population registers and the HSN, not all siblings were found in the population registers in which the RPs appeared. Apart from missed migrants, LINKS seems to contain well-reconstructed families, so that not only the number of children, but also the siblings are identified in the dataset. For both databases based on sources with passive and active registration, it seems best to only include observations on siblings or offspring when separate observations indicate that observations are not censored.

Population estimates on demographic behaviour are strongly affected by whether observations are missed due to migration. Ruggles (1992) used simulation methods to show that – even in the absence of healthy migrant effects – cessation of observation on individuals due to out-migration causes underestimation of the ages at which demographic events occur[15]. Later work pointed out that migration at young ages, or because of a marriage at the same age as in the population of origin, do not bias estimations of age at marriage[40]. As more individuals were lost from observation due to migration in LINKS than in the HSN due to the provincial scope of LINKS and the national scope of the HSN, we expected that the mean age at which life-course transitions occur would be lower in LINKS than in the HSN. Indeed, we found that not only age at death, but also age at first marriage, first childbirth, and last childbirth were higher in the HSN than in LINKS. More generally, this implies that mean estimates, such as average age at death of a study population, show a stronger downward bias when the loss of observation due to migration increases. However, estimates of age at marriage are much less affected by migration. This contrasts earlier work which showed that in some populations migration patterns may not distort estimations of age at marriage altogether, as individuals migrating out may migrate for marriage specifically, or very early in life, before they are at risk of marrying[14,40]. In addition, in LINKS, more men and women had no identified children than in the HSN. Censoring of observations due to migration - and not passive registration - thus has a significant effect on population estimates.

Ages at death were identical in the HSN and LINKS for 304 out of 306 cases, indicating the validity and comparability of the life-course reconstructions in both databases. In line with earlier observations from Hacker (1997), migration seemed to have a strong effect on mortality estimates[13]. We expected that we would find a lower mean age at death in LINKS than in the HSN, as it has been shown that migrants are often healthier than the native population, a phenomenon known as the healthy migrant effect. Indeed, we found a lower mean lifespan for the RPs in LINKS than in the HSN, which was attributable to the almost 100 extra observations of lifespans that were available in the HSN in comparison to LINKS. These observations mainly concerned out-migrated adults, increasing the length of the mean life span in the HSN. Moreover, we found that individuals who were observed outside the province of Zeeland during their life course, but who returned to Zeeland, died at earlier ages than individuals who never migrated or who migrated within the province of Zeeland. This suggests that return migration occurred for health considerations, contributing to the problem of underestimation of ages at death in LINKS. In sum, this means that reliable estimates on mortality rates in the general population cannot be derived from regions with pronounced out-migration, unless subgroups are studied (e.g. infants or those 50+) or moments of censoring after the last observation are inferred[4,5,16,41]. However, one can wonder how useful the latter method is, seen that it only corrects mortality estimates for when individuals migrate, i.e. 15-50, and not for when migrants have left, ages 50 and up.

In the literature, a number of earlier studies have reported findings in line with the "salmon bias", which states that the relative health advantage of migrants in comparison to the native population may at least partially be caused by return-migration movements of unhealthy migrants. Earlier work has found that healthier individuals tend to migrate more and further in contemporary as well as historical populations[42,43]. Work from England has shown that migrants affected by pulmonary tuberculosis tended to return to their regions of origin, leading to high mortality rates in sending regions and relatively low mortality rates in receiving regions[44]. At the same time, a historical study on Rotterdam did not find evidence for either a healthy migrant effect or a salmon bias[45] Evidence for the current data is in line with both healthy migrant effects and a salmon bias. Possibly, in Rotterdam, healthy migrant effects were counterbalanced by a heavy urban penalty affecting migrant's health, which is absent in the small towns of Zeeland. Alternatively, for salmon bias to occur a disease has to be in chronic rather than a short sickbed before death[46]. The occurrence of salmon bias may therefore be related to spatial differences in disease patterns.

This paper has illustrated that life-course and family reconstructions based on linked, passive registration on individuals constitute a reliable alternative to such reconstructions based on active registration. Through the further integration of existing sources databases for innovative new research may be generated. Information from different datasets can be combined to gain new and more complete insights into demographic behaviour. The extensive family networks found in LINKS can contribute to more detailed kinship information in the HSN; for instance, with regard to lifeless reported infant siblings and children or more detailed observations on socioeconomic status. In current versions of the HSN, marriage certificates - which are also included in LINKS - are already used to enrich information on relationship formation found in population registers. Second, differences between the two databases may itself be of interest for family historians and historical demographers. Deviating information on siblings and children within households in the HSN and regardless of household in LINKS may provide researchers with clues on non-co-resident kin, a phenomenon on which neither database

alone provides information. Similarly, supplemental observations on socioeconomic status in the HSN may enrich our understanding of the development of the status of individuals over time. As the current analyses have shown, it should be taken into account that information for certain individuals may more readily be matched between databases, such as index persons from LINKS who remained in their province of origin.

References

1 Bousquet, G.-H. & Henry, L. Manuel de démographie historique. *Rev. économique* **21**, 513 (1970).

2 RUGGLES, S. The limitations of English family reconstitution: English population history from family reconstitution 1580–1837. *Contin. Chang.* **14**, 105–130 (1999).

3 Song, X. & Campbell, C. D. Genealogical Microdata and Their Significance for Social Science. *Annu. Rev. Sociol.* **43**, 75–99 (2017).

4 Gill, R. D. Nonparametric estimation under censoring and passive registration. *Stat. Neerl.* (1997). doi:10.1111/1467-9574.00036

5 Alter, G., Devos, I. & Kvetko, A. Completing Life Histories with Imputed Exit Dates: A Method for Historical Data from Passive Registration Systems. *Popul. (english Ed.* (2009). doi:10.3917/pope.902.0293

6 Wisselgren, M. J., Edvinsson, S., Berggren, M. & Larsson, M. Testing methods of record linkage on Swedish censuses. *Hist. Methods* (2014). doi:10.1080/01615440.2014.913967

7 Delger, H. & Kok, J. Bridegrooms and biases: A critical look at the study of intergenerational mobility on the basis of marriage certificates. *Hist. Methods* (1998). doi:10.1080/01615449809601194

8 Gavrilov, L. A. & Gavrilova, N. S. Biodemographic Study of Familial Determinants of Human Longevity. *Popul. An English Sel.* **13**, 197–221 (2001).

9 Mandemakers, K. Building life courses datasets from population registers by the historical sample of the Netherlands. *in History and Computing* **14**, 87–107 (2002).

10 Kok, J., Mandemakers, K. & Bras, H. Van geboortebank tot collaboratory. Een reflectie op twintig jaar dataverzameling en onderzoek met de HSN. *Tijdschr. voor Soc. en Econ. Geschiedenis/ Low Ctries. J. Soc. Econ. Hist.* **6**, 3 (2009).

11 Schellekens, J. & van Poppel, F. Early-life conditions and adult mortality decline in Dutch cohorts born 1812–1921. *Popul. Stud. (NY).* (2016). doi:10.1080/00324728.2016.1223336

12 Adams, J. W., Kasakoff, A. & Kok, J. Migration over the life course in XIX th century netherlands and the American north: A comparative analysis based on genealogies and population registers. *Ann. Demogr. Hist. (Paris).* (2002). doi:10.3917/adh.104.0005

13 Hacker, J. D. Trends and Determinants of Adult Mortality in Early New England: Reconciling Old and New Evidence from the Long Eighteenth Century. *Soc. Sci. Hist.* (1997). doi:10.1017/s014555320001782x

14 Voland, E. & Dunbar, R. I. M. The impact of social status and migration on female age at marriage in an historical population in north-west Germany. *J. Biosoc. Sci.* **29**, 355–360 (1997).

15 Ruggles, S. Migration, Marriage, and Mortality: Correcting Sources of Bias in English Family Reconstitutions. *Popul. Stud. (NY).* **46**, 507–522 (1992).

16 Jonker, M. A. & van der Vaart, A. W. Correcting missing-data bias in historical demography. *Popul. Stud. (NY).* **61**, 99–113 (2007).

17 Kasakoff, A. B. & Adams, J. W. The effect of migration on ages at vital events: A critique of family reconstitution in historical demography. *Eur. J. Popul.* (1995). doi:10.1007/BF01264948

18 Thorvaldsen, G. Using NAPP Census Data to Construct the Historical Population Register for Norway. *Hist. Methods A J. Quant. Interdiscip. Hist.* **44**, 37–47 (2011).

19 van Dijk, I. K. *Death and the family. High mortality families and the life course, The Netherlands, 1812-1912.* (Enschede: Ipskamp, 2019).

20 Massey, C. G. Playing with matches: An assessment of accuracy in linked historical data. *Hist. Methods A J. Quant. Interdiscip. Hist.* **50**, 129–143 (2017).

21 Ruggles, S., Fitch, C. A. & Roberts, E. Historical Census Record Linkage. *Annu. Rev. Sociol.* (2018). doi:10.1146/annurev-soc-073117-041447

22 Vulsma, R. F. Burgerlijke stand en bevolkingsregister *[Civil Register and Population Register].* (Centraal Bureau voor Genealogie, 1988).

23 van den Berg, N. et al. Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nat. Commun.* **10**, 35 (2019).

24 Priester, P. *Geschiedenis van de Zeeuwse landbouw circa 1600-1910.* ('t Goy-Houten: HES Uitgevers, 1998).

25 Knotter, A. & Meijer, A. C. De gemeentelijke bevolkingsregisters, 1850-1920. *Den Haag Inst. voor Ned. Geschied.* (1995).

26 Stadsarchief Amsterdam (2018). Handleiding Gezinskaarten 1893-1939. Available at: https://archief.amsterdam/indexen/gezinskaarten_1893-1939/handleiding/index.nl.html.

27 Mandemakers, K. Historical sample of the Netherlands. In P. K. Hall, R. McCaa, & G. Thorvaldsen (Eds.). in *Handbook of International Historical Microdata for Population Research* 149–177 (2000).

28 IISG. (2018). About HSN: Sampling. Retrieved from https://socialhistory.org/en/hsn/sampling.

29 Janssens, *A. Family and social change. Canadian Journal of Sociology / Cahiers canadiens de sociologie* (Cambridge University Press, 1993). doi:10.1017/CBO9780511628696

30 Mandemakers, K. & Laan, F. LINKS dataset Genes Germs and Resources. WieWasWie Zeeland. Civil Certificates. (2017).

31 *HSN (2010b). Historical Sample of the Netherlands (HSN): Data set life courses release 2010.01 [Data file and code book].*

32 *HSN (2010a). Historical Sample of the Netherlands (HSN): Data set civil certificate release 2010.01 [Data file and code book].*

33 Mandemakers, K. https://socialhistory.org/en/hsn/hsn-releases. *HSN 2010.01 release* (2010).

34 Fu, Z., Boot, H. M., Christen, P. & Zhou, J. Automatic Record Linkage of Individuals and Households in Historical Census Data. *Int. J. Humanit. Arts Comput.* (2014). doi:10.3366/ijhac.2014.0130

35 Van Leeuwen, M. H. D. & Maas, I. *HISCLASS: A historical international social class scheme.* (Universitaire Pers Leuven, 2011).

36 Mandemakers, K. et al. HSN standardized HISCO-coded and classified occupational titles. (2018).

37 Janssen, F. & van Poppel, F. The Adoption of Smoking and Its Effect on the Mortality Gender Gap in Netherlands: A Historical Perspective. *Biomed Res. Int.* **2015**, 1–12 (2015).

38 van dijk, I. K. & Mandemakers, K. Like mother, like daughter. Intergenerational transmission of infant mortality clustering in Zeeland, the Netherlands, 1833-1912. *Hist. Life Course Stud.* 1–26 (2018).

39 Ekamper, P., van Poppel, F. & Mandemakers, K. Widening Horizons? The Geography of the Marriage Market in Nineteenth and Early-Twentieth Century Netherlands. in *Navigating Time and Space in Population Studies* (2011). doi:10.1007/978-94-007-0068-0_6

40 Desjardins, B. Bias in Age at Marriage in Family Reconstitutions: Evidence from French-Canadian Data. *Popul. Stud. (NY).* **49**, 165–169 (1995).

41 Jonker, M. A. & Van Der Vaart, A. W. A semi-parametric model for censored and passively registered data. *Bernoulli* (2001). doi:10.2307/3318600

42 Alter, G. & Oris, M. Childhood conditions, migration, and mortality: migrants and natives in 19th-century cities. *Soc. Biol.* (2005). doi:10.1080/19485565.2005.9989108

43 Lassetter, J. H. & Callister, L. C. The Impact of Migration on the Health of Voluntary Migrants in Western Societies: A Review of the Literature. *J. Transcult. Nurs.* (2009). doi:10.1177/1043659608325841

44 Hinde, A. Sex differentials in phthisis mortality in England and Wales, 1861–1870. *Hist. Fam.* (2015). doi:10.1080/1081602X.2015.1051077

45 Puschmann, P., Donrovich, R. & Matthijs, K. Salmon Bias or Red Herring? *Hum. Nat.* **28**, 481–499 (2017).

46 Reid, A. & Garrett, E. Mortality, Work and Migration. A Consideration of Age-specific Mortality from Tuberculosis in Scotland, 1861-1901. *Hist. life course Stud.* **6**, 111–132 (2018).

47 Campbell, C. D. Demographic Techniques: Family Reconstitution. *in International Encyclopedia of the Social & Behavioral Sciences* 138–142 (Elsevier, 2015). doi:10.1016/B978-0-08-097086-8.31012-1

CHAPTER 4

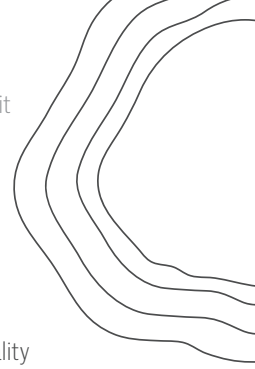
LONGEVITY DEFINED AS TOP 10% SURVIVORS AND BEYOND IS TRANSMITTED AS A QUANTITATIVE GENETIC TRAIT

Niels van den Berg[1,2,3]
Mar Rodríguez-Girondo[6]
Ingrid K. van Dijk[3]
Rick J. Mourits[3]
Kees Mandemakers[4]
Angelique A.P.O. Janssens[3]
Marian Beekman[1]
Ken Robert Smith[2] †
P. Eline Slagboom[1,5] †

[1] Department of Biomedical Data Sciences, section of Molecular Epidemiology,
Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands
[2] Department of Family and Consumer Studies; Population Sciences, Huntsman Cancer Institute,
University of Utah, 225 S. 1400 E. Rm 228 Salt Lake City, United States of America
[3] Radboud Group for Historical Demography and Family History,
Radboud University, Erasmusplein 1, 6525 HT Nijmegen, the Netherlands
[4] International Institute of Social History,
Cruquiusweg 31, 1019 AT Amsterdam
[5] Max Planck Institute for Biology of Ageing,
Joseph-Stelzmann-Str. 9b, D-50931 Cologne, Germany
[6] Department of Biomedical Data Sciences, section of Medical Statistics,
Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden the Netherlands

†. Jointly supervising authors

Nature Communications (January 2019)
https://doi.org/10.1038/s41467-018-07925-0



Abstract

Survival to extreme ages clusters within families. However, identifying genetic loci conferring longevity and low morbidity in such longevous families is challenging. There is debate concerning the survival percentile that best isolates the genetic component in longevity. Here, we use three-generational mortality data from two large datasets, UPDB (US) and LINKS (Netherlands). We study 20,360 unselected families containing index persons, their parents, siblings, spouses, and children, comprising 314,819 individuals. Our analyses provide strong evidence that longevity is transmitted as a quantitative genetic trait among survivors up to the top 10% of their birth cohort. We subsequently show a survival advantage, mounting to 31%, for individuals with top 10% surviving first and second-degree relatives in both databases and across generations, even in the presence of non-longevous parents. To guide future genetic studies, we suggest to base case selection on top 10% survivors of their birth cohort with equally long-lived family members.

Main

Human lifespan has a low heritability (12-25%)[1–4], whereas survival into extreme ages (longevity) clusters within families[5–9]. Studies showed that parents, siblings[5–7,9–12], and children[7,13–17] of longevous persons lived longer than first degree relatives of non-longevous persons or population controls. In addition, members of these longevous families seem to delay or even escape age-related diseases[18–21] and in fact, healthy ageing in such families is marked by well attuned immune systems and good metabolic health[22–24]. Understanding the genetic factors influencing longevity may provide novel insights into the mechanisms that promote health and minimize disease risk[1,25]. Identifying longevity loci, however, has been challenging and only a handful of genetic variants have been shown to associate with longevity across multiple independent studies[25–32]. The most consistent evidence has been obtained for variants in APOE and FOXO3A genes[25–30,33] in either genome-wide association studies (GWAS) or candidate gene studies.

The lack of consistent findings in longevity studies hampers comparative research and may be explained by genetic and environmental heterogeneity on one hand and uncertainty in defining the longevity trait itself, as illustrated by the large variation of longevity definitions on the other hand[1,3,20,25–32,34,7,35–38,10,13–17,19]. Establishing a threshold that best isolates the genetic component of longevity and including mortality information of family members is important because the environmentally-related increase in lifespan over recent decennia has caused an increase in longevity phenocopies. As a result, genetic longevity studies generally focus on singletons (i.e. individuals without longevous family members), selected based on one generation of mortality data[27,28,31,32,39]. Here, we aim to establish the threshold for longevity in unselected (for survival) multigenerational families and determine the importance of longevous family members for case selection so that those insights can be used in genetic studies to identify novel longevity loci.

We use the data available in the Utah Population Database (UPDB,Utah) and the LINKing System for historical family reconstruction (LINKS,Zeeland) based on US and Dutch citizens, respectively. Zeeland was a region with difficult living conditions compared to Utah (see methods section). In these datasets we identify 20,360 three-generational families (F1-F3) containing index persons (IPs, F2), their parents (F1), siblings (F2), spouses (F2), and children (F3) comprising 314,819 persons in total. First, we examine the association between the survival, measured as age at death, of IPs (F2) and the number of parents (F1) and siblings (F2) belonging to the top 1-60% of their birth cohort, in a cumulative way (comparing mutually inclusive percentile groups). Second, we determine the survival percentile threshold that drives the cumulative effects as a criterion for defining human longevity by investigating IP (F2) survival when divided into mutually exclusive groups based on the longevity of their parents (F1) and siblings (F2). Third, we focus on the top 10% parents and siblings to investigate whether longevous and non-longevous parents, with increasing numbers of longevous siblings, transmit longevity to the IPs. Fourth, we confirm our findings in the next generation (F3) by examining the association between the survival, measured as age at death or last observation, of IPs' children (F3) and longevity of IPs (F2), their spouses (parents, F2) and siblings (aunts and uncles, F2). Finally, we explore potential environmental influences by studying spouses (F2) of longevous IPs (F2).

Results

Study population

We identified three generations of families in the UPDB and LINKS covering 10,246 and 10,114 families, respectively, who were centered around a single index person (IPs,F2) per family *(Figure 1)*.

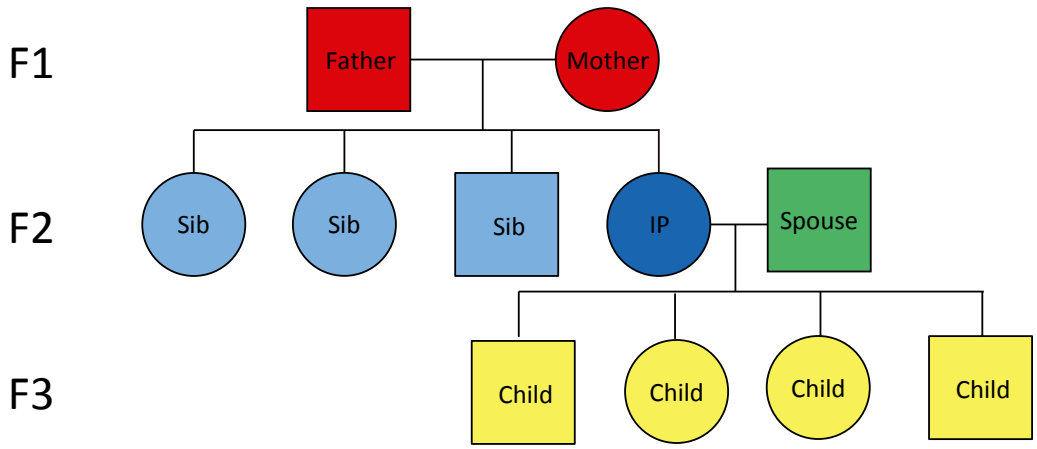


Figure 1: Conceptual pedigree of a 3 filial (F) generation family in the study.

This figure represents a hypothetical family from the UPDB or LINKS covering 3 filial (F) generations. Circles represent women, Squares represent men. Dark blue: Index persons (F2), red: Parents (F1), light blue: Siblings of IP (F2), green: Spouses of IP (F2), yellow: Children of IP (F3). IP: Index Person, Sib: Sibling, F: Filial.

We identified parents (F1, N_{UPDB}=20,492 & N_{LINKS}=20,228), siblings (F2, N_{UPDB}=54,144 & N_{LINKS}=53,978), spouses (F2, N_{UPDB}=11,230 & N^{LINKS}=10,788), and children (F3, N^{UPDB}=61,104 & N^{LINKS}=62,495) for all IPs in both datasets *(Table 1)*. IPs were born between 1767 and 1902 in the UPDB, and between 1797 and 1902 in LINKS. In the UPDB, 51% of the IPs were female, compared to 53% in LINKS. The IPs' mean age at death was 70.88 (SD=16.03) years in the UPDB and 63.86 (SD=17.99) years in LINKS. No IPs were censored, as they were selected to have an available birth and death date. In addition, *Supplementary Figure 1* shows the age at death distribution for the IPs in both datasets. In the following sections we explore associations between IP survival and the number of 1-60% surviving parents and siblings in a cumulative analysis and subsequently identify in mutually exclusive IP groups the survival percentile threshold that drives the cumulative effect and demarcates longevity (see methods section).

Table 1: Overview of UPDB and LINKS IPs and their first degree relatives and spouses

	Parents F1	IPs F2	Siblings F2	Spouses F2	Children F3
UPDB					
Number, N	20492	10246	54144	11230	61104
Deceased, N (%)	19191 (94)	10246 (100)	45701 (84)	10256 (91)	54076 (88)
Female, N (%)	10246 (50)	5193 (51)	26159 (48)	5742 (51)	29675 (49)
Range birth cohorts	1753 - 1884	1767 - 1902	1756 - 1932	1768 - 1922	1792 - 1937
Mean ad or al, years (SD)	68.96 (16.10)	70,88 (16.03)	44.32 (33.60)	69.11 (17.58)	54.87 (32.09)
Mean ad, years (SD)	70.05 (15.23)	70.88 (16.03)	49.98 (32.45)	70.75 (16.43)	57.98 (31.28)
Missing age, N (%)	403 (2)	0 (0)	799 (1)	345 (3)	306 (1)
Censored , N (%)	898 (4)	0 (0)	7644 (14)	629 (6)	6722 (11)
LINKS					
Number, N	20228	10114	53978	10788	62495
Deceased, N (%)	15536 (77)	10114 (100)	40093 (74)	8819 (82)	43896 (70)
Female, N (%)	10114 (50)	5338 (53)	25946 (48)	5193 (48)	30347 (49)
Range birth cohorts	1740 - 1877	1797 - 1902	1796 - 1916	1775 - 1907	1818 - 1952
Mean ad or al, years (SD)	54.65 (20.66)	63.86 (17.99)	20.84 (27.99)	59.04 (21.23)	24.86 (30.06)
Mean ad, years (SD)	62.64 (16.15)	63.86 (17.99)	23.94 (30.76)	65.70 (17.20)	29.59 (33.63)
Missing age, N (%)	49 (<1)	0 (0)	14 (<1)	27 (<1)	21 (<1)
Censored , N (%)	4643 (23)	0 (0)	13878 (26)	1942 (18)	18578 (30)

ad = age at death, al = age at last observation, IPs = Index Persons. Missing age means that we have no observation at all. Number death refers to the number of deceased individuals.

IP survival advantage with top 1-60% parents and siblings

For a first examination of the association between the number of parents (1 or 2,F1) and siblings (1 or 2+,F2) and IP (F2) survival and to explore if a larger level of family aggregation, in terms of numbers of parents (F1) and siblings (F2), was more evident at extreme survival percentiles, we fitted Cox regressions for each subsequent survival percentile (1^{st} to 60^{th} percentile).

Figure 2A and C show that IPs with 1 parent belonging to the top 1-60%, had a survival advantage over IPs without a parent belonging to the top 1-60%. This was shown by the lowest observed statistically significant hazard ratio (HR) of 0.80 (95% $CI_{max\text{-}top\ 1\%}$=0.73-0.88) in the UPDB and 0.74 (95% $CI_{max\text{-}top\ 1\%}$=0.65-0.85) in LINKS where max refers to the age with the largest effect and CI to confidence interval. These HRs indicate a 20% and 26% lower hazard of dying respectively and from here we will refer to this as a 20% and 26% survival advantage. Having 2 parents belonging to the top 1-60% provides a stronger survival advantage to IPs ($HR_{max\text{-}top\ 2\%\text{-}UPDB}$=0.64 (95%CI=0.50-0.84) and $HR^{max\text{-}top\ 14\%\text{-}LINKS}$=0.72 (95% CI=0.65=0.80)), although *Figure 2* shows that the power to detect survival effects of IPs with 2 longevous parents up to the 10th percentile was weak for LINKS due to low group sizes.

The association of IP survival with longevous siblings is shown in *Figure 2B and D*. The maximum statistically significant HRs for IPs with 1 longevous sibling were 0.76 (95% $CI_{max\text{-}top\ 1\%}$=0.62-0.92) and 0.79 (95% $CI_{max\text{-}top\ 5\%}$=0.67-0.93) in the UPDB and LINKS respectively. For IPs with 2 or more longevous siblings these HRs were 0.65 (95% $CI_{max\text{-}top\ 3\%\text{-}UPDB}$=0.51-0.84) and 0.67 (95% $CI_{max\text{-}top\ 8\%\text{-}LINKS}$=0.50-0.90). The slopes in *Figure 2A-D* show a slight increase of IP survival advantage with the increase in percentile score. For example, IPs with parents with the best survival (the left most end of the x-axis) had lower hazard rates than IPs with the least survival (the right most end of the x-axis). We conclude that IP survival when expressed in HRs, both in the UPDB and LINKS, increased with the number of longevous parents, with the number of longevous siblings and, though modestly, with the increase of parent and sibling survival percentile scores as observed in *Figure 2*.

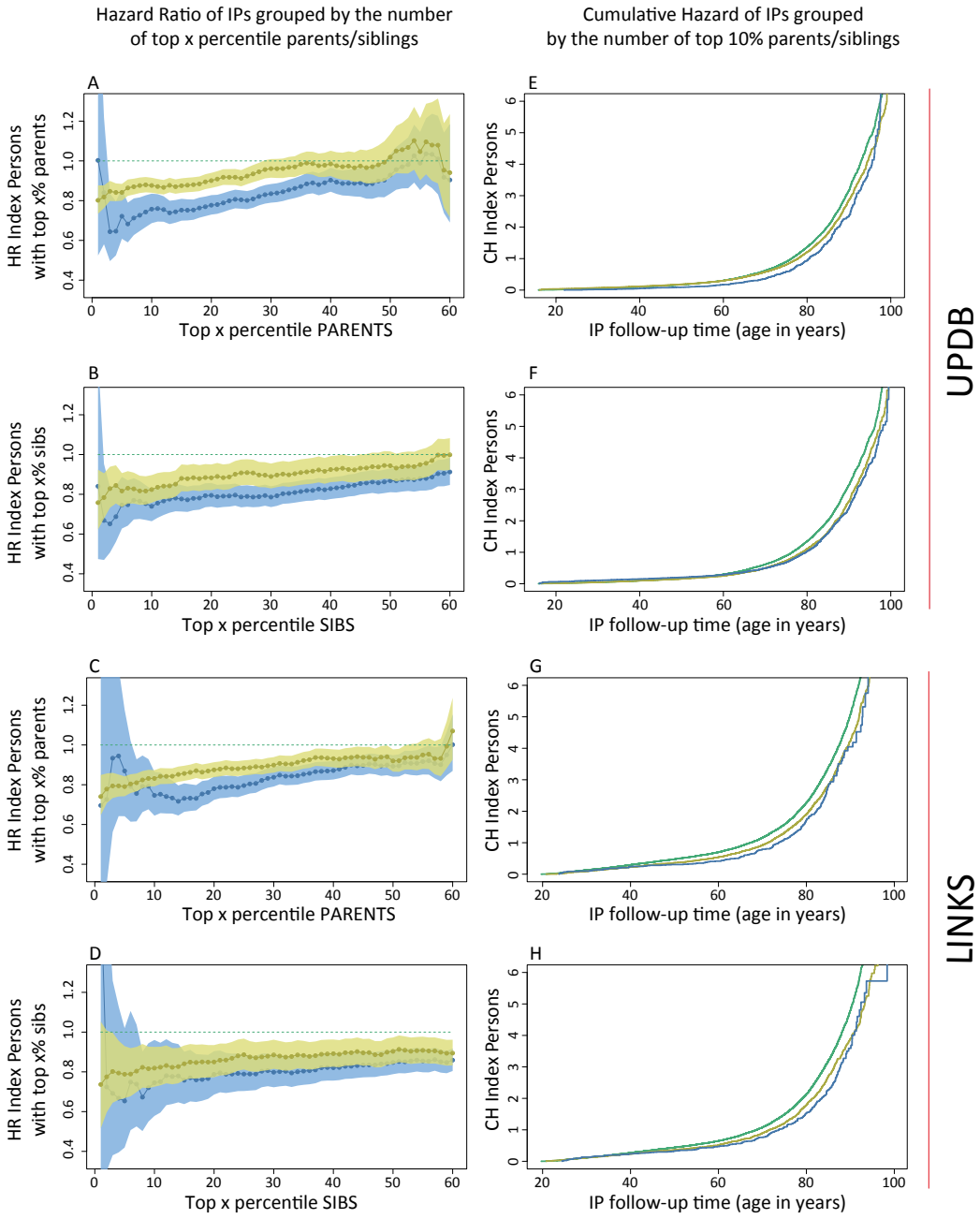


Figure 2: Survival of IPs with relatives belonging to the 1st until 60th percentile survivors of their birth cohort.

This figure depicts the Hazard Ratio (HR) for IPs (left column, panel A-D) with 1 and 2 parents (panel A and C) or 1 and 2+ siblings (panel B and D) belonging to the top x percentile (x = 1,2,3, ..., 60) survivors of their birth cohort. The percentile groups (x-axis) are mutually inclusive, meaning that a first degree family member who belonged to the top 1% also belonged to the top 5% etc. The figure also depicts the Cumulative Hazard (CH) for

index persons (IPs, right column, panel E-H) with 1 and 2 parents (panel E and G) or 1 and 2+ siblings (panel F and H) who belong to the top 10%. Green (dotted) lines present the reference group of 0 top x percentile parents or siblings, yellow lines represent 1 top x percentile parents or siblings, blue lines represent 2 or 2+ top x percentile parents or siblings. left column: x-axes represent the top x birth cohort based survival percentile, the y-axes represent the hazard ratio (HR) of dying for IPs having 1 and 2 or 2+ top x percentile parents or siblings compared to having 0 top x percentile parents or siblings. Right column: x-axes represent IP years of survival, y-axes represent the IPs' cumulative hazard of dying while having 1 and 2 or 2+ top 10th percentile parents or siblings compared to having 0 top 10th percentile parents or siblings. All estimates are adjusted for religion (UPDB only), sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, twin birth, and number of top 10% parents or number of top 10% siblings for the sibling and parent analyses respectively. Error bars represent confidence intervals.

Top 10-15% surviving family members demarcates longevity

To determine the survival percentile threshold that drove the survival advantage of IPs (F2) with the number of top 1-60% parents (F1), as shown in *Figure 2*, we constructed 6 mutually exclusive IP (F2) groups (g) based on the survival percentiles of F1 parents (g1=[$\geq 0^{th}$ & $\leq 1^{th}$ percentile], g2=[$\geq 1^{th}$ & $\leq 5^{th}$ percentile], g3=[$\geq 5^{th}$ & $\leq 10^{th}$ percentile], g4=[$\geq 10^{th}$ & $\leq 15^{th}$ percentile], g5=[$\geq 15^{th}$ & $\leq 20^{th}$ percentile], g6=[$\geq 20^{th}$ & $\leq 100^{th}$ percentile], see methods section) and compared groups 1-5 with group 6. *Figure 3A and B* show the HRs of IP groups for the UPDB and LINKS and is supplemented by the IP age at death and survival percentile variation, as depicted in *Supplementary Figure 2. Figure 3A and B* illustrate that IPs in group 1, 2, 3, and 4 had a significant survival advantage compared to group 6, with the lowest HR for group 1 in both the UPDB and LINKS ($HR_{max\text{-}UPDB}$=0.76 (95% CI=0.67-0.86) and $HR_{max\text{-}LINKS}$=0.72 (95% CI=0.60-0.86)). Group 5 did not statistically differ from group 6 ($HR_{group5\text{-}UPDB}$=1 (95% CI=0.91-1.10) and $HR_{group5\text{-}LINKS}$=0.96 (95% CI=0.87-1.05)) and thus, these effects indicate that the top 15% surviving parents drove the association with the survival advantage of IPs as shown in *Figure 2*.

In the same way we investigated the association of IPs' (F2) survival with that of siblings (F2). *Figure 3C and D* show a survival advantage of IPs in UPDB group 1-3 and LINKS group 2 and 3 as compared to group 6 with the lowest HR for group 1 (UPDB) and group 2 (LINKS) ($HR_{group1\text{-}UPDB}$=0.70 (95% CI=0.59-0.85) and $HR_{group2\text{-}LINKS}$=0.77 (95% CI=0.65-0.92)), respectively. Group 4 and 5 did not significantly differ from group 6 ($HR_{group4\text{-}UPDB}$=0.99 (95% CI=0.88-1.12) and $HR_{group4\text{-}LINKS}$=0.86 (95% CI=0.73-1.02)) which indicated that both in the UPDB and LINKS the top 10% surviving siblings drove the association with the survival advantage of IPs as shown in *Figure 2*.

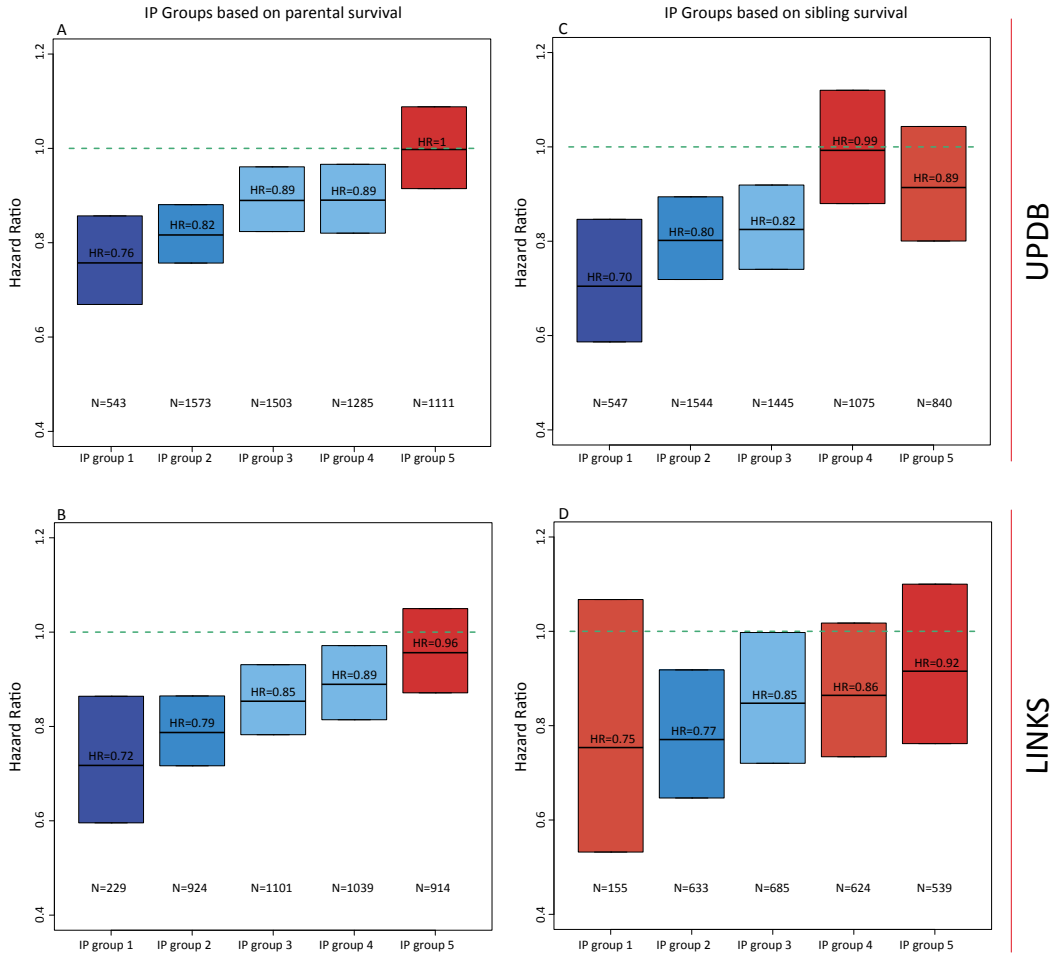


Figure 3: Hazard ratio for IPs grouped by their relatives' survival in mutually exclusive groups:

Parent and Sibling Groups: group 1 = IPs of whom the longest lived parent/sibling belonged to the [≥0th & ≤1th percentile] of their birth cohort, group 2 = IPs of whom the longest lived parent/sibling belonged to the [≥1th & ≤5th percentile], group 3 = IPs of whom the longest lived parent/sibling belonged to the [≥5th & ≤10th percentile], group 4 = IPs of whom the longest lived parent/sibling belonged to the [≥10th & ≤15th percentile], group 5 = IPs of whom the longest lived parent/sibling belonged to the [≥15th & ≤20th percentile], group 6 = IPs of whom the longest lived parent/sibling belonged to the [≥20th & ≤100th percentile]. The left column (Panel A and B) show the HRs of IP group 1-5 compared to group 6 and depicts a parental grouping. The right column (Panel C and D) show the HRs of IP group 1-5 compared to group 6 and depicts a sibling grouping. Groups were colored by the extremity of the HR. The darker the blue the stronger the survival benefit, the darker the red, the weaker the survival benefit and the effect was not significant in with the red colors. The green lines represent the reference category, which is group 6. $N_{green\ line}$ at the top-right = 4759, $N_{green\ line}$ at the top-left = 4227, $N_{green\ line}$ at the bottom-right = 7477, $N_{green\ line}$ at the bottom-left = 5907. All estimates are adjusted for religion (UPDB only), sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, twin birth, and number of top 10% parents or number of top 10% siblings for the sibling and parent analyses respectively. Error bars represent confidence intervals.

Based on the results presented in the cumulative and mutually exclusive group analyses we focused on the top 10% surviving family members because the mutually exclusive group analysis (analysis 2, *Figure 3*) indicated longevity effects up to the top 10% and 15% for siblings and parents respectively. Using the top 10% is consistent between the two groups and is a conservative choice. Furthermore, the cumulative analysis (analysis 1, *Figure 2*) indicated that the top 10% was a reasonable trade-off between effect size and group size (power) within and between the UPDB and LINKS. Hence, we explored the familial clustering of longevity and the influence of covariates for the top 10% surviving parents and siblings and verified all results in the subsequent generation (F3). Next to the top 10% we also conducted our analyses on the top 5% which are illustrated in Supplementary *Figures 3-5* and *Supplementary Tables 1-5*.

Additive association between 10% surviving relatives and IPs

Figure 2E-H shows the cumulative hazard (CH) curves for IPs (F2) with 0, 1 and 2 or more, or exactly 2 parents/siblings (F1/F2) belonging to the top 10% of their birth cohorts and we show Kaplan-Meier and Nelson-Aalen baseline measures in *Supplementary Figure 6*. Both in the UPDB and LINKS, the survival advantage associated with the number of top 10% siblings appeared to start during the beginning (45 years in LINKS) and end (65 years in the UPDB) of the mid-life period. In both the UPDB and LINKS, the survival advantage of IPs with the number of top 10% parents started at the age of 40 years. It should be noted that early life effects could not be tested for, because IPs were selected on having a child for the construction of three generation families.

Table 2 accompanies *Figure 2E-H* by showing the HRs for the number of top 10% parents (F1) and siblings (F2) and for the covariates we used to adjust the analyses. IPs with 1 top 10% parent had a maximum survival advantage of 12% and 18% compared to IPs without such a parent (HR=0.88$_{\text{max-UPDB}}$ (95%CI=0.83-0.92) and HR=0.82$_{\text{max-LINKS}}$ (95%CI=0.78-0.86)). The maximum statistically significant survival advantage for IPs with 2 top 10% parents was 27% and 31% (HR$_{\text{max-UPD}}$B=0.73 (95%CI=0.65-0.83) and HR$_{\text{max-LINKS}}$=0.69 95%CI=0.58-0.82)). The maximum statistically significant HR for having 1 top 10% sibling was 0.82 (95% CI$_{\text{UPDB}}$=0.76-0.90) and 0.82 (95% CI$_{\text{LINKS}}$=0.73-0.93). For 2+ top 10% siblings the HR was 0.74 (95% CI$_{\text{UPDB}}$=0.66-0.82) and 0.75 (95% CI$_{\text{LINKS}}$=0.58-0.96). The survival advantage of IPs with 1 and 2 or more, or exactly 2 top 10% siblings and parents respectively was independent of covariates such as sibship size and religion (LDS church affiliation, *Table 2* and *Supplementary Table 6 and 7*). Religious IPs from Utah had a lower HR than non-religious persons (HR$_{\text{UPDB}}$=0.73 (95% CI=0.65-0.81)) and in the UPDB we observed that sibship size had a small influence on the survival of IPs (HR$_{\text{UPDB}}$=1.01 (95% Ci=1.00-1.02) whereas in LINKS, sibship size had no significant effect HR$_{\text{LINKS}}$=1.01 (95% CI=1.00-1.02)). The survival of IPs increased with the increase of birth cohort (HR$_{\text{UPDB and LINKS}}$=0.99 (95% CI=[>0.99<1.00])) and women had a better survival than men in the UPDB, (HR$_{\text{UPDB}}$=0.71 (95% Ci=0.67-0.76)) but not in LINKS (HR$_{\text{LINKS}}$=1.01 (95% Ci=0.96-1.06)). Furthermore, In Utah, high socio-economic status IPs outlived low socio-economic status IPs whereas this was not the case in LINKS. The association between the number of longevous parents/ siblings and the survival of IPs were independent of each other and no other statistically significant effect was observed for having both longevous parents and siblings. Moreover, the number of longevous siblings showed a strong association with the survival of IPs when both parents were non-longevous. The HR for 1 longevous sibling was 0.85 (95% CI=0.79-0.91) and the HR for 2 or more longevous siblings was 0.78 (95% CI=0.67-0.90) in the UPDB. The HR for 1 longevous sibling was 0.78 (95% CI=0.72-0.85) and the HR for 2 or more longevous siblings was 0.72 (95% CI=0.53-0.99) in LINKS *(Supplementary Table 8)*. In a final step, we observed no evidence

that the association of IP survival and parental longevity depended on maternal or paternal effects, for example through transmission preferentially via the mother or father *(Supplementary Table 9).* Likewise, the association of IP survival and parental longevity did not depend on the sex of the IPs, meaning that this association was equal for sons and daughters *(Supplementary Figure 7 and Supplementary Table 10).*

Table 2: Survival analysis for IPs with top 10% parents and siblings

	UPDB			LINKS		
	N (mean)	HR (95% CI)	p-value	N (mean)	HR (95% CI)	p-value
Top 10% parents (F1)						
0 (ref)	6640 (0.65)			7861 (0.78)		
1	3167 (0.31)	0.88 (0.83-0.92)	$2.94*10^{-7}$	2096 (0.20)	0.82 (0.78-0.86)	$1.27*10^{-13}$
2	439 (0.4)	0.73 (0.65-0.83)	$4.38*10^{-7}$	184 (0.2)	0.69 (0.58-0.82)	$1.91*10^{-5}$
Top 10% sibs (F2)						
0 (ref)	6720 (0.66)			8644 (0.85)		
1	2495 (0.24)	0.82 (0.76-0.90)	$6.85*10^{-6}$	1256 (0.13)	0.82 (0.73-0.93)	$1.38*10^{-3}$
2+	1031 (0.10)	0.74 (0.66-0.82)	$4.15*10^{-8}$	214 (0.2)	0.75 (0.58-0.96)	$2.30*10^{-2}$
LDS (F2)						
0 - non-religious (ref)	2753 (0.27)					
1 - baptized	512 (0.05)	0.73 (0.65-0.81)	$1.49*10^{-8}$	NA	NA	NA
2 - baptized +	6736 (0.66)	0.80 (0.76-0.85)	$2.33*10^{-15}$	NA	NA	NA
endowment	245 (0.02)	0.85 (0.73-0.99)	$4.24*10^{-2}$	NA	NA	NA
3 - missing	10246 (6.28)	1.01 (1.00-1.02)	$2.23*10^{-2}$	10114 (6.34)	1.01 (1.00-1.02)	$9.02*10^{-2}$
Sibship size (F2)	10246 (1868)	0.99	$2.66*10^{-09}$	10114 (1835)	0.99	$<1.00*10^{-15}$
		(>0.99<1.00)			(>0.99<1.00)	
Birth cohort, years (F2)						
Sex (F2)	5053 (0.49)			4776 (0.48)		
Man (ref)	5193 (0.51)	0.71 (0.67-0.76)	$<1.00*10^{-15}$	5338 (0.52)	1.01 (0.96-1.06)	$7.53*10^{-1}$
Women						
SES – OCC_1950 (F2)	315 (0.03)			67 (0.01)		
0 - High (ref)	1482 (0.14)	1.16 (1.01-1.34)	$3.95*10^{-2}$	645 (0.06)	0.88 (0.68-1.14)	$3.42*10^{-1}$
1	400 (0.04)	1.19 (1.00-1.40)	$4.95*10^{-2}$	536 (0.05)	0.97 (0.75-1.27)	$8.42*10^{-1}$
2	352 (0.03)	1.24 (1.05-1.48)	$1.38*10^{-2}$	62 (0.01)	0.76 (0.53-1.10)	$1.45*10^{-1}$
3	187 (0.02)	1.14 (0.93-1.40)	2.09*10-1	71 (0.01)	0.99 (0.70-1.40)	$9.41*10^{-1}$
4	891 (0.09)	1.31 (1.13-1.52)	$4.22*10^{-4}$	733 (0.07)	0.80 (0.62-1.04)	$9.19*10^{-2}$
5	668 (0.07)	1.34 (1.15-1.56)	$2.13*10^{-4}$	311 (0.03)	0.86 (0.65-1.13)	$2.71*10^{-1}$
6	522 (0.05)	1.27 (1.08-1.50)	$4.14*10^{-3}$	759 (0.08)	0.84 (0.65-1.10)	$2.01*10^{-1}$
7	168 (0.02)	1.21 (0.97-1.50)	$8.91*10^{-2}$	574 (0.06)	0.85 (0.65-1.11)	$2.35*10^{-1}$
9 - Low	562 (0.05)	1.48 (1.26-1.73)	$1.70*10^{-6}$	3656 (0.36)	0.83 (0.65-1.07)	$1.56*10^{-1}$
999 - missing	4699 (0.46)	1.61 (1.40-1.84)	$9.54*10^{-12}$	2700 (0.26)	0.93 (0.72-1.20)	$5.95*10^{-1}$
Log likelihood	-60719	-72239				

Table corresponds to the CH curves in the top and bottom right panel of Figure 2. Means represent a mean for a continuous variable and a proportion for a categorical variable. Additional covariates are: age mom at birth, birth order, birth intervals (in years), twin birth. When the p-value was lower than 1.00e-15 we indicated the P-value as <1.00*10-15. LDS = the church of Jesus Christ of latter-day saints (Mormon church), SES = socio-economic status, OCC = occupational coding scheme of 1950, CI = confidence interval, CH = cumulative hazard. P-values are estimated with cox regression.

Survival advantage for children with longevous relatives

We explored the robustness of our findings in F1 and F2 by examining the association between the longevity of IPs (F2), their spouses (F2) and siblings (F2) and the survival of IPs' children (F3). We investigated whether longevity was transmitted from IPs (F2) to their children (F3) and if the children (F3) with longevous aunts and uncles (siblings of the IPs,F2) had a survival advantage compared to children (F3) without longevous aunts and uncles (F2). To test this, we fitted Cox regressions, with a random effect (frailty) to adjust for within-family relations of the F3 children. *Table 3* shows that children of a top 10% surviving IP had a HR of 0.86 (95% CI_{UPDB}=0.84-0.89) in the UPDB and 0.85 in LINKS (95% CI_{LINKS}=0.82-0.88) compared to children without a top 10% IP. Moreover, results indicated that children with two top 10% parents (IPs and spouses) had a HR of 0.77 (95% CI_{UPDB}=0.72-0.82) in the UPDB and 0.77 (95% CI_{LINKS}=0.71-0.84) in LINKS. Similar to the IPs, we observed: (1) that the survival of children did not depend on maternal or paternal effects *(Supplementary Table 9)* and (2) that the association between parents and offspring was equal for sons and daughters *(Supplementary Table 10)*.

Children with 1 or more top 10% aunts or uncles had a 4-16% survival advantage compared to children without such aunts or uncles ($HR_{min\text{-}UPDB}$=0.96 (95% CI=0.93-0.99) and $HR_{max\text{-}LINKS}$=0.84 (95% CI=0.78-0.92)), and this effect was independent of having a top 10% parent (either the IP or the IP's spouse). A stratified analysis showed that the survival benefit for children with the number of top 10% aunts and uncles was still strongly present when the IP and the IP's spouse were non-longevous ($HR_{min\text{-}UPDB\ -\ 1\ aunt/uncle}$=0.96 (95% CI=0.93-0.99) and $HR_{max\text{-}LINKS\ -\ 2+\ aunts/uncles}$=0.81 (95% CI=0.73-0.90)) *(Supplementary Table 11)*. Lastly, Supplementary *Figure 8* shows that the survival benefit for children of a longevous IP and a longevous IP with a longevous spouse (i.e. 1 or 2 longevous parents) started from birth (LINKS) and very early in life (UPDB).

Table 3: Frailty survival analysis for Children of IP's with top 10% IPs and aunts and uncles

		UPDB			LINKS	
	N (mean)	**HR (95% CI)**	**p-value**	**N (mean)**	**HR (95% CI)**	**p-value**
Top 10% IP (F2)						
0 non LL (ref.)	48619 (0.80)			53378 (0.85)		
1 LL	12179 (0.20)	0.86 (0.84-0.89)	$<1.00*10^{-15}$	9096 (0.15)	0.85 (0.82-0.88)	$<1.00*10^{-15}$
Top 10% aunts and uncles (F2)						
0 (ref.)	39474 (0.65)			53228 (0.85)		
1	15134 (0.25)	0.96 (0.93-0.99)	3.19*10-3	7817 (0.12)	0.96 (0.92-0.99)	$1.90*10^{-2}$
2+	6190 (0.10)	0.92 (0.88-0.96)	$4.33*10^{-5}$	1429 (0.3)	0.84 (0.78-0.92)	$5.47*10^{-5}$
Sibshipsize (F3)	60798 (8.89)	1.02 (1.01-1.02)	$<1.00*10^{-15}$	62474 (8.52)	1.00 (>0.99<1.00	$7.87*10^{-1}$
Birth year (F3)	60798 (1892)	0.99 (>0.99<1.00)	$<1.00*10^{-15}$	62474 (1867)	0.99 (>0.99<1.00)	$2.37*10^{-11}$
Sex (F3)						
Man (ref.)	31258 (0.51)			32136 (0.52)		
Women	29540 (0.49)	0.62 (0.60-0.63)	$<1.00*10^{-15}$	30338 (0.48)	0.64 (0.63-0.66)	$<1.00*10^{-15}$
Famid intercept (variance)	60798 (1.00)	0.34 (0.11)		62474 (1.00)	0.34 (0.11)	
BIC	60798 (1.00)	-23756.54		62474 (1.00)	-21477.23	

Additional covariates are: birth order, birth intervals (years), age mom at birth. Religion, Socio-economic status, twin birth have been stratified. When the p-value was lower than 1.00e-15 we indicated the P-value as <1.00e-15. BIC = Bayesian Information Criterion. Famid = family identifier, CI = confidence interval, LL = long lived. P-values are estimated with cox regression.

Spouses live longer in Zeeland but not in Utah

Familial clustering of longevity may depend on (later life) shared environmental effects which could also provide survival benefits to the spouses (F2) of longevous IPs (F2). Hence, we divided the spouses (F2) into mutually exclusive groups according to the survival percentiles of the IPs (see methods). *Figure 4A and 4C* show that none of the spouse groups in the UPDB differed from reference group 6 or from any of the other groups, indicating no survival benefit for spouses. In LINKS *(Figure 4B and 4D)*, spouses of IPs with the highest survival percentile (group 2) had a 14% ($HR_{group2\text{-}LINKS}$=0.86 (95% CI=0.78-0.94)) survival advantage compared to group 6 spouses. This survival advantage was similar for spouses of IPs in group 3, 4, and 5 ($HR_{group3\text{-}LINKS}$=0.86 (95% CI=0.80-0.94) ; $HR_{group4\text{-}LINKS}$=0.92 (95% CI=0.85-0.99) ; $HR_{group5\text{-}LINKS}$=0.86 (95%CI=0.79-0.93)). For Group 1 the effect was comparable but not significant ($HR_{group1\text{-}LINKS}$=0.85 (95% CI=0.70-1.04)), the test in group 4 did not meet Bonferroni correction for multiple testing.

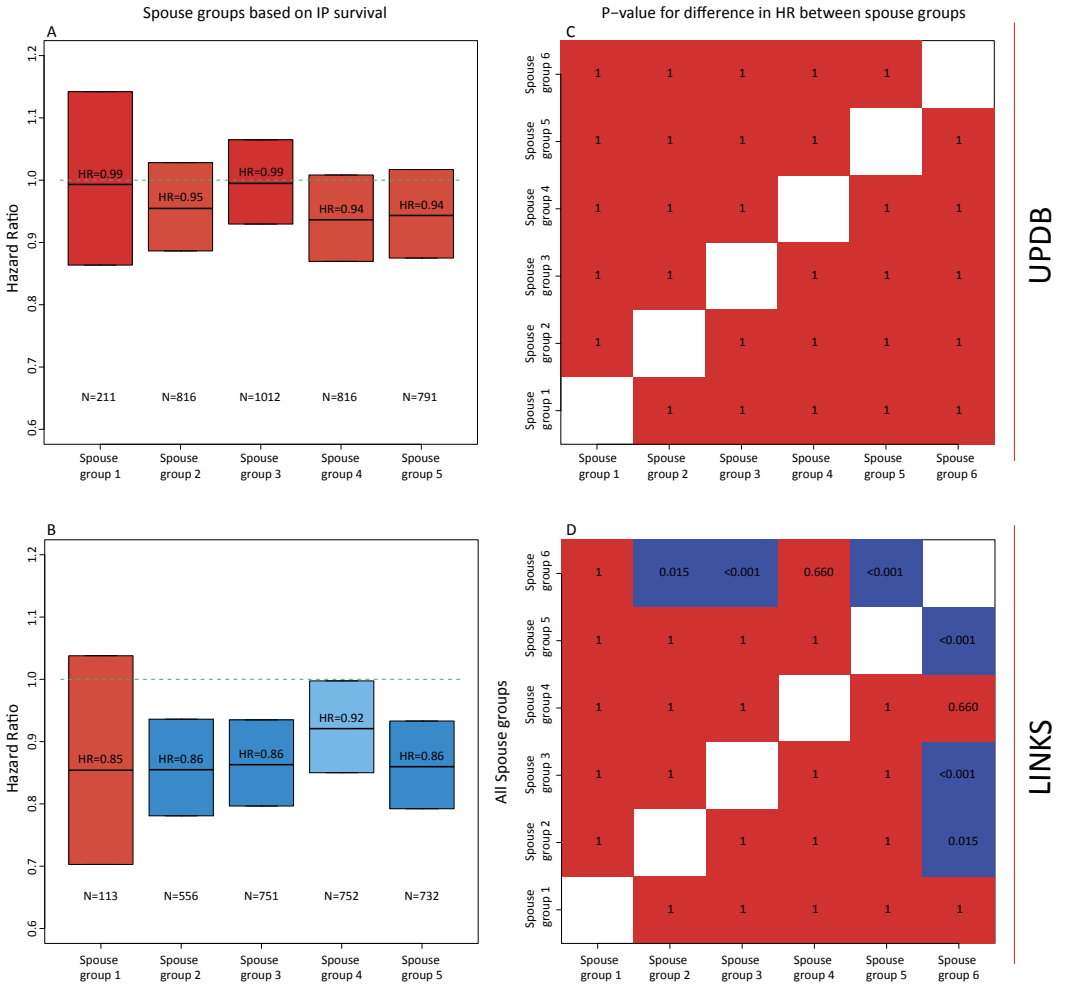


Figure 4: Hazard ratio for spouses grouped by IP survival in mutual exclusive groups.

Spouse Groups: group 1 = Spouses of whom the IP belonged to the [≥0th & ≤1th percentile] of their birth cohort, group 2 = Spouses of whom the IP belonged to the [≥1th & ≤5th percentile] of their birth cohort, group 3 = Spouses of whom the IP belonged to the [≥5th & ≤10th percentile] of their birth cohort, group 4 = Spouses of whom the IP belonged to the [≥10th & ≤15th percentile] of their birth cohort, group 5 = Spouses of whom the IP belonged to the [≥15th & ≤20th percentile] of their birth cohort, group 6 = Spouses of whom the IP belonged to the [≥20th & ≤100th percentile] of their birth cohort. The left column (Panel A and B) show the HRs of group 1-5 compared to group 6. Groups were colored by the extremity of the HR. The darker the blue the stronger the survival benefit, the darker the red, the weaker the survival benefit and the effect was not significant in with the red colors. The green lines represent the reference category, which is group 6. $N_{green\ line}$ at the top-left = 8065, $N_{green\ line}$ at the bottom-left = 7887. The right column (panel C and D) represents a post-hoc test of all groups and illustrates the p-values for the differences in HR between the spouse groups. P-values are estimated with cox regression. Blue color indicates a statistically significant effect after bonferroni correction, red color indicates a non-statistically significant effect after bonferroni correction. All estimates are adjusted for religion (UPDB only), sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, and twin birth. Error bars represent confidence intervals.

Discussion

Human longevity clusters within specific families. Insight into this clustering is important, especially to improve our understanding of genetic and environmental factors driving healthy aging and longevity. The analyses of the UPDB and LINKS datasets, which cover different environmental circumstances, provide strong evidence that for longevous (up to the top 10%) survivors and their families, longevity is transmitted as a quantitative genetic trait, regardless of parental and offspring sex. The main observations supporting this notion are (1) in both datasets the survival of F2 index persons (IPs), and their F3 children, increased with each additional longevous parent (F1 and F2) and sibling (F2), (2) in both datasets the survival of IPs (F2) increased with the number of longevous siblings (F2) in the absence of longevous parents (F1) and likewise the survival of IPs' children (F3) increased with the number of longevous aunts and uncles in the absence of longevous parents. Finally, (3) both datasets indicate an absence of a sex specific pattern.

Previous studies of smaller sample size than the current study, usually focusing on two generations of selected data (for mortality or geographical locations) identified (1) an increase in the heritability of lifespan with parental age[8,40,41] and showed high recurrence risks between parental and offspring or sibling longevity. Thus, providing indications that the heritability of longevity may be stronger than that of lifespan[5,7,13,14,42], (2) that sibling relative risks beyond the top 5% survivors might not increase in a linear fashion[12] and that this non-linearity may indicate the existence of a longevity threshold[43], and (3) longevity recurrence risks for siblings or parents of selected longevous individuals[5–7,9–12] and showed increased survival probabilities and longevity recurrence risks for children of longevous parents[7,13–17]. Here we used two unique, large three-generational datasets (314,819 individuals in 20,360 families) which were unselected for survival and cover multiple geographical areas. We utilized these datasets to robustly identify a longevity threshold by showing that the association between IP survival and the survival of parents and siblings was not linear but in fact was driven by the oldest, up to the top 10%, surviving parents and siblings. We further showed that the survival of F2 IPs (and their F3 children) increased with each additional longevous parent (F1 and F2), and sibling (F2). We extended these analyses by showing that the survival for children of IPs increased with each additional longevous aunt or uncle. We also extended the analyses by investigating the association between IP survival and sibling longevity, and between the survival of IPs' children and the longevity of their aunts or uncles in the absence of longevous parents.

Longevity was transmitted even if parents themselves did not become longevous, which supports the notion that a beneficial genetic component was transmitted. Likewise, the identified associations are additive in the sense that an increase in the number of parents, siblings, or aunts and uncles is associated with an increase in the survival of IPs and the children of IPs. This additive pattern is not necessarily expected if the findings are due to other, non-genetic, factors that cluster within families (for example wealth). This evidence is strengthened by the fact that similar additive associations were identified for IPs and children of IPs without longevous parents but with longevous siblings or aunts and uncles (where the latter generally share less environmental influences with the IPs). Further evidence for the transmission of a genetic component was shown by the fact that none of the tested environmental confounders affected the associations between parental/sibling longevity and IP/children survival, as will be discussed further on in the discussion section. In addition, the fact that we observed very similar results between the two databases, which cover populations with vastly

different environmentally related mortality regimes, significantly adds to the generalizability of our observations regarding the associations between parental/sibling longevity and IP (F2) and children (F3) survival.

We showed that spouses (F2) who married longevous IPs (F2) did not live significantly longer than spouses (F2) who married a non-longevous IP (F2) in the UPDB while they did in LINKS. Previous studies showed diverse results regarding a possible survival benefit for spouses of longevous persons[6,7,9,44,45]. In the Long Life Family study, Pedersen et al. (2017) identified a survival benefit for spouses of longevous siblings. The authors compared the spouses to sex and birth cohort matched controls and suggest assortative mating as an explanation for the observed survival benefit of the spouses[6]. A Quebec study, focused on the spouses of 806 centenarians, also reported a survival benefit[44] and a study of Southern Italy demonstrated that male nonagenarians outlived their spouses, whereas this was not the case for female nonagenarians[45]. A recent study showed that the spouses of 944 nonagenarians had no survival benefit but a life-long sustained survival pattern similar to the general population[9]. An explanation for the difference between the UPDB and LINKS datasets may possibly be that Zeeland had a higher level of relatedness than in Utah. Zeeland had poor living conditions[46] and was characterized by out migration to other provinces or abroad, but limited mobility within the province to other places[47]. Utah at that time had better living conditions[48] with continuous streams of freshly incoming migrants, ensuring a steady influx of new genes49, creating high genetic diversity. Hence, it could be that in Zeeland, spouses and IPs were often related to each other and thus shared some of the genetic component contributing to longevity.

Unlike observations we previously made in the Leiden Longevity Study (LLS)[9] concerning maternal effects on longevity in the generation of the nonagenarians and their parents, we did not observe evidence for a stronger transmission from either parent to the IPs (F1 to F2), or from IPs to their children (F2 to F3) in our current study. We cannot draw final conclusions on this aspect because for the F1-2 transmission we may have missed parental influences on early life mortality since IPs were selected for having survived to an age at which they had one child. However, we did capture early life mortality for F2-F3 but in those generations the selection pressure on child mortality was already slightly decreasing[50]. In the same way we observed no differential association between parental longevity and the survival of sons or daughters (F2 and F3). This equal distribution is in line with our observations in the LLS[9], but so far, the literature was less conclusive on this point[15–17].

In all our analyses, except for the spouse analysis, we adjusted for religion (UPDB only), sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, and twin birth. Some of these biological, social, and demographic factors associated with the mortality of IPs (F2) and their children (F3). Nevertheless, these covariates neither confounded the association between parental (F1) and sibling (F2) longevity and IP (F2) survival, nor that between IP (F2) and spouse (F2) longevity and their children's (F3) survival or between longevity of aunts and uncles (F2) and the survival of IPs' children (F3). This is in line with previous studies showing only a minor[51] or no[52] influence of environmental covariates on the association between parental longevity and offspring survival. It was also shown that a range of early life factors, such as farm ownership, parental literacy, and parental occupation did not affect the association between parental and offspring mortality[52]. We, however, cannot completely rule out that other, unobserved non-genetic familial effects may affect our results. Furthermore, using either Swedish or Dutch lifetables to determine survival percentiles was quite strict for Zeeland because of the hazardous environment[46]. As a result, the number of longevous persons was quite low in LINKS relative to

the UPDB. Although the IPs were randomly selected, we could not completely rule out selection effects, for example related to early life mortality. However, confirmation of the F1-F2 results in the next generation F2-F3 significantly strengthens the results and allowed us to cope with the potential selection effects for IPs. In addition, a sensitivity analysis for sample size, in which we fitted all our statistical models on half of the UPDB and half of the LINKS data, provided similar results to the full sample results, indicating that our results are robust for a reduced sample size *(Supplementary Figure 9)*.

Human Lifespan (defined as age at death) has a low heritability in the population at large[1–4]. Studies estimated the heritability of lifespan between 12 and 25%[1–3] and a recent study estimated that the heritability of lifespan was even lower, ~7%, after adjustment for the lifespans of nongenetic (in-law) relatives[4]. Therefore lifespan based gene mapping may not be fruitful. In addition, the genetic component of lifespan includes the heritability of early life mortality, which is mainly due to disease and external causes. Despite the low heritability and polygenic architecture[26,37] of lifespan, recent genetic studies have identified[31,32] and replicated[53] some lifespan loci of which the rare alleles lower the risk of age-related diseases. Hence, using the lifespan trait hampers the identification of genetic loci contributing to survival into extreme ages (longevity). Longevity however, clusters strongly within families as shown by previous studies[5–9] and robustly quantified in this study. Hence, the longevity trait is much more promising and appropriate for the identification of genetic loci contributing to survival into extreme ages and should not be confused with the lifespan trait[1,12]. Our results imply that to find loci that promote survival to the highest ages in the population, genetic studies should be based on long-lived cases including at least parental mortality information but preferably also mortality information of siblings and other first and second degree relatives. The longevity threshold should include cases belonging up to the top 10% survivors, with parents belonging up to the top 15% survivors of their birth cohort and siblings belonging up to the top 10% survivors of their birth cohort. To sharpen the longevity effect, the percentile threshold applied may be made more extreme but would likely lead unnecessarily to a sample size with limited power. If our proposed longevity definition is consistently applied across studies, the comparative nature of longevity studies may improve and facilitate the discovery of novel genetic variants.

Methods

Utah Population Database

The Utah Population Database (UPDB) contains demographic and genealogical information which is linked to medical records. The data construction began in the mid-1970s with genealogy records from the archives at the Utah Family History Library and was initially based on the founding members of the Utah population, their descendants, and then subsequently all individuals living in Utah. These records contain demographic and mortality information on the pioneers of Utah (United States), their parents and children, and have been linked into multigenerational pedigrees. The founding families were selected for the UPDB when at least one member had a vital event (birth, marriage, or death) on the Mormon pioneer trail or in Utah. The UPDB has been expanded to incorporate other high-quality, state-wide data sources, such as birth and death certificates, cancer records, driver license records, and census records. Currently the UPDB contains information on more than 11 million individuals and covers a maximum of 17 generations[54,55].

LINKing System for historical family reconstruction

The LINKing System for historical family reconstruction (LINKS) data contains demographic and genealogical information which was derived from linked vital event registers (birth, marriage, and death certificates). The data indexing began in 1995 by the "Zeeuws" archive and the results were published by way of "WieWasWie". The data currently covers over 25 million Dutch vital event records[56,57]. Data construction has been completed for the province of Zeeland and is still ongoing for the other provinces in the Netherlands. Currently LINKS Zeeland (henceforth referred to as LINKS) contains 739,453 birth, 387,102 marriage, and 641,216 death certificates which were linked together to reconstruct intergenerational pedigrees and individual life courses[47]. In total the Zeeland data contains 1,930,157 persons covering a maximum of 7 generations[58].

Historical context of Utah and Zeeland

Both Utah and Zeeland were high fertility populations[46,48,59], with a mean number of children of around 7 during the period of this study (1740-1952). In general, Utah was marked by healthy living conditions and Zeeland by contrast, was a much unhealthier place to live. One of the main reasons for the unhealthy living conditions in Zeeland was the lack of clean drinking water, the high prevalence of waterborne diseases and of malaria[46,60,61]. In Utah the quality of the drinking water was good, since water from melting snow, that was filtered running of the mountains, was used to drink[48]. The differences in living conditions between Utah and Zeeland were reflected by a relatively low infant and childhood mortality in Utah[62] and high mortality rates for infants and children in Zeeland[61], especially before 1900. Moreover, Utah was known to be a high in-migration population49 whereas there were indications that Zeeland had a low influx and outflux of migrants[47].

Study selection

For the current study we used 3 Filial (F) generations (F1-3) from the UPDB and LINKS ($N_{UPDB+LINKS}$=314,819). We reconstructed families in both datasets and denote generation 1 as the starting point of the pedigrees in the data. The starting point for this study was generation 3 because starting here minimized missing family links and birth or death dates due to the nature of the source material underlying the data. We denote generation 3 as filial generation 1 (F1). Subsequently, the children ($N_{UPDB+LINKS}$=123,599) of the F1 parents were identified (F2) so that unique families were represented by 2 parents

(F1) and their offspring (F2). Next an index person (IP,F2) was randomly selected per F2 sibship ($N_{UPDB+LINKS}$=20,360) meeting the following criteria: (1) The date of birth and death had to be available, (2) At least one child, sibling, and spouse had to be available, (3) sex had to be available, (4) for the UPDB data only, the IP should preferably be identifiable on a genealogy record *(Supplementary Table 12)*. From there we identified the siblings (F2, $N_{UPDB+LINKS}$=108,122), spouses (F2, $N_{UPDB+LINKS}$=22,018), and the children (F3,$N_{UPDB+LINKS}$=123,599) of the IPs *(Table 1 and Figure 1)*. To summarize, both in the UPDB and LINKS we identified IPs (F2), their parents (F1), siblings (F2), spouses (F2), and children (F3).

All individuals in LINKS have at some point in their lives lived in Zeeland, this is because the data were constructed based on vital event records from Zeeland. Utah was first settled in 1847 and in the UPDB mortality information for ancestors of Utah associated persons are available. As a result not all persons necessarily had to live in Utah. *Supplementary Table 13* shows that in our data, 97 percent of the IPs lived in Utah. This percentage is lower for their fathers (80%) and mothers (87%), and is an expected pattern given the historic nature of how Utah was settled. Furthermore, 70% of the siblings, 97% of the spouses, and 92% of the children lived in Utah. The majority of the persons from our sample who lived in Utah, migrated from another state in the US to Utah (87%), 12% came from Europe and 1% from the rest of the world.

Lifetables

We used cohort lifetables to calculate birth cohort and sex specific survival percentiles for each individual in the UPDB and LINKS. This approach prevents against the effects of secular mortality trends over the last centuries and enables comparisons across study populations[1,12]. We could not use United States (US) lifetables because cohort lifetables were not available and period lifetables were only available from 1933 onward. Moreover, the US birth cohort based central death rates were generally incomplete at the earlier cohorts (up to 1900) and proved to be of limited use for our analyses. However, for Sweden and the Netherlands, population based cohort lifetables were available from 1751 and 1850 until 2018 respectively[63–66]. These lifetables contained, for each birth year and sex, an estimate of the hazard of dying between ages x and x + n (hx) based on yearly intervals (n=1) up to 99 years of age. Conditional cumulative hazards (Hx) and survival probabilities (Sx) were derived using these hazards. In turn, we could determine the sex and birth year specific survival percentile for each person in our study. Swedish cohort lifetables date back furthest of all available lifetables and were shown to be consistent with the lifetables of multiple industrialized societies[67]. In addition, we ensured that the survival percentiles were calculated in the same way for the UPDB and LINKS to make a fair comparison between the survival percentiles. Hence, the Swedish cohort lifetables were used for both datasets and for the LINKS data the Dutch lifetables were used as a sensitivity analysis. *Supplementary Figure 10* shows the ages at death corresponding to the top 10, 5, and 1 percent survivors for the UPDB and LINKS. This figure can be used to map the percentiles, which are based on percentile-age pairings from the Swedish lifetables, to absolute ages. For example: a top 10% female in 1750 matched an age of 76 years whereas this was 74 years for males. In 1850 a top 10% female and male matched an age of 83 years and 81 years respectively.

Statistical analyses

Statistical analyses were conducted using R version 3.4.1[68]. We reported 95% confidence intervals (CIs) and considered p-values statistically significant at the 5% level (α = 0.05).

IP survival at increasing survival percentiles of relatives

Analysis 1: To determine if (1) the association between the survival (measured as age at death) of IPs and the survival percentiles of their parents and siblings increased with increasing survival percentiles, and (2) a larger level of family aggregation, in terms of numbers of parents and siblings, was more evident at extreme survival percentiles, we investigated the association between IP survival and the number of parents and siblings reaching increasingly more extreme survival percentiles. We sequentially identified the number of parents and siblings belonging to the top x (x = 1,2,3, ..., 60) percentiles of their birth cohorts (from here: percentiles) and we analyzed their association with the survival of the IPs for each subsequent survival percentile using a Cox proportional hazard model:

$$\lambda(t_{ij}) = \lambda_0(t_{ij})\exp(\boldsymbol{\beta Z}_{ij} + \boldsymbol{\gamma X}_{ij}) \qquad (1)$$

where t_{ij} is the age at death for IP *j* in family *i*. λ_0 (t_{ij}) refers to the baseline hazard, which is left unspecified in a Cox-type model. $\boldsymbol{\beta}$ is the vector of regression coefficients for the main effects of interest ($\boldsymbol{Z}$) which correspond to: (1) the number of parents belonging to the top x percentile, (2) and the number of siblings belonging to the top x percentile. $\boldsymbol{\gamma}$ is a vector of regression coefficients for the effects of covariates and possible confounders ($\boldsymbol{X}$) which are: IPs' religion (UPDB only), sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, and twin birth.

Identifying a survival threshold that demarcates longevity

Analysis 2: The previous analysis, based on the cumulative effects, does not allow us to identify a specific threshold to define longevity, since the top x percentiles were not mutually exclusive, i.e., if a person belonged to the top 1% survivors, this person also belonged to the groups of top 5% and top 10% survivors. To determine the survival percentile threshold that drove the cumulative top x percentile effects described in the previous section, we grouped IPs according to the survival of their parents and siblings for two separate analysis. More specifically, we constructed mutually exclusive groups of IPs based on having at least one parent or sibling belonging to group g (g = 1,2,3, ..., 6): group 1 = [≥0^{th} & ≤1^{th} percentile], group 2 = [≥1^{th} & ≤5^{th} percentile], group 3 = [≥5^{th} & ≤10^{th} percentile], group 4 = [≥10^{th} & ≤15^{th} percentile], group 5 =[≥15^{th} & ≤20^{th} percentile], group 6 = [≥20^{th} & ≤100^{th} percentile]. Group membership was defined by the most long-lived parent or sibling of the IP. Using Cox proportional hazards models (see expression (1)), we compared the effects of all groups to reference group 6, corresponding to IPs with all parents or siblings belonging to the 20th or less extreme survival percentile and multiple combinations of defining group 6 were tested. Here, the $\boldsymbol{\beta}$ is the vector of regression coefficients for the main effects of interest (Z) which correspond to: (1) the IPs who were divided into mutually exclusive groups by their parental mortality and (2) the IPs who were independently grouped by their sibling mortality. Other parts of the expression are the same as noted in expression 1.

Top 10% relatives and covariates in an integrated design

Analysis 3: Based on the analyses expressed in the previous section we chose the top 10% survivors for specific follow-up analyses. Based on the results presented in the cumulative and mutually exclusive group analyses we focused on the top 10% surviving family members because the mutually exclusive group analysis (analysis 2) indicated longevity effects for siblings beyond the top 10% and 15% for siblings and parents respectively. Using the top 10% is consistent between the two groups and is a conservative choice. Furthermore, the cumulative analysis (analysis 1) indicated that the top 10% was

a good trade-off between effect size and group size (power) within and between the UPDB and LINKS. Hence, we focused on top 10% parents and siblings in an integrated design to investigate the association between IP survival and the number of parents and siblings belonging to the top 10%. We subsequently investigated the association between the number of top 10% siblings and IP survival for IPs without top 10% parents, using Cox regression (see expression (1)). Here the $\boldsymbol{\beta}$ is the vector of regression coefficients for the main effects of interest ($\boldsymbol{Z}$) which correspond to: (1) the number of parents and siblings belonging to the top 10% and (2) the number of siblings belonging to the top 10% for IPs without top 10% parents. Other parts of the expression are the same as noted in expression 1.

In all Cox regression analyses, based on expression 1, we accounted for the fact that IPs were selected to have a spouse and at least one child (left truncation) by using an IP specific age at entry in the study based on the IP's age at first child or the age at marriage, whichever was later. A similar approach was followed for the spouses of the IPs and no adjustment for left truncation was necessary for the children of the IPs, since they were not selected in any way. Moreover, we accounted for right censoring in all relatives of the IPs. We furthermore adjusted for religion (UPDB only), sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, and twin birth since these are known to influence human survival[1]. socio-economic status was constructed according to the Integrated Public Use Microdata Series (IPUMS) occupational coding scheme of 1950 (OCC1950)[69]. Importantly, for the sibling contribution to the cumulative percentile analysis (analysis 1), the sibling contribution to the top 10% analyses (analysis 3), and in all mutually exclusive group analyses (analysis 3), we used analytical weights when fitting the Cox models to avoid family size confounding. Adjustment was not necessary for the number of parents because this number is two by definition. However, sibship sizes vary. For example, a hypothetical IP with 4 siblings belonging to percentiles 1, 6, 8 and 30 will contribute with a weight w=3/4 in the first analysis, based on the cumulative percentiles, when considering the top 10 percent. This same IP, when considering the top 5 percent will contribute with less weight, namely w=1/4. In this way, each person contributed the same to the overall analysis across all percentiles. In the second analysis based on mutually exclusive groups, this same hypothetical IP would be assigned to g1, and will contribute to the analysis with a weight w=1/4. In analysis 3, based on the top 10%, the IP will contribute with a weight of w=3/4 In this way we avoid a potential advantage of larger families to be represented in more extreme groups. Finally, we checked the proportional hazards and linearity assumptions in all fitted Cox models. We did not find evidence that model assumptions were violated for the main effects (parent/sibling and IP/children of IPs associations). The proportional hazards assumption was violated for some covariates. In such a case, stratification was applied for that covariate and this was mentioned in the legend of the table/figure.

Verification of the results in a subsequent generation

Analysis 4: To verify our results regarding the top 10% parents and siblings (analysis 3) in a subsequent generation (children, F3), we investigated whether children of top 10% IPs had a survival advantage compared to children of non-longevous IPs and whether this effect is stronger if the spouse of the IP also belonged to the top 10%. We further investigated familial clustering of longevity by studying the number of top 10% aunts and uncles of the children of IPs. A Cox-type random effect model was used:

$$\lambda(t_{ij}) = u_i \lambda_0(t_{ij}) \exp\left(\boldsymbol{\beta Z}_{ij} + \boldsymbol{\gamma X}_{ij}\right) \qquad (2)$$

where t_{ij} is the age at death or the age at last follow-up for child *j* in family *i*, λ_0 (t_{ij}) refers to the baseline hazard, which is left unspecified, $\boldsymbol{\beta}$ is a vector of regression coefficients for the main effects of interest ($\boldsymbol{Z}$) which correspond to: (1) having a parent top 10% survivor in a first analysis and (2) the effect of the number of uncles/aunts (F2) top 10% in a second analysis. u > 0 refers to an unobserved random effect (frailty) shared by F3 children of a given IP. This unobserved heterogeneity shared within sibships was assumed to follow a log-normal distribution. $\boldsymbol{\gamma}$ contains the effect of person-specific covariates $\boldsymbol{X}$, similar to those included in the previous analyses.

Survival of spouses by the longevity of the index persons

Analysis 5 To investigate the survival of spouses, we applied a group approach, similar to that used above, and analyzed the groups with Cox regression. We grouped the spouses by the survival of the IPs creating 6 different groups g (g = 1,2,3, …, 6): group 1 = [≥0th & ≤1th percentile], group 2 = [≥1th & ≤5th percentile], group 3 = [≥5th & ≤10th percentile], group 4 = [≥10th & ≤15th percentile], group 5 =[≥15th & ≤20th percentile], group 6 = [≥20th & ≤100th percentile]. We compared the groups in two steps: (1) group 6 was the reference category and (2) comparing all groups with each other (post-hoc), applying a Bonferroni correction for multiple testing.

$$\lambda(t_{ij}) = \lambda_0(t_{ij})\exp(\boldsymbol{\beta Z}_{ij}) \qquad (3)$$

where t_{ij} is the age at death or the age at last follow-up for spouse *j* in family *i*. $\lambda_0(t_{ij})$ refers to the baseline hazard, which is left unspecified in a Cox-type model. $\boldsymbol{\beta}$ is the regression coefficient referring to the main effects of interest ($\boldsymbol{Z}$), which are the spouses who were divided into mutually exclusive groups by the IPs mortality.

Code availability

The scripts containing the code for data pre-processing and data analyses can be freely downloaded at: https://git.lumc.nl/molepi/PUBLIC/Longevity_top10perc_survivors. This repository describes the main analyses done.

Ethical regulations

We complied with all relevant ethical regulations. For the Utah data (UPDB) the study was approved by the Resource for Genetic and Epidemiologic Research (RGE). For the Zeeland (LINKS) data the study was approved by the International Institute of Social History. For this study, no informed consent needed to be obtained.

Acknowledgments

This research was supported by Netherlands Organization for Scientific Research [360-53-180], the Professor van Winterfonds, and the Leiden University Fund [6223/07-06-2016]. I thank Professor Ken R smith for hosting my stay in Utah (US) to work with the Utah Population Database. I also thank Dr. Joris Deelen and Dr. Thies Gehrmann for providing their input to this work as independent readers.

Author contributions

Niels van den Berg is the study investigator and was responsible for initiating the study, data management, data analyses, writing the first draft of the manuscript and finalizing it, and obtaining funding to visit Ken R. Smith in Utah. Angelique Janssens and P. Eline Slagboom are the study principal investigators who conceived and obtained funding for the project, which this study is a part of. Ken R. Smith is the head of the UPDB and he hosted the stay of Niels van den Berg in Utah, provided access to the UPDB, and supervised all that concerned the UPDB within this study. P. Eline Slagboom and Marian Beekman provided overall project coordination and supervision. Mar Rodriguez-Girondo provided overall statistical analyses coordination and supervision and assisted in the overall project coordination. Kees Mandemakers is the head of LINKS and provided access and support to the LINKS data. Rick Mourits provided access to documentation and conversion tables of Socio Economic Status coding schemes between the two databases. Ingrid van Dijk assisted with the initial family reconstruction process and assisted in working with the LINKS data.

Competing interests

The authors declare no competing interests.

Data availability

The UPDB and LINKS data that support the findings of this study are available from the UPDB and the IISG but restrictions apply to the availability of these data, which were used under license for the current study, and so are not publicly available. Data are however available upon reasonable request and with approvals of the UPDB and the IISG. The LINKS data is available upon request to Dr. Kees Mandemakers (kma@iisg.nl), International Institute of Social History, Cruquiusweg 31, 1019 AT Amsterdam, the Netherlands. The UPDB is also available upon request and approval by the Resource for Genetic and Epidemiologic Research (RGE). More information on making this request can be obtained from one of the authors, Dr. Ken R Smith (ken.smith@fcs.utah.edu), University of Utah, 225 S. 1400 E. Rm 228 Salt Lake City, United States.

References

1 van den Berg, N., Beekman, M., Smith, K. R., Janssens, A. & Slagboom, P. E. Historical demography and longevity genetics: Back to the future. *Ageing Res. Rev.* **38**, 28–39 (2017).

2 Herskind, A. M. *et al.* The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Hum. Genet.* **97**, 319–323 (1996).

3 Kaplanis, J. *et al.* Quantitative analysis of population-scale family trees with millions of relatives. *Science* **360**, 171–175 (2018).

4 Ruby, J. G. *et al.* Estimates of the Heritability of Human Longevity Are Substantially Inflated due to Assortative Mating. *Genetics* **210**, 1109–1124 (2018).

5 Perls, T. T. *et al.* Life-long sustained mortality advantage of siblings of centenarians. *Proc. Natl. Acad. Sci.* **99**, 8442–8447 (2002).

6 Pedersen, J. K. *et al.* The Survival of Spouses Marrying Into Longevity-Enriched Families. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **72**, 109–114 (2017).

7 Schoenmaker, M. *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur. J. Hum. Genet.* **14**, 79–84 (2006).

8 Ljungquist, B., Berg, S., Lanke, J., McClearn, G. E. & Pedersen, N. L. The Effect of Genetic Factors for Longevity: A Comparison of Identical and Fraternal Twins in the Swedish Twin Registry. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **53A**, 441–446 (1998).

9 Berg, N. van den *et al.* Longevity Around the Turn of the 20th Century: Life-Long Sustained Survival Advantage for Parents of Today's Nonagenarians. *Journals Gerontol. Ser. A* **73**, 1295–1302 (2018).

10 Willcox, B. J., Willcox, D. C., He, Q., Curb, J. D. & Suzuki, M. Siblings of Okinawan centenarians share lifelong mortality advantages. *J. Gerontol. A. Biol. Sci. Med. Sci.* **61**, 345–54 (2006).

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

12 Sebastiani, P., Nussbaum, L., Andersen, S. L., Black, M. J. & Perls, T. T. Increasing Sibling Relative Risk of Survival to Older and Older Ages and the Importance of Precise Definitions of "Aging," "Life Span," and "Longevity". *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **71**, 340–346 (2016).

13 Kerber, R. A., Brien, E. O., Smith, K. R. & Cawthon, R. M. Familial Excess Longevity in Utah Genealogies. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **56**, 130–139 (2001).

14 Gudmundsson, H., Gudbjartsson, D. F. & Kong, A. Inheritance of human longevity in Iceland. *Eur. J. Hum. Genet. EJHG* **8**, 743–749 (2000).

15 Houde, L., Tremblay, M. & Vézina, H. Intergenerational and Genealogical Approaches for the Study of Longevity in the Saguenay-Lac-St-Jean Population. *Hum. Nat.* **19**, 70–86 (2008).

16 Kemkes-Grottenthaler, A. Parental effects on offspring longevity—evidence from 17th to 19th century reproductive histories. *Ann. Hum. Biol.* **31**, 139–158 (2004).

17 Deluty, J. A., Atzmon, G., Crandall, J., Barzilai, N. & Milman, S. The influence of gender on inheritance of exceptional longevity. *Aging (Albany. NY).* **7**, 412–418 (2015).

18 Andersen, S. L., Sebastiani, P., Dworkis, D. a., Feldman, L. & Perls, T. T. Health Span Approximates Life Span Among Many Supercentenarians: Compression of Morbidity at the Approximate Limit of Life Span. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **67A**, 395–405 (2012).

19 Ash, A. S. *et al.* Are Members of Long-Lived Families Healthier Than Their Equally Long-Lived Peers? Evidence From the Long Life Family Study. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **70**, 971–976 (2015).

20 Christensen, K., McGue, M., Petersen, I., Jeune, B. & Vaupel, J. W. Exceptional longevity does not result in excessive levels of disability. *Proc. Natl. Acad. Sci. U. S. A.* **105**, 13274–13279 (2008).

21 Evert, J., Lawler, E., Bogan, H. & Perls, T. Morbidity Profiles of Centenarians: Survivors, Delayers, and Escapers. *Journals Gerontol. - Med. Sci.* **58**, 232–237 (2003).

22 Vaarhorst, A. A. M. *et al.* Lipid metabolism in long-lived families: the Leiden Longevity Study. *Age (Omaha).* **33**, 219–227 (2011).

23 Wijsman, C. A. *et al.* Familial longevity is marked by enhanced insulin sensitivity. *Aging Cell* **10**, 114–121 (2011).

24 Rozing, M. P. *et al.* Low serum free triiodothyronine levels mark familial longevity: The leiden longevity study. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **65A**, 365–368 (2010).

25 Eline Slagboom, P., van den Berg, N. & Deelen, J. Phenome and genome based studies into human ageing and longevity: An overview. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1864**, 2742–2751 (2018).

26 Shadyab, A. H. & LaCroix, A. Z. Genetic factors associated with longevity: A review of recent findings. *Ageing Res. Rev.* **19**, 1–7 (2015).

27 Deelen, J. *et al.* Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* **23**, 4420–4432 (2014).

28 Broer, L. *et al.* GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **70**, 110–118 (2015).

29 Willcox, B. J. *et al.* FOXO3A genotype is strongly associated with human longevity. *Proc. Natl. Acad. Sci.* **105**, 13987–13992 (2008).

30 Flachsbart, F. et al. Association of FOXO3A variation with human longevity confirmed in German centenarians. Proc. Natl. Acad. Sci. 106, 2700–2705 (2009).

31 Pilling, L. C. *et al.* Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany. NY).* **9**, 2504–2520 (2017).

32 Joshi, P. K. *et al.* Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nat. Commun.* **8**, 1–13 (2017).

33 Sebastiani, P. *et al.* Four Genome-Wide Association Studies Identify New Extreme Longevity Variants. *J. Gerontol. A. Biol. Sci. Med. Sci.* **72**, 1453–1464 (2017).

34 Sebastiani, P. *et al.* A Family Longevity Selection Score: Ranking Sibships by Their Longevity, Size, and Availability for Study. *Am. J. Epidemiol.* **170**, 1555–1562 (2009).

35 Dutta, A. *et al.* Longer lived parents: Protective associations with cancer incidence and overall mortality. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **68**, 1409–1418 (2013).

36 Terry, D. F. *et al.* Lower All-Cause, Cardiovascular, and Cancer Mortality in Centenarians' Offspring. *J. Am. Geriatr. Soc.* **52**, 2074–2076 (2004).

37 Christensen, K., Johnson, T. E. & Vaupel, J. W. The quest for genetic determinants of human longevity: challenges and insights. *Nat. Rev. Genet.* **7**, 436–48 (2006).

38 Boyle, E. A., Li, Y. I. & Pritchard, J. K. An Expanded View of Complex Traits: From Polygenic to Omnigenic. *Cell* **169**, 1177–1186 (2017).

39 Joshi, P. K. *et al.* Variants near CHRNA3/5 and APOE have age- and sex-related effects on human lifespan. *Nat. Commun.* **7**, 11174 (2016).

40 Gögele, M. *et al.* Heritability analysis of life span in a semi-isolated population followed across four centuries reveals the presence of pleiotropy between life span and reproduction. *J. Gerontol. A. Biol. Sci. Med. Sci.* **66**, 26–37 (2011).

41 Gavrilova, N. *et al.* Evolution, mutations, and human longevity European royal and noble families. *Hum. Biol.* **70**, 799–804 (1998).

42 Hjelmborg, J. vB *et al.* Genetic influence on human lifespan and longevity. *Hum. Genet.* **119**, 312–321 (2006).

43 Gavrilova N.S., G. L. A., Gavrilov, L. A. & Gavrilova, N. S. When does human longevity start?: Demarcation of the boundaries for human longevity. *Rejuvenation Res.* **4**, 115–124 (2001).

44 Jarry, V., Gagnon, A. & Bourbeau, R. Survival advantage of siblings and spouses of centenarians in 20th-century Quebec. *Can. Stud. Popul.* **39**, 67 (2013).

45 Montesanto, A. *et al.* The genetic component of human longevity: analysis of the survival advantage of parents and siblings of Italian nonagenarians. *Eur. J. Hum. Genet.* **19**, 882–886 (2011).

46 Hoogerhuis, O. W. *Baren op Beveland. Vruchtbaarheid en zuigelingensterfte in Goes en omliggende dorpen gedurende de 19e eeuw.* (Wageningen Afd. Agrarische Geschiedenis Wageningen Universiteit 2003, 2003).

47 van den Berg, N. *et al.* Families in Comparison: An individual-level comparison of life course and family reconstructions between population and vital event registers. *SocArXiv* (2018). at <https://osf.io/preprints/socarxiv/h2w8t/>

48 Bean, L. L., Mineau, G. P. & Anderton, D. L. *Fertility Change on the American Frontier: Adaptation and Innovation.* **4**, (Univ of California Press, 1990).

49 Toney, M. B., Stinner, C. M. & Kan, S. Mormon and Nonmormon Migration in and out of Utah. *Rev. Relig. Res.* **25**, 114 (1983).

50 Fogel, R. W. & Costa, D. L. A Theory of Technophysio Evolution, With Some Implications for Forecasting Population, Health Care Costs, and Pension Costs. *Demography* **34**, 49 (1997).

51 Danzhen You, Danan Gu & Zeng Yi. Familial Transmission of Human Longevity Among the Oldest-Old in China. *J. Appl. Gerontol.* **29**, 308–332 (2010).

52. Gavrilov, L. A. & Gavrilova, N. S. Predictors of Exceptional Longevity: Effects of Early-Life and Midlife Conditions, and Familial Longevity. *North Am. Actuar. J.* **19**, 174–186 (2015).

53 Timmers, P. R. H. J. *et al.* Genomic underpinnings of lifespan allow prediction and reveal basis in modern risks. *bioRxiv* 1–71 (2018). at <http://biorxiv.org/content/early/2018/07/06/363036.abstract>

54 https://healthcare.utah.edu/huntsmancancerinstitute/research/updb/data/family-records.php.

55 Bean, L. L., May, D. L. & Skolnick, M. The Mormon Historical Demography Project. *Hist. Methods A J. Quant. Interdiscip. Hist.* **11**, 45–53 (1978).

56 https://socialhistory.org/en/hsn/linking-system-historical-family-reconstruction-links.

57 Mandemakers, K. *LINKing System for historical family reconstruction.* (2007).

58 Mandemakers, K. & Laan, F. LINKS dataset Genes Germs and Resources. WieWasWie Zeeland. Civil Certificates. (2017).

59 Heaton, T. B. How Does Religion Influence Fertility?: The Case of Mormons. *J. Sci. Study Relig.* **25**, 248 (1986).

60 van Poppel, F., Jonker, M. & Mandemakers, K. Differential infant and child mortality in three Dutch regions, 1812–1909. *Econ. Hist. Rev.* **58**, 272–309 (2005).

61 van dijk, I. K. & Mandemakers, K. Like mother, like daughter. Intergenerational transmission of infant mortality clustering in Zeeland, the Netherlands, 1833-1912. *Hist. Life Course Stud.* **1**–26 (2018).

62 Lynch, K. A., Mineau, G. P. & Anderton, D. L. Estimates of Infant Mortality on the Western Frontier: The Use of Genealogical Data. *Hist. Methods A J. Quant. Interdiscip. Hist.* **18**, 155–164 (1985).

63 Wilmoth, J. R. & Shkolnikov, V. Human Mortality Database. *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)* (2017).

64 Van Der Meulen, *A. Life tables and Survival analysis.* (2012). at <https://www.cbs.nl/en-gb/our-services/methods/statistical-methods/output/output/life-tables>

65 Carolina, T., Uijvenhoven, L. & van der Laan, J. Overlevingstafels en longitudinale analyse. *CBS* 1–25 (2009). at <https://www.cbs.nl/nl-nl/onze-diensten/methoden/statistische-methoden/output/output/overlevingstafels-en-longitudinale-analyse-survival-analyse-duurmodellen>

66 Lundstrom, H. *Cohort mortality in Sweden: Mortality statistics since 1861.* (2010). at <https://www.scb.se/statistik/_publikationer/be0701_2010a01_br_be51br1001eng.pdf>

67 Lindahl-Jacobsen, R. *et al.* The male–female health-survival paradox and sex differences in cohort life expectancy in Utah, Denmark, and Sweden 1850–1910. *Ann. Epidemiol.* **23**, 161–166 (2013).

68 R Core Team. R: A language and environment for statistical computing. (2016).

69 Ruggles, S., Genadek, K., Goeken, R., Grover, J. & Sobek, M. *Integrated Public Use Microdata. User's guide* **6**, (2015).

CHAPTER 5

LONGEVITY AROUND THE TURN OF THE 20TH CENTURY:

LIFE-LONG SUSTAINED SURVIVAL ADVANTAGE FOR PARENTS OF TODAY'S NONAGENARIANS

Niels van den Berg[1,2]
Mar Rodríguez-Girondo[3]
Anton J.M. de Craen[4]
Jeanine J Houwing-Duistermaat[3]
Marian Beekman[1]
P. Eline Slagboom[1,5]

[1] Department of Biomedical Data Sciences, section of Molecular Epidemiology,
Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden the Netherlands
[2] Radboud Group for Historical Demography and Family History,
Radboud University, Erasmusplein 1, 6525 HT Nijmegen, the Netherlands
[3] Department of Biomedical Data Sciences, section of Medical Statistics,
Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden the Netherlands
[4] Department of Gerontology and Geriatrics,
Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden the Netherlands
[5] Max Planck Institute for Biology of Ageing,
Joseph-Stelzmann-Str. 9b, D-50931 Cologne, Germany

Journals of Gerontology series a (October 2018)
https://doi.org/10.1093/gerona/gly049

Abstract

Members of longevous families live longer than individuals from similar birth cohorts and delay/escape age related diseases. Insight into this familial component of longevity can provide important knowledge about mechanisms protecting against age-related diseases. This familial component of longevity was studied in the Leiden Longevity Study which consists of 944 longevous siblings (participants), their parents (N=842), siblings (N=2302), and spouses (N=809). Family longevity scores were estimated to explore whether human longevity is transmitted preferentially through the maternal or paternal line. Standardized mortality ratio's (SMRs) were estimated to investigate whether longevous siblings have a survival advantage compared to longevous singletons and we investigated if parents of longevous siblings harbor a life-long sustained survival advantage compared to the general Dutch population by estimating lifetime SMRs (L-SMRs). We found that sibships with long-lived mothers and non-long-lived fathers had 0.41 (P=0.024) less observed deaths than sibships with long-lived fathers and non-long-lived mothers and 0.48 (P=0.008) less observed deaths than sibships with both parents non-long lived. Participants had 18.6% less deaths compared to matched singletons and parents had a life-long sustained survival advantage (L-SMR=0.510 and 0.688). In conclusion, genetic longevity studies may incorporate the maternal transmission pattern and genes influencing the entire life-course of individuals.

Introduction

The average human life expectancy steadily increased over the last 200 years in industrialized countries, with record life expectancy increasing from 43/45 years in 1840 to 79/85 years in 2015 for males and females respectively[1]. Until 1950 the average increase in life expectancy could mainly be attributed to improved living conditions and better healthcare, causing a decrease in childhood and early life mortality[2]. After 1950 the average life expectancy increased due to a delay of mid and late-life mortality[3–6]. Despite the average increase in life expectancy in the industrialized countries, significant individual differences in lifespan, defined as age at death, exist[7,8]. In fact, a small group of individuals is able to survive into exceptionally old ages. This longevity capacity clusters within families[9–11] and on top of that, members of such long-lived families seem to delay or even escape age-related disease[12–15]. Hence, research into long-lived families plays a key role in gaining knowledge about how to prevent age related disease.

Previous research has focused on the survival of first degree relatives and spouses of long-lived persons. Siblings of centenarians and siblings of nonagenarian descendants had a life-long sustained survival advantage compared to sex and birth cohort matched controls[9,10,16]. In addition, siblings, parents, and offspring of nonagenarian siblings lived significantly longer than members of comparable birth cohorts[11]. Multigenerational studies into the sex-specific inheritance pattern of lifespan and longevity showed inconsistent results however, with either paternal or maternal transmission patterns ([17–33], as reviewed in [34]). Despite the generally observed survival advantage of first degree relatives of longevous subjects, observations on the survival of their spouses and on longevity inheritance patterns remain inconclusive[11,35,36].

The limitations in current inheritance pattern studies are twofold. First, secular trends, such as the increase of life expectancy over time, are not taken into account. Second, parent-offspring analysis usually focuses on a single child per family, thereby omitting the potential of a complete sibship per family[37]. Furthermore, studies have selected long-lived persons based on different criteria, focusing either on multiple siblings or singletons[9–11,16]. It remains to be elucidated whether the stringency of long-lived case selection based on the presence or absence of a long-lived sibling provides a survival advantage in the selected persons compared to birth cohort and sex matched long-lived singletons. Apart from this, research into the survival of first degree relatives and spouses of long-lived persons often struggles to obtain an accurate population based control group, sometimes leading to the generalization of a single birth year control group to other birth years[16]. It is also difficult to compare the survival of parents of long-lived persons to population based sex and birth cohort matched controls because representative cohort lifetables preceding 1900 are often unavailable, except for the Netherlands and Sweden[38]. Overall, research is still inconclusive about the following issues: sex-specific inheritance pattern of longevity, the survival advantage of long-lived sibships as compared to long-lived singletons and about the question whether their parents already had a life-long sustained survival advantage.

To investigate these three issues, we used the data available in the Leiden Longevity Study (LLS). The LLS currently contains 421 complete 3 generational families, which we denote with filial 0 until 2 (F0 – F2). First, we grouped complete F1 sibships to their parental longevity. We defined parental longevity as belonging to the top 1% of their birth cohort[34,39] and constructed four parental groups: Group 1: both parents were long-lived (n=1); group 2: mother long-lived and father not

long-lived (n=17); group 3: father long-lived and mother not long-lived (n=21); group 4: both parents were not long-lived (n=371). We subsequently compared the longevity Family Scores (LFS) of the different groups. Next, we investigated whether longevous siblings had a survival advantage over sex and birth cohort matched singletons using standardized mortality ratios (SMR). We compared the survival of spouses of longevous siblings to sex and birth cohort matched controls. Finally, we estimated lifetime standardized mortality ratios (L-SMRs) to determine if parents of longevous siblings had a life-long sustained survival advantage.

Methods

Leiden Longevity Study

The LLS was initiated in 2002 to study genetic determinants of human longevity. The LLS consists of 421 families and covers 2 generations of living subjects (F1 and F2) who were born between 1864 and 2017. Inclusion took place from 2002 until 2006. Men and women could participate if they were alive and aged >=89 and >= 91 respectively. Both men and women were recruited to have a living sibling meeting the same criteria. Furthermore, the parents of the F1 participants had to be of Dutch Caucasian origin, and the siblings in one family had to descend from the same parents. The sex specific age inclusion criteria represented individuals equal to, or beyond the oldest 0.5 percent of the Dutch population in 2001. There were no selection criteria on health or demographic characteristics. In total 944 longevous F1 participants, who provided blood for research purposes, were included in the LLS (F1). In addition, their offspring and the spouses of their offspring were included (F2).

Relevant for the current study is that genealogical information was collected for the siblings (F1; N=2302), parents (F0; N=842) and spouses (F1; N=809) of the longevous F1 participants (henceforth referred to as siblings, parents, spouses, and participants). All genealogical information was verified by birth or marriage certificates and passports whenever possible. Additionally, verification took place via personal cards which were obtained from the Dutch Central Bureau of Genealogy in the Hague. In 2017 we updated the ages at death and last observation via the currently centralized municipal personal records database. For this study we used two generations (F0 and F1) consisting of 4807 individuals in all 421 families *(Figure 1 and Table 1)* because 86% from the third generation (F2) were still alive.

Lifetables

In the Netherlands, population based cohort lifetables are available from 1850 until 2017[40,41]. These lifetables contain, for each birth year and sex, an estimate of the hazard of dying between ages x and x + n (h_x) based on yearly intervals (n=1) up to 99 years of age. Conditional cumulative hazards (H_x) and survival probabilities (S_x) can be derived using these hazards. In turn, we can determine to which sex and birth year based survival percentile each person of our study belonged to. For example: person "A" was born in 1876, was a female, and died at age 92. According to the lifetable information this person belonged to the top three percent survivors of her birth cohort, meaning that only three percent of the women born in 1876 reached a higher age than person A. We used the lifetables to calculate the birth cohort and sex specific survival percentiles for each individual in the LLS. *Figure A1* shows the ages at death corresponding to the top 10, 5, and 1 percent survivors of their birth cohorts for the period 1850-1960.

Statistical analyses

Statistical analyses were conducted using R statistics version 3.3.0[42].

Standardized mortality Ratio's

To indicate excess mortality or excess survival of groups in the LLS compared to a reference population we used Standardized Mortality Ratios (SMRs). An SMR is estimated by dividing the observed number of deaths by the expected number of

deaths. The expected number of deaths are given by the sum of all individual cumulative hazards based on the birth cohort and sex specific lifetables of the Dutch population. An SMR between 1 and 0 indicates excess survival, an SMR of 1 indicates that the study population shows a similar survival to the reference population, and an SMR above 1 indicates excess mortality. The SMR can be estimated conditional on the specific age at which an individual starts to be observed in the study. This was necessary to avoid selection bias if individuals in a study population were not at risk of dying before a specific age of entry.

$$SMR = \frac{observed\ number\ of\ deaths}{expected\ number\ of\ deaths} = \frac{\sum_{i=1}^{N} d_i}{\sum_{i=1}^{N} H_{t0i}(t_i|t_{0i})}$$

d_i=dead status (1=dead, 0=alive), H_{t0i}=sex and birth year specific cumulative hazard based on lifetable, t_i=timing, referring to age at death or last observation, t_{0i}=liftable age conditioning, in this case from birth (t_{0i}=0), N= group sample size

SMRs were estimated for all first degree relatives (F0 and F1) of the LLS participants (F1) to investigate their survival compared to the Dutch population. Direct or indirect selection effects were taken into account when estimating the SMR by conditioning the lifetable hazards to the age at first death of a specific group. SMRs were also estimated for participants by conditioning to age of inclusion, which varies between 89 and 102 years (see *Table A1* for an overview of conditioning criteria). Note that the lifetables do not contain yearly interval information beyond the age of 99. For this reason the SMR estimations were truncated at 99 years.

To estimate the SMR at every possible starting age we restricted age at death or last observation at yearly thresholds between 0 – 99 years for every group in the LLS, except for the participants because they were selected to have survived >=89/91 years (men/women). We will refer to these age conditioned SMRs as L-SMRs. These L-SMRs provided insight into the specific moment the first degree relatives and spouses had a survival advantage during their lifespan. SMR and L-SMR confidence intervals were estimated using 95% family based bootstrap confidence intervals with 500 resampling cycles to correct for familial dependencies in the LLS data.

Longevity Family Score

To summarize the survival of a specific study population or subsample on the level of families we constructed a Longevity Family Score (LFS). The LFS is related to the SMR, but it is estimated by subtracting the sex, birth cohort, and age conditioned specific cumulative hazards by event status (1 if death and 0 if alive) for each individual in the study population. In a next step, the family mean is calculated which adjusts for family size and results in the LFS. The LFS is related to the Family Mortality History Score described by Rozing et al.[43] and the est(SE) described by Sebastiani et al.[44]. The LFS ranges between -1 and infinity. A score of 0 indicates that the familial longevity resembles that of the normal Dutch population. A score above 0 indicates excess survival and below 0 indicates excess mortality. For example: family "A" scores an LFS of 1. This indicates that we observe 1 death less than expected based on the Dutch population.

$$LFS_i = \frac{expected\ deaths - observed\ deaths}{sibship\ size} = \frac{\sum_{j=1}^{N_i}(H_{t0ij}(t_{ij}|t_{0ij}) - d_{ij})}{N_i}$$

d_{ij}=dead status (1=dead, 0=alive) of individual j, H_{t0ij}=sex and birth year specific cumulative hazard based on lifetable, t^{ij}=timing, referring to age at death or last observation, t_{0ij}=liftable age conditioning, in this case from birth (t_{0ij}=0), N_i=sibship size

To identify the presence of a sex specific inheritance pattern, four groups of F1 sibships (participants+siblings) were constructed according to their parental longevity. We defined parental longevity as belonging to the top 1% of their birth cohort. Group 1: both parents were long-lived (n=1); group 2: mother long-lived and father not long-lived (n=17); group 3: father long-lived and mother not long-lived (n=21); group 4: both parents were not long-lived (n=371). Group 1 was omitted from the analyses because the size was too small and 12 sibships could not be grouped due to missing ages at death of their parents. The LFS was used to summarize F1 sibship survival relative to the parental groups. F1 LFS differences between the groups were tested using the non-parametric Mann-Whitney U test and corresponding 95% exact confidence intervals were reported[45].

Results

To investigate sex specific inheritance and the presence of a life-long sustained survival advantage in the LLS we used two generations covering longevous participants (F1; N= 944), their parents (F0; N= 842), siblings (F1; N= 2302), and spouses (F1; N=809) *(Figure 1).*

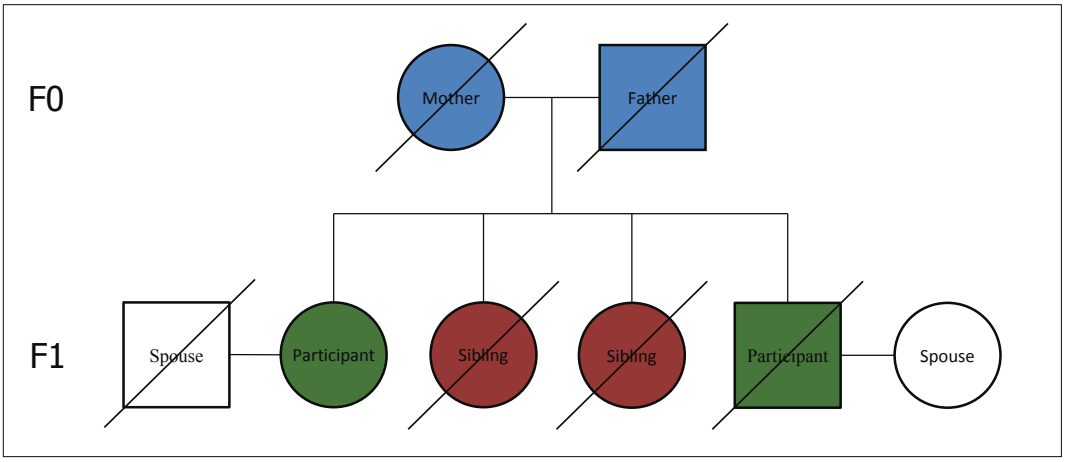


Figure 1: Pedigree map of an example LLS family illustrating the LLS study design.

Circles represent women and squares represent men. Diagonal lines indicate that an individual is deceased. This figure indicates that some participants and their spouses are still alive as of the data of submission. *Table 1* provides an elaborate overview of the LLS data. Colors indicate as follows: BLUE: parental generation (F0); GREEN: participants (F1); RED: siblings (F1); TRANSPARENT: spouses (F1).

The participants were born between 1900 and 1916, and 63% were female (n=595). The participants' mean age at death or at last observation was 97 years and 22 (2%) participants are currently alive. The parents were born between 1850 and 1894 and they are all passed away with a mean age at death of 77 years. We were unable to retrieve the age at death of 22 parents (3%). The siblings were born between 1875 and 1941 and 47% were female (n=1082). The siblings mean age at death was 69 years and the median age at death was 80 years. 365 (16%) siblings are currently still alive while we were unable to retrieve any information on the age at death for 33 (2%) siblings. The mean sibship size for F1 (participants+siblings) was 7.71 (SD=3.4) with a minimum of 2 and a maximum of 17 siblings. The spouses were born between 1882 and 1950. 40% of the spouses were female (n=324) and their mean age at death was 75 years. 27 (3%) spouses are currently alive and for 119 (15%) spouses no age at death or last observation was available *(Table 1).*

Table 1: Leiden Longevity Study sample for participants and first degree relatives

	Parents F0	Participants* F1	Siblings** F1	Spouses F1
Number, N	842	944	2302	809
Deceased, N (%)	820 (97)	922 (98)	1904 (83)	663 (82)
Alive, N (%)	0 (0)	22 (2)	365 (16)	27 (3)
Female, N (%)	421 (50)	595 (63)	1082 (47)	324 (40)
Range birth cohorts	1850-1894	1900-1916	1875-1941	1882-1950
Mean age, years (SD)	77 (14.2)	97 (3.6)	69 (28.3)	75 (14.5)
Median age, years (MAD)	80 (13.3)	97 (4.0)	80 (12.8)	78 (11.0)
Missing age, N (%)	22 (3)	0 (0)	33 (2)	119 (15)

*Participants are enrolled as siblings meeting the age criteria of 89 (men) or 91 years (women). *Siblings are the siblings of participants who did not meet the age criteria yet or who had already been deceased at the time of enrolment. Age refers to either age at death or age at last observation. Missing age means that we have no observation at all. -SD = standard deviation, MAD = median absolute deviation.

LLS data is of high quality

We verified the observations as described by Schoenmaker et al. (2006) based on the first 100 LLS families by estimating SMRs for parents, spouses, and siblings of the complete enrolled LLS *(Table 2)*.

Table 2: Sex specific standardized mortality ratios for 1th degree relatives and spouses of LLS participants

	Sample size	Observed deaths	Expected deaths	SMR (95% CI)
Generation 0 (F0)				
Parents of participants	842	820	1190	0.688 (0.651 – 0.727)
Generation 1 (F1)				
Siblings of participants	2302	1867	2816	0.663 (0.634 – 0.695)
Spouses of participants	809	663	648	1.022 (0.966 – 1.093)

Confidence intervals have been estimated using bootstrapping with 500 cycles. The Dutch life tables do not contain yearly interval information beyond the age of 99. For this reason the SMR calculations have been truncated at 99 years in order to correctly estimate group specific SMR's. No significant differences between men and women have been observed for any category. Observed deaths have been counted after the age of the first death in a group for "parents of participants", "siblings of participants", and "spouses of participants". For the participants observed deaths have been counted after the age of inclusion for each individual separately. This is to correct for selection effects in the data. In line with the counting of the observed deaths, the Dutch lifetables have been age conditioned to match the counting of deaths in the different groups. Equal to the counting of observed deaths, the age conditioning of the lifetables was done to correct for selection effects.

We estimated an SMR of 0.688 (95% CI=0.651–0.727) for parents, indicating that we observed 31.2 percent less deaths than would have been expected based on single individuals from a similar birth cohort and sex. The SMR for siblings was 0.662 (95% CI=0.634–0.695), indicating that we observed 33.8 percent less deaths than would have been expected based on single individuals from a similar birth cohort and sex. Spouses had an estimated SMR of 1.022 (95% CI=0.966–1.093). This indicates that we have not found differences between the survival of spouses and single individuals from similar birth cohorts and sex.

Maternal transmission of longevity

To determine an inheritance pattern based on information of not just single individuals but an entire sibship we used a Longevity Family Score (LFS) to summarize sibship survival. We grouped sibships (F1, participants+siblings) according to their parental (F0) longevity (parental longevity was defined as belonging to the top 1% survivors of their birth cohort) and compared the median group LFS of the complete sibships. *Figure 2* shows that all F1 sibship groups, on average, had an excess survival as compared to single individuals from the same birth cohorts and sex, as indicated by the median scores which were all above 0. Sibships with a long-lived (LL) father and a non-long-lived (NL) mother had 1.21 (median LFS) less observed deaths in reference to the Dutch population and a mean sibship size of 8.34 (SD=3.4). Sibships with an LL mother and an NL father had 1.62 (median LFS) less observed deaths with a mean sibship size of 5 (SD=1.9) and sibships with both parents NL had 1.1 less observed deaths with a mean sibship size of 7.95 (SD=3.4). As a result, sibships with long-lived mothers and non-long-lived fathers showed larger LFSs than sibships with long-lived fathers and non-long-lived mothers (median difference in LFS of 0.41; 95% CI=0.07–0.77 ; P=0.024) Similarly, they showed larger LFSs than sibships with both parents non-long lived (median difference in LFS=0.48; 95% CI=0.15–0.79 ; P=0.008). We did not observe differential survival between sons and daughters with a long-lived mother *(Figure A2).* In conclusion, we observed a maternal transmission pattern of human longevity with no evidence of a differential survival advantage for sons and daughters.

Last life-phase survival advantage of siblings over singletons

To test if longevous F1 participants had a survival advantage over birth cohort, sex and inclusion age matched singletons we estimated sex-specific SMRs for the participants *(Figure 3A).* An SMR of 0.814 (95% CI=0.757–0.884) was estimated for the participants, indicating that as a group the participants had 18.6% less deaths than expected based on single individuals from similar birth cohorts and sex. Female participants had a slightly larger survival advantage (0.804 (95% CI=0.738–0.894)) than male participants (0.828 (95% CI=0.742–0.943)) although this difference was not significant.

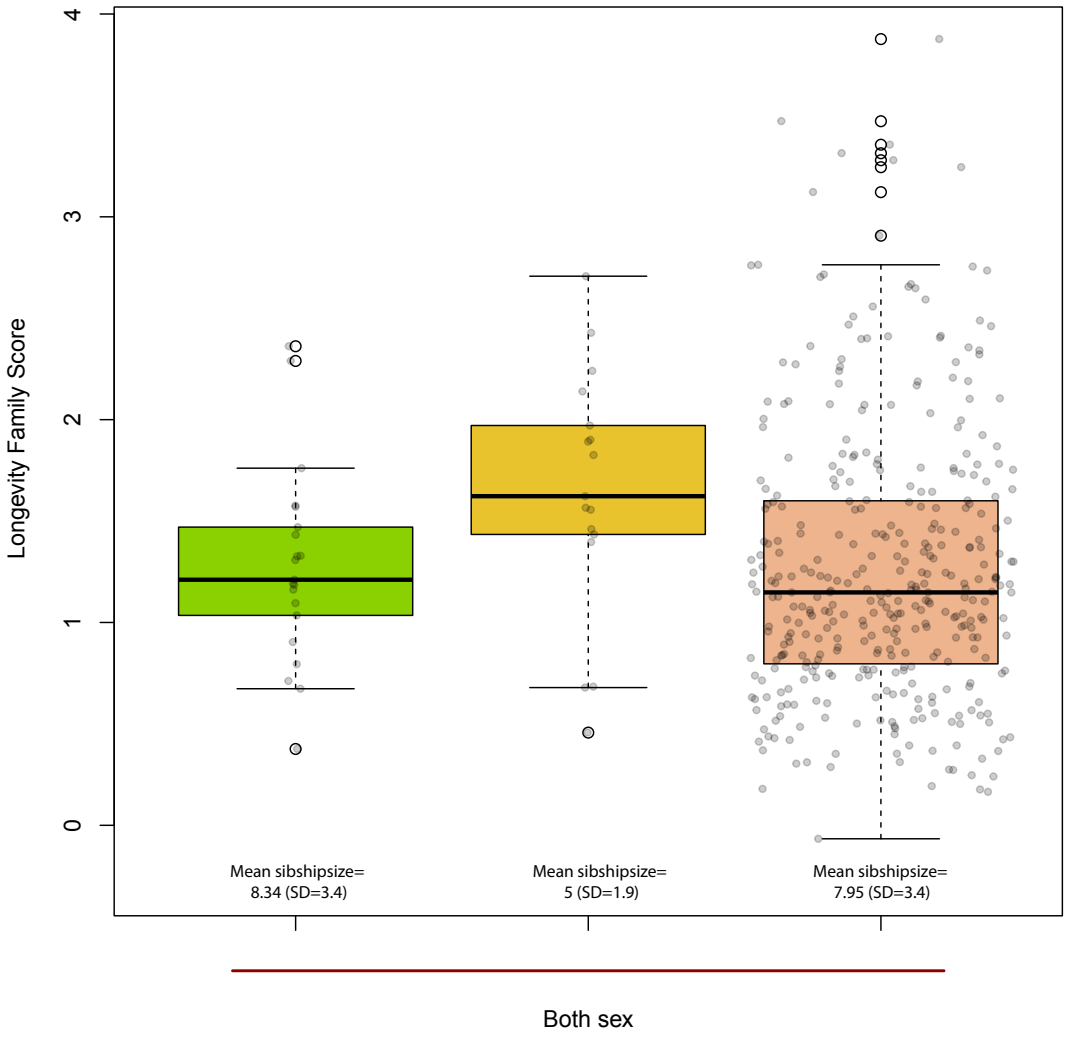


Figure 2: Median Longevity family score per sibship with one or none long-lived parent.

Each gray dot represents a complete sibship. Green boxplot represents the group of sibships with long-lived father and a non-long-lived mother (N_sibships=21 ; N_individuals=177). Orange boxplot represents the group of sibships with a long-lived mother and a non-long-lived father (N_sibships=17 ; N=individuals=85). Light brown boxplot represents the group of sibships with both parents not long-lived (N_sibships=371 ; N_individuals=2949).

Life-long sustained survival advantage of siblings and parents but not for spouses

Whether first degree relatives and spouses of the participants had a survival advantage over their entire lifetime was studied by estimating L-SMRs. *Figure 3B* shows that siblings had a significant survival advantage compared to individuals from similar birth cohorts and sex at any point of their lifetime distribution until the threshold of 97 years, although the SMR at 98 years was again significant. The mean L-SMR was 0.680 and the median L-SMR was 0.660. No sex differences were identified at any age threshold. We observed that spouses had a non-significant L-SMR until age 74, indicating that they were similar to sex and birth cohort matched individuals from the general population. Beyond age 74 there was a small but significant survival disadvantage (min SMR=1.09 and max SMR=1.32) and from age 91 until 94 the effects were not statistically significant anymore. Among spouses, no statistically significant differences between husbands and wives could be detected at any age threshold. The mean L-SMR was 1.050 and the median L-SMR over all age points was 1.030 *(Figure 3C)*. Finally, we were able to study the life-long survival for parents of longevous participants *(Figure 3D)*. Parents had a significant survival advantage compared to individuals from the same birth cohort and sex at any point of the parents' lifetime distribution until 93 years. After 93 years the SMR estimates were still below 1 although not statistically significant, probably due to small sample size. The parental mean and median L-SMR were 0.510 and 0.688 respectively. No sex differences were identified at any age threshold. Exact values corresponding to *Figure 3* can be found in *Table A2*.

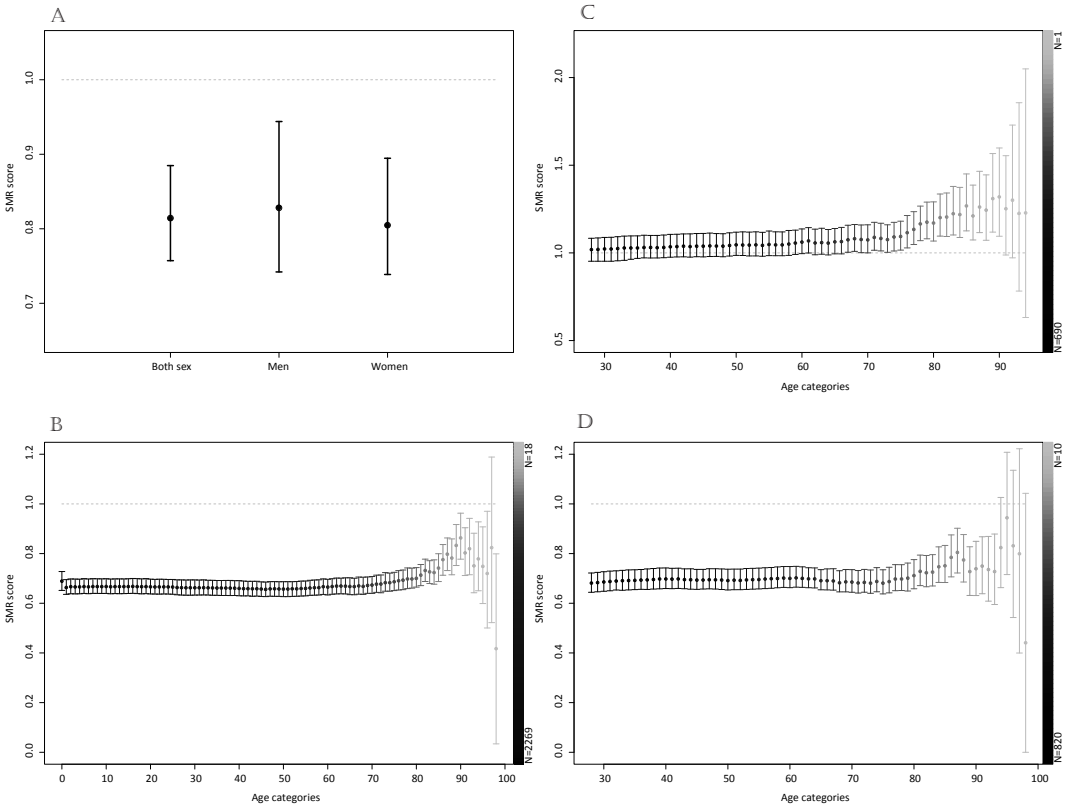


Figure 3: Standardized mortality ratio (SMR) for participants and lifetime SMR for first-degree relatives + spouses

(A) SMR for the LLS participants, (B) all age SMR for sibs (F1) of participants, (C) all age SMR for spouses (F1) of participants, and (D) all age SMR for parents (F0) of participants. The horizontal dotted line illustrates the SMR threshold value of 1. The nodes are SMR point estimates. The error bars represent the family bootstrapped confidence intervals. The colors in (B), (C), and (D) illustrate the sample size at every cutoff. The higher the age threshold, the lower the sample size, and hence, the lighter the color. The bars at the right side of the subfigures show the sample size associated with the colors of the SMRs.

Discussion

We investigated the survival of the longevous F1 LLS participants (who are longevous siblings) selected in the Leiden Longevity Study, and their F1 siblings, F0 parents, and F1 spouses. Based on the lifespan data of entire sibships (F1, participants+siblings), we observed a maternal transmission pattern of longevity with equal probability to sons and daughters. As compared to inclusion age matched singletons from similar birth cohorts and sex, LLS participants had 18.6% less observed deaths than expected, and thus a survival advantage. In the LLS the spouses of the participants had a life-long sustained survival pattern similar to the general population. Finally, we conclude that parents and siblings of the LLS participants had a life-long sustained survival advantage as compared to individuals matched on birth cohorts and sex.

Family longevity scores (FLS) were used to explore whether human longevity was transmitted preferentially through the maternal or paternal line, using the entire sibship information instead of only that of one single child per family. All sibships had an increased survival compared to individuals from the same birth cohort and sex, regardless of their parental longevity, because we selected LLS participants to have lived >= 89 and 91 years for men and women respectively. However, the median FLS for sibships with a long-lived mother and a non-long-lived father was 0.41 (P=0.024) higher than for sibships with a long-lived father and a non-long-lived mother, and 0.48 (P=0.008) higher than for sibships with both parents non-long-lived. This indicates that in the LLS longevity was transmitted preferentially via the maternal line. This maternal transmission of longevity is in concordance with the mitochondrial transmission hypothesis which posits that longevity may be transmitted through mitochondrial DNA from mothers to her offspring[8]. Though, this theory argues that because mitochondria are only maternally inherited they are under selection pressure for optimized compatibility with only the female genome, we have no evidence that there is preferential transmission of longevity from mothers to daughters. Another explanation connects to Fogel's (1997) theory of technophysio evolution which explains that in the turn of the 19th to the 20th, century childhood and early life mortality decreased significantly. This decrease was attributed to an increased birth weight and height of children and young adults respectively[46]. Since mothers are pivotal in this process it might be that the long-lived mothers were able to give birth to such healthy children whereas this may not have been the case for non-long-lived mothers, irrespective of the beneficial effect that 19th century long-lived fathers may have provided. The similarity in LFS for sibships with a long-lived father and a non-long-lived mother (LFS=1.21) and sibships with both parents non-long-lived (LFS=1.14) indicates the small influence of paternal effects compared to maternal effects. This absence may indicate that paternal socio-economic status in the LLS is of marginal influence to the intergenerational transmission of longevity[47,48]. Sibships with a long-lived mother and a non-long-lived father had not only had a higher LFS, they also had a mean sibship size of 5 whereas the two other categories had a mean sibship size of 8.34 and 7.95. In general, the probability of finding long lived subjects in families increases with sibship size[49]. The finding of longevity among children in small sibships (with a long-lived mother) may therefore indicate that the longevity is less likely to be prominent by chance. The smaller sibship size of LL mothers may be explained by a trade off in longevity families, either based on environmental (i.e. limited economic resources) or biological (i.e, reproductive capacity) factors. The discordant parental groups were quite small *(Figure 2)*. We identified sibships with a long lived father but not mother, and vice versa (N_sibships=21 ; N_individuals=177 and N_sibships=17 ; N=individuals=85) which interestingly shows that the maternal transmission effects are found not in

all, but in a subset of LLS families.

To investigate familial clustering of longevity, studies selected long-lived subjects based on multiple siblings or singletons[9–11,16]. So far it was unclear whether a sibling based selection provides a survival advantage over singletons. We showed that longevous siblings (F1 LLS participants) indeed had an 18.6% survival advantage over inclusion age, birth cohort, and sex matched longevous singletons. The effect can be considered large because the observational period focuses on the last stage of life (age >= 89 and 91 for men and women), especially when taking into account that siblings of LLS participants, who's full life course was observed showed a 33.7% survival advantage. It might even be expected that confining the sample to participants consisting of 3 or more longevous siblings increases the survival advantage. We did not, however, have the sample size to stratify our analyses to specific numbers of longevous participants within a family. Furthermore, we accounted for direct selection effects, although we could not directly account for the possibility that more healthy persons enrolled in the LLS than unhealthy persons or vice versa. We, however, did not expect that this has influenced our results since the first participants died only a few weeks after inclusion. We conclude that, when compiling a long-lived study cohort, selecting longevous siblings is a more stringent selection than longevous singletons of the same age.

Literature is inconclusive about the potential survival advantage of spouses of long-lived persons[10,11,35,36]. We showed, in a large group, that spouses of longevous LLS participants (N=809) had an equal survival to the general population until the age of 74. Beyond 74 years we observed a small excess mortality. We have no other explanation for this finding than the fact that this excess mortality beyond 74 years may be a function of small sample size. Pedersen et al. (2017) observed a survival advantage in the long life family study for spouses of long-lived siblings when comparing them to a birth cohort and sex matched control group. The authors point to assortative mating as a factor explaining the survival advantage for spouses of longevous participants[10]. An earlier Quebec study also reported a survival advantage of spouses[35] and a study of Southern Italy found male nonagenarians to outlive their spouses, whereas this was not the case for female nonagenarians[36]. Clearly, biological, environmental, and cultural factors influence survival to advanced ages in longevous families.

Because of unique Dutch lifetables dating back to 1850, we were able to show that parents of longevous LLS participants had a life-long sustained survival advantage compared to birth cohort and sex matched controls, until at least the age of 93 years. Beyond 94 years the confidence intervals increased due to a limited sample size. The life-long sustained survival advantage of first-degree relatives indicates a familial clustering of human longevity, which may be the result of the absence of deleterious genetic mutations[50,51] or the presence of genetic mutations protecting from aging related diseases[52]. Genetic studies aimed at identifying longevity loci promoting a life-long survival advantage up to the highest ages requires a focus on extreme individuals: cases belonging to the top 1-5% survivors with comparable parents. Recent genetic studies in the large UK Biobank[50,51] focused on subjects of 70 years on average without a parental selection[51] or selecting on parents belonging to the top 10% survivors[50]. This selection resulted in loci known to influence healthy ageing and mortality in middle and older age rather than exceptional longevity. As alternative to genetic influences, shared lifestyle or environmental factors may influence the longevity clustering in families. With the SMR analyses we could not adjust for environmental and lifestyle factors. However, the fact that we found spouses to survive comparable to the general population and that first

degree relatives (siblings and parents) had a life-long sustained survival advantage suggests a familial/genetic influence on human longevity, possibly acting from early life onward.

Longevity clusters within specific families and insight into this familial clustering is important in gaining knowledge of factors involved in a life-long survival advantage up to the highest ages. Knowledge about the inheritance pattern of longevity may be useful for genetic studies trying to discover longevity related genes. For example, effects of mitochondrial genes on human longevity should be investigated in those families with a history of maternal transmission of human longevity. Furthermore, research aiming to establish a study cohort of long-lived persons should ideally take family information into account, because we have demonstrated an enhanced survival for longevous siblings (LLS participants) over birth cohort and sex matched singletons. In the LLS, spouses seem comparable to the general population, making them a suitable comparison group for various health-related phenotypes as well as longevity. Lastly, as compared to sex and birth cohort matched individuals, parents of the LLS participants at the turn of the 19th century have a life-long sustained survival advantage up to the highest ages which was previously reported for the 20th century survival of siblings of longevous singletons[9,10,16]. This indicates that when studying the determinants of longevity factors involving the entire lifespan may contribute and emphasize the importance of longitudinal population based studies in the search for protective factor for age-related disease.

Conflict of interest

The authors declare that they have no conflict of interest

Funding

This work was supported by the Netherlands Organization for Scientific Research [360-53-180]

Informed consent

Informed consent was obtained from all Leiden Longevity Study participants

References

1 Oeppen, J. & Vaupel, J. W. Broken limits to life expectancy. *Science* **296**, 1029–1031 (2002).

2 Fogel, R. W. Aging, *Health and Public Policy.* **30**, (2004).

3 Meij, J. J. *et al.* Quality-quantity trade-off of human offspring under adverse environmental conditions. *J. Evol. Biol.* **22**, 1014–1023 (2009).

4 Smith, K. R., Mineau, G. P. & Bean, L. L. Fertility and post-reproductive longevity. *Soc. Biol.* **49**, 185–205 (2002).

5 Christensen, K., Johnson, T. E. & Vaupel, J. W. The quest for genetic determinants of human longevity: challenges and insights. *Nat. Rev. Genet.* **7**, 436–48 (2006).

6 Temby, O. F. & Smith, K. R. the Association Between Adult Mortality Risk and Family History of Longevity: the Moderating Effects of Socioeconomic Status. *J. Biosoc. Sci.* **46**, 703–716 (2014).

7 Mitnitski, A. B., Graham, J. E., Mogilner, A. J. & Rockwood, K. Frailty, fitness and late-life mortality in relation to chonological and biological age. *BMC Geriatr.* **2**, 1–8 (2002).

8 Austad, S. N. & Fischer, K. E. Sex Differences in Lifespan. *Cell Metab.* **23**, 1022–1033 (2016).

9 Perls, T. T. *et al.* Life-long sustained mortality advantage of siblings of centenarians. *Proc. Natl. Acad. Sci.* **99**, 8442–8447 (2002).

10 Pedersen, J. K. *et al.* The Survival of Spouses Marrying Into Longevity-Enriched Families. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **72**, 109–114 (2017).

11 Schoenmaker, M. *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur. J. Hum. Genet.* **14**, 79–84 (2006).

12 Andersen, S. L., Sebastiani, P., Dworkis, D. a., Feldman, L. & Perls, T. T. Health Span Approximates Life Span Among Many Supercentenarians: Compression of Morbidity at the Approximate Limit of Life Span. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **67A**, 395–405 (2012).

13 Ash, A. S. *et al.* Are Members of Long-Lived Families Healthier Than Their Equally Long-Lived Peers? Evidence From the Long Life Family Study. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **70**, 971–976 (2015).

14 Christensen, K., McGue, M., Petersen, I., Jeune, B. & Vaupel, J. W. Exceptional longevity does not result in excessive levels of disability. *Proc. Natl. Acad. Sci.* **105**, 13274–13279 (2008).

15 Evert, J., Lawler, E., Bogan, H. & Perls, T. Morbidity Profiles of Centenarians: Survivors, Delayers, and Escapers. *Journals Gerontol. - Med. Sci.* **58**, 232–237 (2003).

16 Willcox, B. J., Willcox, D. C., He, Q., Curb, J. D. & Suzuki, M. Siblings of Okinawan centenarians share lifelong mortality advantages. *J. Gerontol. A. Biol. Sci. Med. Sci.* **61**, 345–54 (2006).

17 Gavrilov, L. a., Gavrilova, N. S., Olshansky, S. J. & Carnes, B. a. Genealogical data and the biodemography of human longevity. *Biodemography Soc. Biol.* **49**, 160–173 (2002).

18 Kerber, R. A., Brien, E. O., Smith, K. R. & Cawthon, R. M. Familial Excess Longevity in Utah Genealogies. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **56**, 130–139 (2001).

19 Matthijs, K., Putte, B. & Vlietinck, R. The Inheritance of Longevity in a Flemish Village. *Eur. J. Popul. Eur. Démographie* **18**, 59–81 (2002).

20 Westendorp, R. G. J. & Kirkwood, T. B. L. Maternal and Paternal Lines of Familial Longevity. *Popul. An English Sel.* **13**, 223–235 (2001).

21 Cournil, A., Legay, J. M. & Schächter, F. Evidence of sex-linked effects on the inheritance of human longevity: a population-based study in the Valserine valley (French Jura), 18-20th centuries. *Proc. Biol. Sci.* **267**, 1021–5 (2000).

22 Gavrilov, L. A. & Gavrilova, N. S. Biodemographic Study of Familial Determinants of Human Longevity. *Popul. An English Sel.* **13**, 197–221 (2001).

23 Abbott, M. H., Abbey, H., Bolting, D. R. & Murphy, E. A. The Familial Component in Longevity-A Study of Offspring of Nonagenarians : 111 . Intrafamilial Studies. *Am. J. Med. Genet.* **2**, 105–120 (1978).

24 Philippe, P. & Opitz, J. M. Familial correlations of longevity: An isolate-based study. *Am. J. Med. Genet.* 2, 121–129 (1978).

25 Dutta, A. et al. Longer lived parents: Protective associations with cancer incidence and overall mortality. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **68**, 1409–1418 (2013).

26 Bocquet-Appel, J. P. & Jakobi, L. Familial transmission of longevity. *Ann. Hum. Biol.* **17**, 81–95 (1990).

27 Danzhen You, Danan Gu & Zeng Yi. Familial Transmission of Human Longevity Among the Oldest-Old in China. *J. Appl. Gerontol.* **29**, 308–332 (2010).

28 Lach, S., Riton, Y. & Simhon, A. *Longevity across generations.* (2006).

29 Parman, J. Gender and Intergenerational Mobility : *Using Health Outcomes to Compare Intergenerational Mobility Across Gender and Over Time.* (2010).

30 Hallett, M., Human, R. a T. & Physiology, C. familial longevity: the older you are the older your father may have been. *J. Am. Geriatr. Soc.* **53**, 357–358 (2005).

31 Salaris, L., Tedesco, N. & Poulain, M. Familial transmission of human longevity : a population-based study in an inland village of Sardinia (Italy), 1850 – 2010. *Vienna Yearb. Popul. Res.* **11**, 325–349 (2013).

32 Caselli, G. et al. Family clustering in Sardinian longevity: A genealogical approach. *Exp. Gerontol.* **41**, 727–736 (2006).

33 Kemkes-Grottenthaler, A. Parental effects on offspring longevity—evidence from 17th to 19th century reproductive histories. *Ann. Hum. Biol.* **31**, 139–158 (2004).

34 van den Berg, N., Beekman, M., Smith, K. R., Janssens, A. & Slagboom, P. E. Historical demography and longevity genetics: Back to the future. *Ageing Res. Rev.* **38**, 28–39 (2017).

35 Jarry, V., Gagnon, A. & Bourbeau, R. Survival advantage of siblings and spouses of centenarians in 20th-century Quebec. *Can. Stud. Popul.* **39**, 67 (2013).

36 Montesanto, A. *et al.* The genetic component of human longevity: analysis of the survival advantage of parents and siblings of Italian nonagenarians. *Eur. J. Hum. Genet.* **19**, 882–886 (2011).

37 Deluty, J. A., Atzmon, G., Crandall, J., Barzilai, N. & Milman, S. The influence of gender on inheritance of exceptional longevity. *Aging (Albany. NY).* **7**, 412–418 (2015).

38 Wilmoth, J. R. & Shkolnikov, V. Human Mortality Database. *University of California, Berkeley (USA), and Max Planck Institute for Demographic Research (Germany)* (2017).

39 Sebastiani, P., Nussbaum, L., Andersen, S. L., Black, M. J. & Perls, T. T. Increasing Sibling Relative Risk of Survival to Older and Older Ages and the Importance of Precise Definitions of "Aging," "Life Span," and "Longevity". *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **71**, 340–346 (2016).

40 Van Der Meulen, A. *Life tables and Survival analysis*. (2012).

41 Carolina, T., Uijvenhoven, L. & van der Laan, J. Overlevingstafels en longitudinale analyse. *CBS* 1–25 (2009).

42 R Core Team. R: A language and environment for statistical computing. (2016).

43 Rozing, M. P. *et al.* Familial longevity is associated with decreased thyroid function. *J. Clin. Endocrinol. Metab.* **95**, 4979–4984 (2010).

44 Sebastiani, P. *et al.* A Family Longevity Selection Score: Ranking Sibships by Their Longevity, Size, and Availability for Study. *Am. J. Epidemiol.* **170**, 1555–1562 (2009).

45 Bauer, D. F. Constructing Confidence Rank Sets Using Statistics. *J. Am. Stat. Assoc.* **67**, 687–690 (1972).

46 Fogel, R. W. & Costa, D. L. A Theory of Technophysio Evolution, With Some Implications for Forecasting Population, Health Care Costs, and Pension Costs. *Demography* **34**, 49 (1997).

47 Razzell, P. & Spence, C. The hazards of wealth: Adult mortality in pre-twentieth-century England. *Soc. Hist. Med.* **19**, 381–405 (2006).

48 Schenk, N. & Van Poppel, F. Social class, social mobility and mortality in the Netherlands, 1850-2004. *Explor. Econ. Hist.* **48**, 401–417 (2011).

49 Rodríguez-Girondo, M., Deelen, J., Slagboom, E. P. & Houwing-Duistermaat, J. J. Survival analysis with delayed entry in selected families with application to human longevity. *Stat. Methods Med. Res.* **0**, 1–22 (2016).

50 Pilling, L. C. *et al.* Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany. NY).* **9**, 2504–2520 (2017).

51 Joshi, P. K. *et al.* Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nat. Commun.* **8**, 1–13 (2017).

52 Beekman, M. *et al.* Genome-wide association study (GWAS)-indentified disease risk alleles do not compromise human longevity. *Proc. Natl. Acad. Sci. USA* **107**, 18046–18049 (2010).

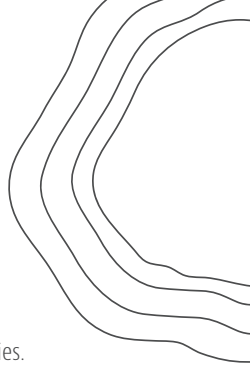
CHAPTER 6

LONGEVITY RELATIVES COUNT SCORE DEFINES HERITABLE LONGEVITY CARRIERS AND SUGGESTS CASE IMPROVEMENT IN GENETIC STUDIES.

Niels van den Berg[1,2]
Mar Rodríguez-Girondo[3]
Kees Mandemakers[4]
Angelique A.P.O. Janssens[2]
Marian Beekman[1]
P. Eline Slagboom[1,5]

[1] Department of Biomedical Data Sciences, section of Molecular Epidemiology,
Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands
[2] Radboud Group for Historical Demography and Family History,
Radboud University, Erasmusplein 1, 6525 HT Nijmegen, the Netherlands
[3] Department of Biomedical Data Sciences, section of Medical Statistics,
Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands
[4] International Institute of Social History,
Cruquiusweg 31, 1019 AT Amsterdam, the Netherlands
[5] Max Planck Institute for Biology of Ageing,
Joseph-Stelzmann-Str. 9b, D-50931 Cologne, Germany

BioRxiv (April 2019)
https://doi.org/10.1101/609891



Abstract

Longevity loci represent key mechanisms of a life-long decreased mortality and decreased/compressed morbidity. However, identifying such loci is challenging. One of the most plausible reasons is the uncertainty in defining long-lived cases with the heritable longevity trait amongst long-living phenocopies. To avoid phenocopies, family selection scores have been constructed but these have not yet been adopted as state of the art in longevity research. Here we aim to identify individuals with the heritable longevity trait by using current insights and a novel family score based on these insights. We use a unique dataset connecting living study participants to their deceased ancestors covering 37,825 persons from 1,326 five-generational families, living between 1788 and 2019. Our main finding suggests that longevity is transmitted for at least 2 subsequent generations only when at least 20% of all relatives are long-lived. This proves the importance of family data to avoid phenocopies in genetic studies.

Main

In contrast to the low heritability of human lifespan[1–4], human longevity is strongly heritable as illustrated by the familial clustering of survival into extreme ages[5,6,15–17,7–14]. Identifying longevity loci is important because these loci likely represent key mechanisms of a life-long decreased mortality[12,13], decreased morbidity[9,17,18] and compression of morbidity towards the end of the lifespan[19–21]. Currently, genome wide linkage and association studies (GWAS) identified a limited number of loci promoting longevity[22–31], for example the APOE and FOXO3A genes (more details can be found in current review papers[22,23,30]). However, many of the identified loci could not be replicated in independent studies as yet. In addition, the largest and most recent longevity GWAS, based on cases belonging to the top 10% oldest survivors, again only replicated association of the APOE locus[32].

One of the main reasons for the limited success of longevity genetic studies[24–26,31–34] is the uncertainty in defining the heritable longevity trait itself[1,13]. Given the increased life expectancy of the past 200 years due to non-genetic factors (improved hygiene, nutrition and medication) there are likely many phenocopies among the long-lived cases selected for our genetic studies[35,36]. The presence of phenocopies is illustrated by the increase of centenarians in the United States between 1994 and 2012 from 1 in 10,000 to 1 in 5,000[37]. To avoid phenocopies, family selection scores, such as the Family Longevity Selection Score (FLoSS) and the Family Excess Longevity (FEL) score have been constructed[38,39]. The use of such scores is substantiated by novel studies which showed that that including family history information can provide valuable information about an individual's genetic liability for a trait and is likely to increase the power to detect genetic[40–42]. The scores focus, in different ways, on selecting multiple family members with the same trait[12,38,39,43,44] and usually focus on a single group of relatives, such as parents[12,43] or siblings[39] of cases.

As the definition of heritable longevity was not yet established, the construction and application of the family selection scores have not yet been adopted as state of the art in longevity research. As such, the majority of genealogical[5–11,45] and genetic studies[24–26,31–34] focus only on single, and thus including sporadic, long-lived individuals (singletons), with some exceptions focusing for example on parental age[28,29] or multiple siblings[15,25]. In previous work, we showed that longevity defined as top 10% survivors or more extreme is transmitted to subsequent generations[13]. With this, a consistent definition of longevity was provided that is also adopted in the largest longevity GWAS up to now[32]. In addition, we showed that every additional long-lived relative independently contributes to the survival advantage of study participants, according to their genetic distance[13]. As such, there is room to incorporate these novel insights into family selection scores to gain knowledge about the extent that longevity needs to cluster in families in order to include individuals with the heritable longevity trait and increase the power of genetic studies.

Here, we aim to establish the proportion of ancestral blood relatives that should be long-lived (top 10% survivors of their birth cohort or more extreme) in order to observe a survival advantage in their descendants and incorporate these insights into a novel family score to define cases with the heritable longevity trait for inclusion in genetic studies. For our analyses we use the data available in the Historical Sample of the Netherlands (HSN) for the period between 1860 and 1875 which is based on Dutch citizens[46–48]. We primarily identify cases who died beyond 80 years (N=884, on average top 10% survivors

of their birth cohort), allowing us to select on more extreme ages at death, and controls who died between 40 and 59 years (N=442). We extend this filial (F) 1 generation data with a parental and 3 descendant generations of individual life course and mortality data and refer to the data as the HSN case/control dataset. We subsequently exclude groups with high rates of missing mortality information and where the majority was still alive *(Supplementary Figure 4)*. This study covers 37,825 persons from 1,326 three-generational families (F1-F3) and contains F1 index persons (IPs), 2 consecutive generations of descendants (F2-F3) and 2 generations of spouses (F2-F3) *(Table 1)*. The dataset is unique in that it covers multiple generations and connects alive persons to at least two generations of deceased ancestors.

Results

Outline

We analyzed the data across multiple steps *(Supplementary Figure 5)* in two phases. In the first phase, we used Standardized Mortality Ratios (SMRs) to compare the transmission of longevity for cases (died beyond 80 years) and controls (died between 40 and 59 years) as defined in the original approach *(Figure 1A)*, focusing on the F1 index persons (IPs) and two generations of descendants.

In the second phase of our study (the combined approach), we combined original cases and controls and their descendants into one combined group and focused on the survival of the F3 descendants in relation to their F2 and F1 ancestral family members *(Figure 1B)*. First, we constructed the Longevity Relatives Count (LRC) score. We used the LRC score to investigate the proportion of long-lived (top 10% survivors of their birth cohort) F1 and F2 ancestors required for F3 descendants to express a survival advantage compared to members of the same birth cohort and sex (family method, *Figure 1B*). On the basis of these observations we defined a new case and control group in F3, where we labeled F3 descendants with ≥30% long-lived ancestors as family cases and those without long-lived ancestors as family controls. Subsequently, these F3 family cases and controls were compared for their survival, that of their spouses (to investigate environmental influences), and for survival differences with the F3 descendants, selected to have at least one (singleton) long-lived ancestor or at least one average-lived ancestor. This means that they could have more than 1 long or average lived ancestor but we actively selected for the presence of only 1 such ancestor. *Supplementary Figure 3A* provides a conceptual overview of this selection. To this end, we selected either F3 descendants with at least one top 10% grandparent, at least one top 10% parent, or with grandparents who died between 40 and 59 years (their children (parents) resembled the general population). In a final step, we focused on the F3 descendants with at least one long-lived parent and calculated LRC scores within this F3 group to determine if parents transmitted their longevity more frequently if they were part of a long-lived (LRC ≥ 0.30) family *(Figure 1B)*. The analysis steps are summarized in *Supplementary Figure 5* and an overview of the available data per group and generation is shown in *Table 1*.

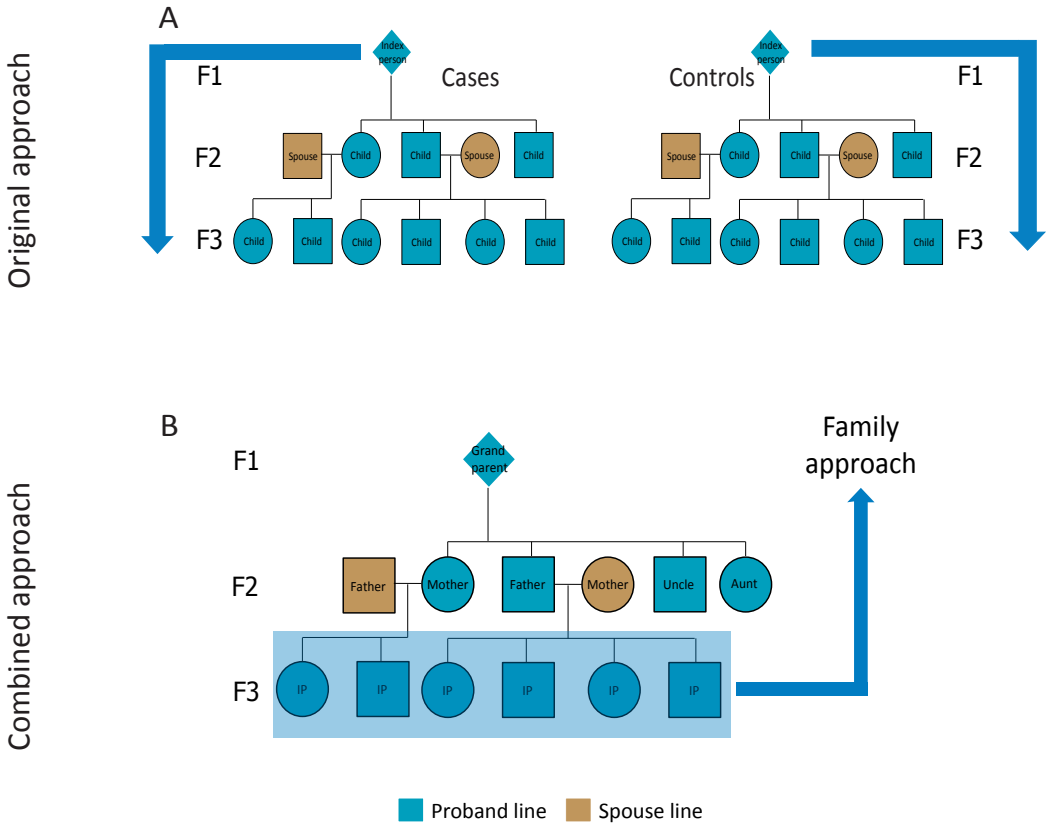


Figure 1: Pedigree overview of the data structure.

This figure illustrates the two approaches; 1. the original approach and 2. the combined approach. The original approach refers to the case and control group based on the F1 IPs where cases died at 80 years or older and controls died between 40 and 59 years (panel A). Panel B shows a pedigree of the data from the perspective of F3 children (combined approach). The combined approach refers to the dataset where we combined the cases and controls form the original design and constructed a new case and control group in the F3 descendants. To this end, F3 descendants with ≥30% long-lived ancestors were labeled as family cases and those without long-lived ancestors as family controls. F3 spouses were left out of this figure but this group was used to confirm a genetic enrichment in the F3 descendants.

Table 1: Overview study sample for groups in all generations based on the proband and F3 perspective

Role	Number	Deceased (%)	Alive (%)	Femail (%)	Range Birth cohort	Mean age (sd)	Median age (sd)	Missing age (%)
Cases (Original design)								
F1 IPs	884	884 (100)	0 (0)	422 (50)	1860-1875	85.79 (4.59)	84.99 (4.95)	0 (0)
F2 descendants	4916	4405 (90)	11 (1)	2435 (50)	1879-1941	63.04 (31.11)	75.51 (17.72)	500 (9)
F2 spouses	3899	1500 (38)	16 (1)	1504 (38)	1873-1934	76.2 (15.09)	78.78 (12.83)	2383 (61)
F3 descendants	9910	4869 (49)	4146 (42)	4733 (48)	1901-1973	70.35 (19.54)	74.77 (11.38)	895 (9)
F3 spouses	3431	1289 (38)	792 (23)	1963 (57)	1900-1959	77.14 (11.31)	79.25 (10.1)	1350 (39)
F4 descendants*	9001	746 (8)	7172 (80)	3937 (44)	1922-1995	57.7 (10.68)	58.21 (9)	1083 (12)
Controls (Original design)								
F1 IPs	442	442 (100)	0 (0)	214 (48)	1860-1875	51.71 (5.71)	52.88 (6.21)	0 (0)
F2 descendants	2488	2202 (89)	1 (<1)	1217 (49)	1881-1925	58.17 (32.49)	71.72 (21.37)	285 (11)
F2 spouses	1877	690 (37)	7 (<1)	734 (39)	1875-1935	76.02 (14.77)	78.34 (13.76)	1180 (63)
F3 descendants	4761	2540 (53)	1813 (38)	2265 (48)	1904-1966	69.39 (20.38)	74.49 (11.36)	408 (9)
F3 spouses	1778	721 (41)	376 (21)	972 (55)	1893-1965	76.54 (11.5)	78.66 (10.47)	681 (38)
F4 descendants*	4710	387 (8)	3744 (80)	2099 (45)	1871-1992	57.72 (11.17)	58.37 (9.35)	579 (12)
F3 perspective (Combined design)								
F3 descendants	14671	7409 (51)	5959 (41)	6998 (48)	1901-1973	70.03 (19.82)	74.68 (11.38)	1303 (8)
F3 spouses	5209	2010 (38)	1168 (22)	2935 (55)	1893-1965	76.93 (11.38)	79.07 (10.24)	2031 (40)
F2 parents	9728	6139 (63)	23 (1)	4137 (43)	1873-1935	76.8 (13.4)	78.9 (12.31)	3566 (36)
F2 aunts & uncles	7036	6382 (91)	10 (1)	3456 (49)	1879-1941	61.81 (31.47)	74.4 (18.67)	644 (8)
F1 grandparents	1181	1181 (100)	0 (0)	560 (47)	1860-1875	74.88 (16.6)	81.94 (9.72)	0 (0)

The Cases and Controls rows provide an overview of the groups of persons from the original case/control perspective of the data, described as part a. The F3 perspective rows provide an overview of the groups of persons from the perspective of F3 descendants, described as part b. mean and missing age refer to an unknown age at death or an unknown age at last observation. For the F0 and F1 groups we assume everyone is dead because the birth cohorts date back further than 120 years. From the F2 generations we requested Personal Records Data indicating if a person was still alive or not and if not, what the date of death was. The F1 IPs are the focal persons in the pedigrees as they are selected to be 80 years or older (cases) or to have died between 40 and 59 years (controls). * indicates that the group is excluded for this study, sd refers to standard deviation.

Longevity is transmitted in the case group and not in the control group

Focusing on the original approach *(Figure 1A)*, we determined to what extent longevity is transmitted in the original case and the control group by estimating SMRs per generation for all cases and controls separately. *Table 2* shows that F1 cases had a similar survival pattern to birth cohort members of the same sex, indicating that they resemble a representative group of random Dutch persons aged ≥ 80 years and born between 1860 and 1875. The SMR for the descendants of the cases (F2 case descendants) was 0.87 (95%CI=0.84-0.89), indicating 13% less deaths than expected based on individuals from a similar birth cohort and sex. From here we refer to this as 13% excess survival (or, if appropriate, excess mortality) compared to the general population. The descendants of controls (F2 control descendants) had a similar survival pattern to the general population (SMR=1.01 (95%CI=0.96-1.05)). The spouses of the F2 case and control descendants surprisingly also showed a pattern of excess survival ($SMR_{case\_F2spouses}$=0.89 (95%CI=0.85-0.94) and $SMR_{control\_F2spouses}$=0.9 (95%CI=0.83-0.97)). Next we observed 14% (95%CI=11%-16%) excess survival compared to the general population for F3 descendants of the F1 cases, whereas F3 control descendants resembled the general population (SMR=0.96 (95%CI=0.93-1.00)) just as observed in the F2 generation. The spouses of both F3 groups resembled the general population ($SMR_{case\_F3spouses}$=1.00 (95%CI=0.95-1.05) & $SMRc_{ontrol\_F3spouses}$=1.07 (95%CI=0.99-1.15)). We conclude that two descendant generations of cases, who belong on average to the top 10% survivors, have 13-14% excess survival compared to the general populations and that the descendants of controls resemble the general population.

Table 2: Standardized mortality ratios for original case and control group individuals

Role	Case group SMRs	Number (N)	Control group SMRs	Number (N)	Adjustment for right truncation
F1 IPs	1.06 (0.99-1.13)	884	NA	NA	80 years
F2 descendants	0.87 (0.84-0.89)	4416	1.01 (0.96-1.05)	2203	No adjustment
F2 spouses	0.89 (0.85-0.94)	1516	0.9 (0.83-0.97)	697	20 years
F3 descendants	0.86 (0.84-0.89)	9015	0.96 (0.93-1.00)	4353	No adjustment
F3 spouses	1.00 (0.95-1.05)	2081	1.07 (0.99-1.15)	1097	20 years

Original cases (F1 IPs) died at 80 years or older, original controls (F1 IPs) died between 50 and 69 years. If persons could not die before a specific age due to direct or indirect selection, due to for example that all persons in a group were selected to have a child an adjustment for right truncation was applied so that a fair comparison could be made with their birth cohort members. An SMR for F1 control IPs could not be estimated due to a combination of left and right truncation in the data. The lifetables can only be adjusted for right or left truncation, but not a combination between the two.

To explore to what extent the survival of F2 and F3 descendants depends on the extremity of the longevity of their parents, we calculated SMRs for F2 and F3 case and control descendants with increasing parental longevity (for example, a parent belonged to the top 10%, 5%, or 1% survivors). We observed that the SMR decreased in descendants when defining parental longevity in terms of more extreme survival percentiles. This was the case for descendants of both the IP cases and controls although the effects were stronger in the descendants of the cases, especially in F3, since this group is now selected to have long-lived parents and grandparents *(Supplementary Table 1)*. This illustrates that selection on single long-lived persons belonging on average to the top 10% survivors, as we did for the IP selection, leads only to a modest transmission of

longevity in two generations (max 14%). Likely, the control group includes misclassified persons of which the descendants do live longer, whereas the case group includes long-lived persons that do not transmit longevity to their descendants (potentially these are phenocopies). Such misclassification can jeopardize genetic studies immensely. To be able to evaluate living persons as potential carriers of the heritable longevity trait in genetic studies, we constructed and validated a familial longevity score.

Constructing the Longevity Relatives Count score

We now look at the HSN data from a different perspective, the combined approach *(Figure 1B)*. In the combined approach we consider the F3 generation as the focal point of the pedigree, instead of the F1 generation, as was the case in the original approach. To identify individuals with the heritable longevity trait, we constructed the LRC score.

$$LRC_i = \frac{\text{weighted number of top 10\% ancestors}}{\text{weighted total number of ancestors}} = \frac{\sum_{k=1}^{N_i} w_k \cdot I(P_k \geq 0.9)_i}{\sum_{k=1}^{N_i} w_k}$$

Where k=1,...,Ni are all the available ancestral blood relatives (from here: ancestors) of F3 descendant i used to build the score (parents, aunts and uncles and grandparent of the F3 descendants, *Figure 1B*), Pk is the sex and birth year-specific survival percentile, based on lifetables, of ancestor k, and I(Pk ≥ 0.9) indicates if ancestor k belongs to the top 10% survivors. $\sum_{k=1}^{N_i} w_k$ is the weighted total number of ancestors of F3 descendant i. The relationship coefficients are used as weights wk. The LRC score indicates the proportion of ancestors that has become long-lived. For example, an LRC of 0.5 indicates 50% long-lived ancestors (see methods for a more detailed and general description of the LRC score).

Longevity is transmitted when at least 20% of all ancestors are long-lived

To determine what proportion of long-lived ancestors could be associated with the survival of F3 descendants, we calculated LRC scores for all F3 descendants and subsequently defined 9 mutually exclusive LRC groups (g) of F3 descendants: LRC_g1=0, LRC_g2=[>0 & <0.1], LRC_g3=[≥0.1 & <0.2], LRC_g4=[≥0.2 & <0.3], LRC_g5=[≥0.3 & <0.4], LRC_g6=[≥0.4 & <0.5], LRC_g7=[≥0.5 & <0.6], LRC_g8=[≥0.6 & <0.7], LRC_g9=[≥0.7 & ≥1.0]. For each group of F3 descendants we explored whether they have a survival benefit compared to the general population by estimating SMRs *(Figure 2)*. F3 descendants without any long-lived ancestors (LRC score of 0) had a survival pattern that resembled the general population (SMR=0.97 (95%CI=0.93-1.01)). Similarly, we observed a survival pattern that resembled the general population for F3 descendants with up to 20% long-lived ancestors (group 2 and 3, SMR=0.97 (95%CI=0.91-1.04) and SMR=0.95 (95%CI=0.91-1.00) respectively). This shows that the long-lived ancestors of group 2 and 3 F3 descendants were likely phenocopies instead of genetically enriched long-lived persons. We observed a pattern of excess survival for F3 descendants with more than 20% long-lived ancestors. The weakest significant effect was observed for group 3, with an SMR of 0.84 (95%CI=0.80-0.89) which is comparable to the excess survival of the F3 descendants of the singleton F1 cases in the original approach (first part of the results). The strongest significant effect was observed for group 8, with an SMR of 0.56 (95%CI=0.45-0.69). Hence, the higher the degree of long-lived ancestors, the lower the SMR. This indicates that the more long-lived ancestors an F3 descendant has, the higher the level of excess survival of these F3 descendants is compared to the general population, and the more likely that genetic effects drive the transmission of longevity.

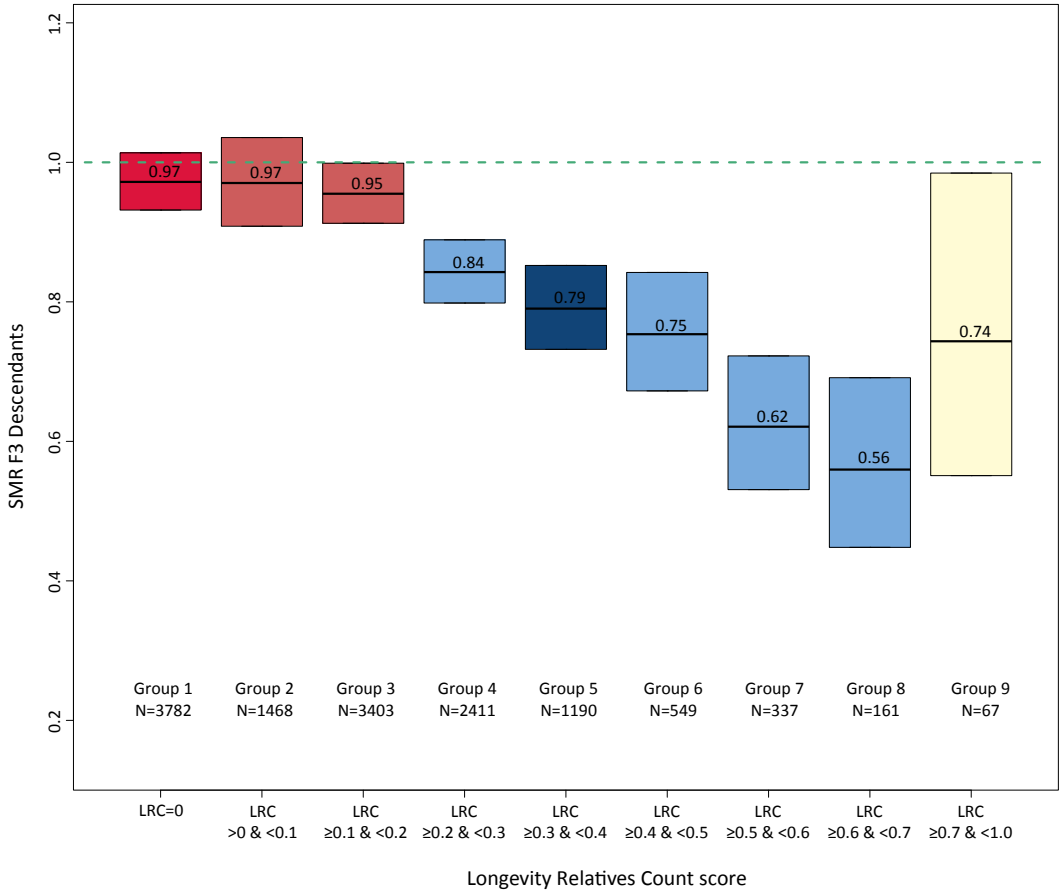


Figure 2: LRC score in mutually exclusive F3 descendant groups.

The figure shows Standardized Mortality Ratios for all F3 descendants without missing mortality information. The F3 descendants are grouped into mutually exclusive groups based on the Longevity Relatives Count (LRC) score. The LRC score represents the family approach as illustrated in Figure 1B. The dark red color of group one represents F3 descendants without any long-lived (top 10%) ancestors and are denoted as family controls. The light red represents F3 descendants who had more than 0 and less than 20% long-lived ancestors. The light blue colors represent the F3 descendants with 20% or more long-lived ancestors. The dark blue color represent our cut-off point for the family case definition. Hence all F3 descendants with 30% or more long-lived ancestors were considered family cases. The beige color of group 9 shows that this bar represents all F3 ancestors with more than 70% long-lived ancestors as their sample size was very low, we grouped them into one group.

Using the LRC score family method we defined a new case and control group in the F3 generation, which is based on the presence or absence of longevity among the ancestors of the F3 generation and potential excess survival or mortality in the F3 generation itself *(Figure 1B)*. The F3 family controls include all F3 descendants without any long-lived ancestors (LRC score of 0, N=4,166). To define the F3 family cases we chose an LRC cutoff based on a trade-off between the size and the uncertainty, given by the sample size, of the SMR. The F3 family cases include all F3 descendants with at least 30% long-lived ancestors (LRC score ≥ 0.30 (N=2,526)). Even if F3 family cases are not long-lived themselves, their survival reflects the presence of longevity of their ancestors, which is transmitted by their parents. Similarly, F3 controls reflect the

absence of longevity of their ancestors. *Supplementary Figure 1* shows the variation in lifespan of the F3 family case and control descendants. F3 descendants with more than 0% and up to 20% long-lived ancestors (LRC score >0 and < 0.2) did not express excess survival (N=5,340). The F3 descendants with an LRC score ≥0.2 and < 0.30 showed some excess survival compared to the general population, but the size of the SMR was considered too low to enter our family case definition. Hence, we denoted them as non-classified (N=2,639).

Strong survival advantage and genetic enrichment for F3 family cases

To validate the LRC score, we investigate survival differences, measured as age at death or last observation, between the F3 family cases and controls and used a Cox-type random effects (frailty) regression model to adjust for within-family relations of the F3 descendants. *Figure 4 and Table 3A* show that F3 cases have a 25% (95%CI=18-31%) lower hazard of dying than F3 controls, even after adjustment for sibship size, birth year, and sex. The difference between the cases and controls became increasingly more pronounced when confining the cases to a higher proportion of long-lived ancestors, for example an LRC score of 0.40, 0.50, or 0.60, reflecting 40%, 50%, or 60% long-lived ancestors *(Supplementary Figure 2)*. The strongest effect was observed for those with an LRC score ≥ 0.60 (hazard ratio (HR) of 0.62 (95%CI=0.50-0.77)). The mortality pattern for the spouses of these F3 cases resembled that of the F3 controls (HR=0.94 (95%CI=0.82-1.07), *Table 3B*) and the general population (SMR=0.92 (95%CI=0.83-1.02)). The survival of the spouses, equal to the F3 controls and the general population, in addition to the absence of effects of environmental covariate adjustment, indicates that environmental factors were likely of limited influence to the observed survival benefit of the F3 cases as defined by our novel family based definition. Hence, the observed survival benefit of F3 cases likely represents a genetic longevity component.

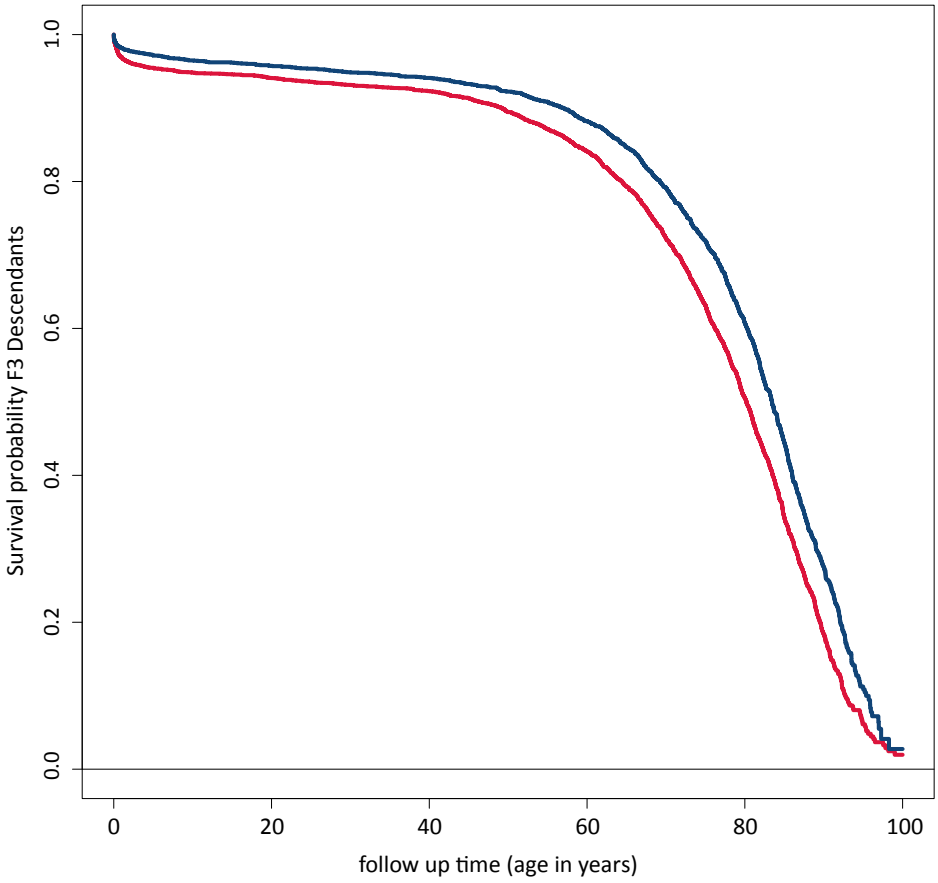


Figure 4: Survival differences between family based cases and their spouses.

This figure shows the survival curve for the difference in survival between the F3 family cases and controls. The figure is connected to Table 3A which shows the Hazard Ratios corresponding to the difference between the two curves. Blue color represent the cases, red color represents the controls.

Table 3: Mortality difference between family cases and controls and their spouses

	A			B		
	N (mean)	HR (95% CI)	P-value	N (mean)	HR (95% CI)	P-value
Family based case/control group						
Control group (ref)	3714 (0.62)			3714 (0.50)		
Case group	2282 (0.38)	0.75 (0.69-0.82)	1.75e-10	2282 (0.30)	0.74 (0.68-0.80)	4.08e-12
Spouses of cases				541 (0.07)	0.94 (0.82-1.07)	3.44e-01
Spouses of controls				937 (0.13)	1.12 (1.00-1.25)	4.07e-02
Birth year	5996 (1933)	0.99 (0.98-0.99)	1.99e-05	7474 (1932)	0.98 (0.98-0.99)	1.39e-12
Sex						
Males (ref)	3133 (0.52)			3364 (0.45)		
Females	2863 (0.48)	0.56 (0.52-0.61)	<1.00e-15	4110 (0.55)	0.49 (0.46-0.53)	<1.00e-15
Sibship size						
Small - 1-2 sibs (ref)	1531 (0.26)					
Medium - 3-5 sibs	1770 (0.30)	1.17 (1.04-1.32)	8.51e-03			
Large - 6-8 sibs	927 (0.15)	1.22 (1.04-1.43)	1.21e-02			
Exceptional - 9-15 sibs	441 (0.07)	1.36 (1.09-1.68)	5.84e-03			
Single child - 0 sibs	1327 (0.22)	1.81 (1.62-2.02)	<1.00e-15			

Table 3A corresponds to the CH curves of panel a of Figure 4. Means represent a mean for a continuous variable and a proportion for a categorical variable. When the p-value was lower than 1.00e-15 we indicated the P-value as <1.00*10-15. SES = socio-economic status, OCC = occupational coding scheme of 1950, CI = confidence interval, CH = cumulative hazard. P-values are estimated with cox regression. F3 children with relatives who were still alive and had no last moment of observation ≥ 100 years were removed to assure an equal comparison between cases and controls. In *Table 3B* the spouses of cases and controls are adjusted for the fact that they could not die before the birth of at least their first child (left truncation). We adjusted for this left truncation by entering the spouses of cases and controls in the model based on the first observed death in the groups (cases: 30 years and controls: 25 years). In model A no adjustment for left truncation was necessary. In both models we adjusted for right censoring by including a censoring indicator in the cox model.

Family cases live longer than those with one long-lived parent or grandparent

Next, we test if the F3 descendants with 30% long-lived ancestors (the family cases) have a stronger survival advantage than F3 descendants with at least 1 long-lived (top 10%) parent or grandparent. We actively selected this group of F3 descendants to have 1 long-lived parent or grandparent, meaning that other ancestors could also be long-lived but there was no active selection on the presence of their longevity *(Supplementary Figure 3A and 3B)*, hence the designation 'at least' for this group. Subsequently, we tested if F3 descendants without long-lived ancestors (the family controls) had a similar survival pattern to the F3 descendants with parents resembling the general population (those with a grandparent who died between 40 and 59 years). *Table 4* shows that we observed 14% (95%CI=11%-17%) excess survival compared to the general population for F3 descendants with at least one long-lived grandparent (F1). When identifying F3 descendants with at least one long-lived parent (F2), we observed 16% (95%CI=8%-24%) excess survival compared to the general population. Using the family method at 30% long-lived family members to identify F3 family cases, we observed 26% (95%CI=22%-30%) excess

survival compared to the general population and this increased to 38% (95%CI=31%-45%) when applying a 50% threshold to the family method. For the identification of controls both methods seem to preform equally well, with almost identical SMRs of around 1. This indicates that the F3 controls, whether defined by having no long-lived ancestors or by grandparents dying between 40 and 50 years, have a similar survival pattern to the general population. We conclude that, at least for cases, the family method provides a better contrast in excess survival compared to the general population and seems to better represent the heritable longevity trait.

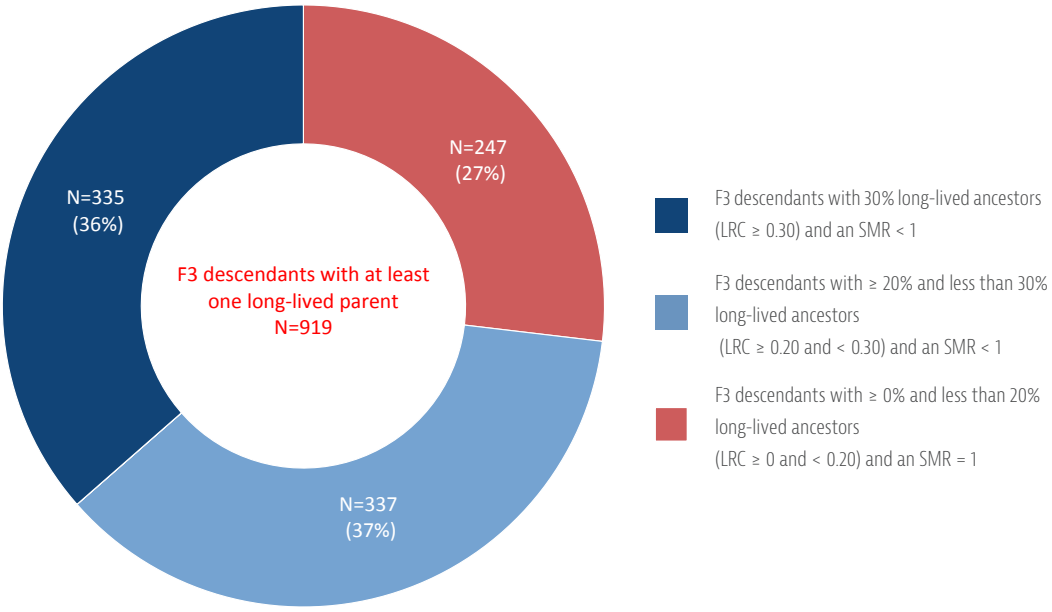


Figure 3: LRC score for F3 descendants with at least one long-lived parent.

This center of this doughnut figure shows all F3 descendants (N=919) with at least one long-lived (top 10%) parent, ignoring the rest of the ancestors. Thus, at least means that they could have more than 1 long-lived ancestor but we actively selected for the presence of only 1 such ancestor. The edges of the doughnut illustrate the number and proportion of these 919 F3 descendants with at least one long-lived parent who had 1. 30% or more long-lived ancestors (LRC ≥ 0.30) and excess survival compared to the general population (SMR < 1), N=335 (36%) 2. between 20% and 30% long-lived ancestors (LRC ≥ 0.20 and < 0.30) and excess survival compared to the general population (SMR < 1), N=337 (37%) and 3. between 0% and 20% long-lived ancestors (LRC > 0.20 and < 0.20) and a similar survival pattern to the general population (SMR ~ 1), N=247(27%).

Since the F3 descendants with ≥ 30% long-lived ancestors have a stronger survival advantage than those with at least one long-lived parent, it is possible to get an indication of how many F3 descendants did not appear to have a survival advantage compared to the general population, even though at least one parent was long-lived. This is relevant in view of case definitions used in large genetic studies into longevity. *Figure 3 and Supplementary Figure 3* show that 919 F3 descendants had a long-lived parent. Out of those 919 F3 descendants, 247 (27%) had more than 0% but less than 20% long-lived ancestors (LRC > 0 and < 0.20) and thus as a group had an SMR that resembled the general population *(Supplementary Figure 3D)*. The other 672 (73%) had exactly, or more than 20% long-lived ancestors (LRC ≥ 0.20) and thus, as a group, showed excess survival compared to the general population *(Supplementary Figure 3B and C)*. These results suggest that if living persons are selected as case in genetic studies on the basis of one long-lived parent, 27% of these persons is unlikely to be a carrier of the longevity trait. Persons defined as 30% long-lived ancestors, on the other hand would be potential carriers.

Table 4: Standardized Mortality Ratio for different F3 descendant groups

Group	SMR	N
Cases		
F3 descendant with at least one long-lived grandparent	0.86 (95%CI=0.83-0.89)	4986
F3 descendant with at least one long-lived parent	0.84 (95%CI=0.76-0.92)	852
F3 descendant with ≥ 30% long-lived ancestors (LRC ≥ 30%)	0.74 (95%CI=0.70-0.78)	2304
F3 descendant with ≥ 50% long-lived ancestors (LRC ≥ 50%)	0.62 (95%CI=0.55-0.96)	565
Controls		
F3 descendant with grandparent who died between 40 and 59 years	0.96 (95%CI=0.93-1.00)	4353
F3 descendant with no long-lived ancestors (LRC = 0)	0.97 (95%CI=0.93-1.01)	3782

Long-lived is defined as belonging to the top 10% survivors of their birth cohort. Note that the group size (N) reflects only those with a known age at death as this was necessary to estimate a standardized mortality ratio.

Discussion

Human longevity is heritable and clusters in specific families. Studying the familial clustering of longevity in these families is important to improve our understanding of genetic factors promoting longevity and healthy aging. The main observations supporting this are (1) In the original approach, we observed 14% excess survival of the cases compared to their birth cohort for two subsequent generations (F2-F3) while in the controls no such benefit was observed, (2) in the combined approach, the excess survival of the F3 cases compared to the general population was 26-38% depending on the proportion of long-lived family members being 30-50% and these estimates strongly overlap to the survival difference between the F3 family cases and controls based on the Cox models, (3) no excess survival as compared to the birth cohort and general population was observed for F3 controls, spouses of cases or controls and neither for F3 cases with up to 20% long-lived ancestors. The analyses in the HSN case/control dataset provides strong evidence that longevity is transmitted for at least 2 subsequent generations and only when at least 20% of all ancestors are long-lived. Moreover, the family cases seem to be genetically enriched for longevity while the controls resemble the general population. Finally, 27% of the F3 descendants showed a survival pattern similar to the general population even though they had at least one long-lived parent.

Previous family studies, usually focusing on 2 generations and single individuals, showed that siblings and children of long-lived persons lived longer than first degree ancestors of non-long-lived persons or population controls[5,6,45,49,7–12,15,17]. This knowledge about the familial clustering of longevity was utilized to construct longevity ranking scores such as the Family Mortality History Score (FMHS)[43], the est(SE) which subsequently was developed into the FLOSS[39,44], the Longevity Family Score (LFS) which is an adaptation to the est(SE) and the FMHS[12], and finally a method was developed to rank individuals by the survival of their ancestors, the Familial Excess Longevity (FEL) score[38]. The FMHS, FLOSS, and LFS all resemble excess survival of a family (FMHS focus on parents and FLOSS and LFS focus on siblings) compared to the general population. The FEL score focuses on excess survival, defined as the difference between a person's attained and expected age, derived from an accelerated failure time model. This excess survival was estimated for ancestors and from this a score was created for individuals. Although these scores all resemble a continuous familial estimate of a lifespan advantage and not necessarily longevity, they might be used as an inclusion tool for cases in genetic (association) studies[39]. However, these scores are not based on a clear longevity definition that represents the heritable longevity trait and they always require an arbitrary and difficult to interpret decision to make a cutoff in the scores so that they resemble longevity. In addition, the majority of the scores are not based on ancestors and thus do not capture the full family history of longevity. As such, the scores are not suitable to establish the proportion of family members that should be long-lived in order to properly define long-lived cases with a heritable longevity trait and thus, increase the power of genetic longevity studies.

To overcome these issues, we developed a novel tool based on mapping the longevity of a person's ancestors, the LRC score. The LRC score can be used to select carriers of the heritable longevity trait (cases) and controls who resemble the general population. Another interesting group, which we did not address in this article, is composed of persons without any long-lived ancestors who themselves are long-lived. It may be interesting to study environmental factors contributing to a long and healthy life in this group. Here we used the LRC score to construct a novel family case and control group and observed

a survival advantage for F3 case descendants, even when their parents were not necessarily long-lived, supporting the idea that a beneficial genetic component was transmitted. Likewise, the increase in the LRC score ≥ 20% associated with an increase in survival advantage for F3 descendants. This indicates that every additional ancestor contributes to the survival advantage of F3 descendants and confirms our previous findings in the LINKing System for historical demography (LINKS) data and the Utah Population Database (UPDB)[13]. This additive pattern is not readily expected if the observations are due to non-genetic factors, such as wealth, that cluster in families. The fact that none of the environmental confounders (sex, birth year, and sibship size) affected the survival differences between the family cases and controls provided additional evidence for the transmission of a genetic component. A final indication for the genetic enrichment of the family cases is based on the observed mortality pattern for the spouses of the family cases and controls which resembled the family controls themselves and the general population.

We observed that F3 descendants with at least one long-lived parent had less excess survival than a subset of these F3 descendants who had at least 30% long-lived ancestors and this difference increased when at least 50% of their ancestors were long-lived. These results indicate that some parents were long-lived but might not have transmitted their longevity to the subsequent F3 generation. In fact, 27% of the F3 descendants with at least one long-lived parent did not have an LRC ≥0.20 and, as a group, did not express excess survival. Hence the parents of theses 27% F3 descendants were sporadically long-lived as they did not transmit their longevity. Thus, genetic studies may benefit from a case definition, where cases are long-lived and have at least 30% long-lived ancestors, as current genetic studies, based on long-lived cases, often not include ancestral longevity in their case selection. Even though our data did not allow for an exact misclassification analysis, studies showed that the level of phenotypic misclassification in case and control annotation has a strong inhibiting effect on the power to identify variants in genetic association studies, including GWAS[42,50–58]. Moreover, it was shown that the power to identify genetic variants decreases at an equal rate to the level of misclassification[42]. For example, a study with 95% power to detect an association based on a sample of 100 cases and controls when there are no phenotypic errors may actually have only 75% power when 20% of the cases are misclassified as controls and vice versa42. Interestingly, when known, methods exist to adjust for the level of phenotypic misclassification[51–53,55,59], providing opportunities for specific application in genetic longevity research.

Due to the nature of the HSN data we could not use the mortality data for the parents (F0), siblings (F1), and spouses (F1) of the F1 IPs. Mortality data was less incomplete for the F2 and F3 spouses *(Table 1A)* but there was still a relatively large number of missing mortality data. Thus, for future studies with this dataset it might be interesting to extend the mortality information for these groups. Furthermore, life course data was only present for persons with an identified personal card or personal list (details in the methods section). Consequently, socio-economic status and religion was only available for a small part (around 15%) of the F3 descendants with an unequal share of availability between men and women. This led to the exclusion of these environmental factors from our analyses. Even though we could not adjust our models for socio-economic status and religion, it is known from other studies that those factors are not influencing the association between parental longevity and offspring survival[13]. Similarly, previous studies showed only a minor[60] or no[13,61] influence of early and mid-life environmental covariates, such as farm ownership, parental literacy, parental and own occupation, and birth intervals, on the association between parental longevity and offspring survival. We, however, cannot completely rule

out that other, unobserved non-genetic familial effects may affect our results. The observed excess survival of F2 case and control group spouses in the original approach seem to be an exception, as we observed a survival advantage for both groups. This is likely a form of ascertainment bias because mortality data for this group was difficult to obtain in the Dutch Personal Records Database, leading to an overrepresentation of high ages at death. These observations add to the mixed results about whether spouses married to a long-lived person have a survival advantage themselves[8,12–15,62].

Our results have two important implications. First, existing studies based on living study participants who have not yet reached the ages to express longevity, but have ancestral survival data, such as UK Biobank, can now better distinguish cases by incorporating a liability based on the LRC score. Second, new studies would obtain a maximum power to identify loci that promote survival to the highest ages in the population when cases are included with at least 30% (LRC≥0.30) ancestors who belong at least to the top 10% survivors of their birth cohort and are themselves among the 10% longest lived. More extreme selections can be made on the survival percentile by for example focusing on the top 5% or 1% survivors, and/or on the proportion of long-lived family members, for example 50%. However, this is not strictly necessary and might unnecessarily lead to limited sample sizes[13]. In addition, controls without any ancestors living to the top 10% survivors of their birth cohort should be included, as their mortality pattern resembles that of the general population. Finally, for future research it may be interesting to study the environmental factors causing the longevity in those individuals who were long-lived but had no long-lived ancestors. If our proposed method is consistently applied across studies, the comparative nature of longevity studies may improve and facilitate the discovery of novel genetic variants.

Methods

Historical Sample of the Netherlands

The Historical Sample of the Netherlands (HSN) Dataset Life Courses, Release 2010.01 is based on a sample of birth certificates and contains complete life course information for 37,137 Dutch individuals (index persons (IPs)) born in and between 1850 and 1922[46–48]. These 37,137 persons were subsequently identified in the Dutch population registers and followed in the registers throughout their entire life course[47,48,63]. The database includes information about the IPs' household, including their siblings, parents, and children, occupation at several points in time and religion. Households were only followed as long as the IP was present in that household meaning that information on kin was only partly covered[48,63]. For this study we selected 884 IPs who died at 80 years or beyond (case group) and 442 IPs who died between 40 and 59 years (control group), representing 1,326 disjoint families. IPs from both groups were born between 1860 and 1875. The case group was defined so that we would obtain a sample with overrepresentation of long-lived individuals. This was interesting since it would potentially allow to select on more extreme ages at death and still guarantee numbers reasonably large. The control group was selected to represent the mortality pattern of the general population of that time as best as possible. Individuals from both groups were selected to have an available date of birth, date of death, and at least one child should be identified. In conclusion, we identified 1,326 IPs (cases and controls), their F0 parents (N=2,652), F1 siblings (N=5,179), F2 descendants (N=7,404) and F1 spouses (N=1,409), covering 3 filial generations (F0 - F2) spanning from 1788 to 1941 *(Figure 1A and Table 1)*. The underlying data for this specific study were released as Kees Mandemakers and Cor Munnik, Historical Sample of the Netherlands. Project Genes, Germs and Resources. Dataset LongLives. Release 2016.01.

Extending the HSN study

For this study we extended the pedigrees until we identified the living descendants for all 1326 families. From the population registers we know the names of all F2 descendants and we subsequently identified the F2 descendants on personal cards (PCs) and personal lists (PLs) which were obtained from the Dutch central bureau of genealogy (CBG). These PLs and PCs were respectively introduced in 1939 and 1994 as the individualized and subsequently, digitized form of the population register[48]. The cards contain similar information to the population registers and because of privacy legislation could only be obtained for deceased persons, one year after they passed away (https://cbg.nl/bronnen/cbg-verzamelingen/persoons kaarten-en-lijsten). Hence, from these cards we obtained similar life course and mortality information for the F2 descendants as for the F1 IPs and we obtained the names of their descendants (F3). We repeated this procedure until no cards could be obtained anymore, which was at the F3 generation. Thus the F4 generation was not identified on the PCs of PLs anymore. In conclusion, we identified and obtained information for the F2 descendants, F2 spouses, F3 descendants, F3 spouses, and F4 descendants *(Figure1A and Table1)*. We will refer to this database as the HSN case/control database.

Obtaining information for the living descendants

In a final step we obtained as much mortality information as possible for the relatives of the identified persons and we obtained addresses, as contact information for the living descendants. This information was obtained through the Personal Records Database (PRD) which is managed by Dutch governmental service for identity information. *https://www.government.nl/topics/personal-data/personal-records-database-brp*

The PRD contains PL information on all Dutch citizens (alive and death) and PC information is continuously added. We were granted permission (permission number: 2016-0000364875) to obtain the date of death, date of last observation, current living address, and identifying information such as names of a person's father and mother to double check if the person identified in the PRD was identical to the person in our HSN case/control database. Using the PRD we were able to obtain addresses for F3 and F4 descendants and additional mortality information for F2 descendants, F2 spouses, F3 descendants, F3 spouses, and F4 descendants *(Figure1A and Table1)*. The final database covers 57,337 persons from 1,326 five-generational families (F0-F4) and contains F1 index persons (IPs), their parents (F0), siblings (F1), spouses (F1), and 3 consecutive generations of descendants (F2-F4) and spouses (F2-F4), connecting deceased persons to their living descendants.

Exclusion criteria and study population

Due to the nature of the source data there is a high rate of missing mortality information for F0 parents, F1 spouses and F1 siblings, which we therefore excluded from analyses. We further excluded F4 descendants because 92% is still alive *(Table 1 and Figure 1B)*. The final study population covers 37,825 persons from 1,326 three-generational families (F1-F3) and contains F1 index persons (IPs), 2 consecutive generations of descendants (F2-F3) and 2 generations of spouses (F2-F3).

Statistical analyses

Statistical analyses were conducted using R version 3.4.1 [64]. We reported 95% confidence intervals (CIs) and considered p-values statistically significant at the 5% level ($\alpha = 0.05$).

Lifetables

In the Netherlands, population based cohort lifetables are available from 1850 until 2019 [65,66]. These lifetables contain, for each birth year and sex, an estimate of the hazard of dying between ages x and x + n (hx) based on yearly intervals (n=1) up to 99 years of age. Conditional cumulative hazards (Hx) and survival probabilities (Sx) can be derived using these hazards. In turn, we can determine to which sex and birth year based survival percentile each person of our study belonged to. For example: a person was born in 1876, was a female, and died at age 92. According to the lifetable information this person belonged to the top three percent survivors of her birth cohort, meaning that only three percent of the women born in 1876 reached a higher age. We used the lifetables to calculate the birth cohort and sex specific survival percentiles for all persons in the HSN case/control study. This approach prevents against the effects of secular mortality trends over the last centuries and enables comparisons across study populations[1,11]. *Supplementary Figure 6* shows the ages at death corresponding to the top 10, 5, and 1 percent survivors of their birth cohorts for the period 1850-1935.

Standardized Mortality Ratios

To indicate excess mortality or excess survival of groups, such as F2 case or control group descendants in the HSN case/control study compared to Dutch birth cohort members of the same sex, we used Standardized Mortality Ratios (SMRs). An SMR is estimated by dividing the observed number of deaths by the expected number of deaths. The expected number of deaths are given by the sum of all individual cumulative hazards based on the birth cohort and sex specific lifetables of the Dutch population. An SMR between 1 and 0 indicates excess survival, an SMR of 1 indicates that the study population shows a similar survival to the reference population, and an SMR above 1 indicates excess mortality. The SMR can be estimated

conditional on the specific age at which an individual starts to be observed in the study (correction for left truncation). This was necessary to avoid selection bias if individuals in a study population were not at risk of dying before a specific age of entry.

$$SMR = \frac{observed\ number\ of\ deaths}{expected\ number\ of\ deaths} = \frac{\sum_{i=1}^{N} d_i}{\sum_{i=1}^{N} H_{t0i}(t_i|t_{0i})}$$

Where d_i=dead status (1=dead, 0=alive), H_{t0i}=sex and birth year specific cumulative hazard based on lifetable, t_i=timing, referring to age at death or last observation, t_{0i}=liftable age conditioning, for example from birth (t_{0i}=0), N= group sample size. Exact CIs were derived[67] and compared to bootstrap CIs for family data[12]. Both methods provided identical CIs and thus, to reduce the amount of computational time necessary to estimate bootstrap CIs, we estimated exact CIs.

Longevity Relatives Count score

Based on the results of a recent study which shows that longevity is heritable beyond the 10% survivors of their birth cohort and that multiple family members, such as parents and/or aunts and uncles, should belong to the top 10% survivors[13] we constructed a novel score that summarizes the familial history of longevity, the Longevity Relatives Count score (LRC).

$$LRC_i = \frac{\text{weighted number of top x percentile relatives}}{\text{weighted total number of relatives}} = \frac{\sum_{k=1}^{N_i} w_k \cdot I(P_k \geq 0.9)_i}{\sum_{k=1}^{N_i} w_k}$$

Where k=1,...,Ni are the available relatives of individual i used to build the score, P_k is the sex and birth year–specific survival percentile based on lifetables of relative k and $I(P_k \geq 0.9)$ indicates if relative k belongs to the top 10% survivors $\sum_{k=1}^{N_i} w_k$ is the weighted total number of relatives of person i. The relationship coefficients are used as weights w_k.. For example, persons share on average 50% of their nuclear DNA with their parents and siblings and this is 25% for aunts, uncles or grandparents. Hence, in the LRC, each parent and sibling contributes 0.5 to the score while each aunt, uncle or grandparent contributes only 0.25. This is consistent to a previous study of us, which shows that distant longevous relatives associate significantly, but less strong to a person's survival than a close long-lived relative[13]. The higher the score, the higher the familial aggregation level of longevity. For example, a score of 0.5 indicates that 50% of a person's relatives were long-lived. We utilized the LRC score to map the proportion of long-lived ancestors for all F3 descendants, select cases with the heritable longevity trait and controls resembling the general population, and compare the survival advantage of F3 descendants who had at least one long-lived parent to those who had at least 30% long-lived descendants. The LRC scores were based on all identified relatives of F3 descendants with sufficient data quality *(Supplementary Figure 4 and 5)*.

Survival analysis (Cox-type random effects regression model)

To investigate the extent of a survival difference between the family F3 case and control group we use a Cox-type random effects model:

$$\lambda(t_{ij}) = u_i \lambda_0(t_{ij}) \exp(\boldsymbol{\beta} \boldsymbol{Z}_{ij} + \boldsymbol{\gamma} \boldsymbol{X}_{ij})$$

where $\boldsymbol{t_{ij}}$ is the age at death for person *j* in family *i*. $\boldsymbol{\lambda_0}(\boldsymbol{t_{ij}})$ refers to the baseline hazard, which is left unspecified in a Cox-type model. $\boldsymbol{\beta}$ is the vector of regression coefficients for the main effects of interest ($\boldsymbol{Z}$). $\boldsymbol{\gamma}$ is a vector of regression

coefficients for the effects of covariates and possible confounders ($\boldsymbol{X}$). $u_i > 0$ refers to an unobserved random effect (frailty). In all Cox models we adjust for sibship size, birth year, and sex.

Code availability

The scripts containing the code for data pre-processing and data analyses can be freely downloaded at: *https://git.lumc.nl/molepi/PUBLIC/LRCscore*

Data availability

Currently all data is cleaned and we are constructing a data description file. As soon as the data description file is completed the data will be made freely available in a data repository.

Competing interests

The authors declare no competing interests.

References

1 van den Berg, N., Beekman, M., Smith, K. R., Janssens, A. & Slagboom, P. E. Historical demography and longevity genetics: Back to the future. *Ageing Res. Rev.* **38**, 28–39 (2017).

2 Herskind, A. M. *et al.* The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Hum. Genet.* **97**, 319–323 (1996).

3 Kaplanis, J. *et al.* Quantitative analysis of population-scale family trees with millions of relatives. *Science* **360**, 171–175 (2018).

4 Ruby, J. G. *et al.* Estimates of the Heritability of Human Longevity Are Substantially Inflated due to Assortative Mating. *Genetics* **210**, 1109–1124 (2018).

5 Houde, L., Tremblay, M. & Vézina, H. Intergenerational and Genealogical Approaches for the Study of Longevity in the Saguenay-Lac-St-Jean Population. *Hum. Nat.* **19**, 70–86 (2008).

6 Perls, T. T. *et al.* Life-long sustained mortality advantage of siblings of centenarians. *Proc. Natl. Acad. Sci.* **99**, 8442–8447 (2002).

7 Willcox, B. J., Willcox, D. C., He, Q., Curb, J. D. & Suzuki, M. Siblings of Okinawan centenarians share lifelong mortality advantages. *J. Gerontol. A. Biol. Sci. Med. Sci.* **61**, 345–54 (2006).

8 Jarry, V., Gagnon, A. & Bourbeau, R. Survival advantage of siblings and spouses of centenarians in 20th-century Quebec. *Can. Stud. Popul.* **39**, 67 (2013).

9 Dutta, A. *et al.* Longer lived parents: Protective associations with cancer incidence and overall mortality. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **68**, 1409–1418 (2013).

10 Gudmundsson, H., Gudbjartsson, D. F. & Kong, A. Inheritance of human longevity in Iceland. *Eur. J. Hum. Genet. EJHG* **8**, 743–749 (2000).

11 Sebastiani, P., Nussbaum, L., Andersen, S. L., Black, M. J. & Perls, T. T. Increasing Sibling Relative Risk of Survival to Older and Older Ages and the Importance of Precise Definitions of "Aging," "Life Span," and "Longevity". *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **71**, 340–346 (2016).

12 Berg, N. van den *et al.* Longevity Around the Turn of the 20th Century: Life-Long Sustained Survival Advantage for Parents of Today's Nonagenarians. *Journals Gerontol. Ser. A* **73**, 1295–1302 (2018).

13 van den Berg, N. *et al.* Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nat. Commun.* **10**, 35 (2019).

14 Pedersen, J. K. *et al.* The Survival of Spouses Marrying Into Longevity-Enriched Families. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **72**, 109–114 (2017).

15 Schoenmaker, M. *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur. J. Hum. Genet.* **14**, 79–84 (2006).

16 Ljungquist, B., Berg, S., Lanke, J., McClearn, G. E. & Pedersen, N. L. The Effect of Genetic Factors for Longevity: A Comparison of Identical and Fraternal Twins in the Swedish Twin Registry. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **53A**, 441–446 (1998).

17 Terry, D. F. *et al.* Lower All-Cause, Cardiovascular, and Cancer Mortality in Centenarians' Offspring. *J. Am. Geriatr. Soc.* **52**, 2074–2076 (2004).

18 Westendorp, R. G. J. *et al.* Nonagenarian Siblings and Their Offspring Display Lower Risk of Mortality and Morbidity than Sporadic Nonagenarians: The Leiden Longevity Study. *J. Am. Geriatr. Soc.* **57**, 1634–1637 (2009).

19 Christensen, K. *et al.* Physical and cognitive functioning of people older than 90 years: a comparison of two Danish cohorts born 10 years apart. *Lancet* **382**, 1507–1513 (2013).

20 Christensen, K., McGue, M., Petersen, I., Jeune, B. & Vaupel, J. W. Exceptional longevity does not result in excessive levels of disability. *Proc. Natl. Acad. Sci.* **105**, 13274–13279 (2008).

21 Andersen, S. L., Sebastiani, P., Dworkis, D. a., Feldman, L. & Perls, T. T. Health Span Approximates Life Span Among Many Supercentenarians: Compression of Morbidity at the Approximate Limit of Life Span. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **67A**, 395–405 (2012).

22 Shadyab, A. H. & LaCroix, A. Z. Genetic factors associated with longevity: A review of recent findings. *Ageing Res. Rev.* **19**, 1–7 (2015).

23 Eline Slagboom, P., van den Berg, N. & Deelen, J. Phenome and genome based studies into human ageing and longevity: An overview. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1864**, 2742–2751 (2018).

24 Deelen, J. *et al.* Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* **23**, 4420–4432 (2014).

25 Broer, L. *et al.* GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **70**, 110–118 (2015).

26 Willcox, B. J. *et al.* FOXO3A genotype is strongly associated with human longevity. *Proc. Natl. Acad. Sci.* **105**, 13987–13992 (2008).

27 Flachsbart, F. *et al.* Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc. Natl. Acad. Sci.* **106**, 2700–2705 (2009).

28 Pilling, L. C. *et al.* Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany. NY).* **9**, 2504–2520 (2017).

29 Joshi, P. K. *et al.* Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nat. Commun.* **8**, 1–13 (2017).

30 Partridge, L., Deelen, J. & Slagboom, P. E. Facing up to the global challenges of ageing. *Nature* **561**, 45–56 (2018).

31 Sebastiani, P. *et al.* Four Genome-Wide Association Studies Identify New Extreme Longevity Variants. *J. Gerontol. A. Biol. Sci. Med. Sci.* **72**, 1453–1464 (2017).

32 Deelen, J. *et al.* A meta-analysis of genome-wide association studies identifies multiple longevity genes. *Nat. Commun.* **10**, 3669 (2019).

33 Flachsbart, F. *et al.* Immunochip analysis identifies association of the RAD50/IL13 region with human longevity. *Aging Cell* **15**, 585–588 (2016).

34 Zeng, Y. *et al.* Novel loci and pathways significantly associated with longevity. *Sci. Rep.* **6**, 21243 (2016).

35 Oeppen, J. & Vaupel, J. W. Broken limits to life expectancy. *Science* **296**, 1029–1031 (2002).

36 Vaupel, J. W. Biodemographic Trajectories of Longevity. *Science (80-.).* **280**, 855–860 (1998).

37 Sebastiani, P. & Perls, T. T. The Genetics of Extreme Longevity: Lessons from the New England Centenarian Study. *Front. Genet.* **3**, 1–7 (2012).

38 Kerber, R. A., Brien, E. O., Smith, K. R. & Cawthon, R. M. Familial Excess Longevity in Utah Genealogies. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **56**, 130–139 (2001).

39 Sebastiani, P. *et al.* A Family Longevity Selection Score: Ranking Sibships by Their Longevity, Size, and Availability for Study. *Am. J. Epidemiol.* **170**, 1555–1562 (2009).

40 Hujoel, M. L. A., Gazal, S., Loh, P., Patterson, N. & Alkes, L. Combining case-control status and family history of disease increases association power. 1–20 (2019).

41 Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case–control association mapping by proxy using family history of disease. *Nat. Genet.* **49**, 325–331 (2017).

42 Gordon, D. Factors affecting statistical power in the detection of genetic association. *J. Clin. Invest.* **115**, 1408–1418 (2005).

43 Rozing, M. P. *et al.* Familial longevity is associated with decreased thyroid function. *J. Clin. Endocrinol. Metab.* **95**, 4979–4984 (2010).

44 Arbeeva, L. S. *et al.* How Well Does the Family Longevity Selection Score Work: A Validation Test Using the Utah Population Database. *Front. Public Heal.* **6**, (2018).

45 Deluty, J. A., Atzmon, G., Crandall, J., Barzilai, N. & Milman, S. The influence of gender on inheritance of exceptional longevity. *Aging (Albany. NY).* **7**, 412–418 (2015).

46 Mandemakers, K. Historical sample of the Netherlands. In P. K. Hall, R. McCaa, & G. Thorvaldsen (Eds.). in *Handbook of International Historical Microdata for Population Research* 149–177 (2000).

47 Mandemakers, K. https://socialhistory.org/en/hsn/hsn-releases. *HSN 2010.01 release* (2010).

48 van den Berg, N. *et al.* Families in Comparison: An individual-level comparison of life course and family reconstructions between population and vital event registers. *SocArXiv* (2018). doi:10.31235/osf.io/h2w8t

49 Kemkes-Grottenthaler, A. Parental effects on offspring longevity—evidence from 17th to 19th century reproductive histories. *Ann. Hum. Biol.* **31**, 139–158 (2004).

50 Buyske, S., Yang, G., Matise, T. C. & Gordon, D. When a Case Is Not a Case: Effects of Phenotype Misclassification on Power and Sample Size Requirements for the Transmission Disequilibrium Test with Affected Child Trios. *Hum. Hered.* **67**, 287–292 (2009).

51 Gordon, D., Haynes, C., Yang, Y., Kramer, P. L. & Finch, S. J. Linear trend tests for case–control genetic association that incorporate random phenotype and genotype misclassification error. *Genet. Epidemiol.* **31**, 853–870 (2007).

52 Barral, S., Haynes, C., Stone, M. & Gordon, D. LRTae: improving statistical power for genetic association with case/control data when phenotype and/or genotype misclassification errors are present. *BMC Genet.* **7**, 24 (2006).

53 Rekaya, R., Smith, S., Hay, E. H., Farhat, N. & Aggrey, S. Analysis of binary responses with outcome-specific misclassification probability in genome-wide association studies. *Appl. Clin. Genet.* **Volume 9**, 169–177 (2016).

54 Ji, F., Yang, Y., Haynes, C., Finch, S. J. & Gordon, D. Computing Asymptotic Power and Sample Size for Case-Control Genetic Association Studies in the Presence of Phenotype and/or Genotype Misclassification Errors. *Stat. Appl. Genet. Mol. Biol.* **4**, (2006).

55 Edwards, B. J., Haynes, C., Levenstien, M. A., Finch, S. J. & Gordon, D. Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. *BMC Genet.* **6**, 18 (2005).

56 Royall, D. R., Chiodo, L. K. & Polk, M. J. Misclassification Is Likely in the Assessment of Mild Cognitive Impairment. *Neuroepidemiology* **23**, 185–191 (2004).

57 Bross, I. Misclassification in 2 X 2 Tables. *Biometrics* **10**, 478 (1954).

58 Platz, E. A., De Marzo, A. M. & Giovannucci, E. Prostate cancer association studies: Pitfalls and solutions to cancer misclassification in the PSA era. *J. Cell. Biochem.* **91**, 553–571 (2004).

59 Smith, S., Hay, E., Farhat, N. & Rekaya, R. Genome wide association studies in presence of misclassified binary responses. *BMC Genet.* **14**, 124 (2013).

60 Danzhen You, Danan Gu & Zeng Yi. Familial Transmission of Human Longevity Among the Oldest-Old in China. *J. Appl. Gerontol.* **29**, 308–332 (2010).

61 Gavrilov, L. A. & Gavrilova, N. S. Predictors of Exceptional Longevity: Effects of Early-Life and Midlife Conditions, and Familial Longevity. *North Am. Actuar. J.* **19**, 174–186 (2015).

62 Montesanto, A. *et al.* The genetic component of human longevity: analysis of the survival advantage of parents and siblings of Italian nonagenarians. *Eur. J. Hum. Genet.* **19**, 882–886 (2011).

63 Mandemakers, K. Building life courses datasets from population registers by the historical sample of the Netherlands. in *History and Computing* **14**, 87–107 (2002).

64 R Core Team. R: A language and environment for statistical computing. (2016).

65 Van Der Meulen, *A. Life tables and Survival analysis*. (2012).

66 Carolina, T., Uijvenhoven, L. & van der Laan, J. Overlevingstafels en longitudinale analyse. CBS 1–25 (2009).

67 Ulm, K. Simple method to calculate the confidence interal of a standardized mortality ratio (SMR). *Am. J. Epidemiol.* **131**, 373–375 (1990).

CHAPTER 7

GENERAL DISCUSSION

The complexity of longevity as a multifactorial trait

Identifying longevity loci is important because these loci likely represent key mechanisms of a life-long decreased mortality[1,2], decreased morbidity[3–5] and compression of morbidity towards the end of life[6–8]. However, the identification of longevity loci has been challenging and only a handful of genetic variants have been shown to associate with longevity across multiple independent studies[9–16]. In fact, genome-wide linkage and association studies identified only a few robust loci promoting longevity[9–17]. The most compelling evidence was obtained for alleles in the APOE and FOXO3A genes as they have been consistently identified with either genome-wide association studies (GWAS) or candidate gene studies[10–15,18].

One of the main reasons for the limited success of genetic longevity studies[12–14,18–20] is the uncertainty in defining the heritable longevity trait itself[2,21]. Given the increased life expectancy of the past 200 years due to non-genetic/social factors[22,23] (improved hygiene, nutrition and medication) there are likely many phenocopies among the long-lived cases selected for our genetic studies. In this thesis we showed that the solution may lie in the familial clustering of longevity and that the inclusion of persons with the heritable longevity trait (the persons with a high familial clustering of longevity) may provide a fruitful basis for future longevity research.

It is important to consider the strong increase in human life expectancy when investigating familial longevity using multigenerational genealogical data, as multiple generations of long-lived individuals experienced significant developments in the knowledge of good hygiene, healthy behavior and health care over the past 200 years6,[22–34] (this is known as the epidemiological transition).

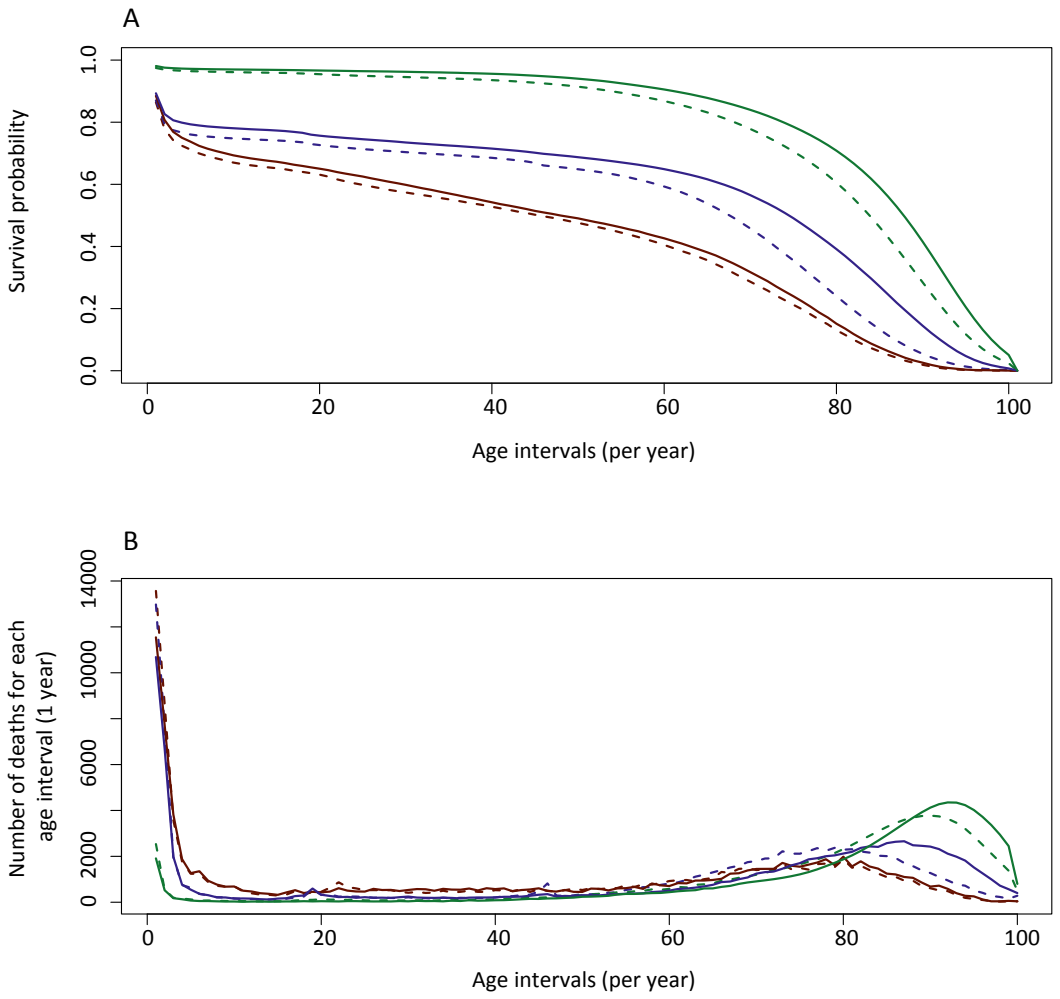


Figure 1: Increase in human lifespan between 1850 and 1950.

Panel A shows the average increase in human survival from 1850 onward in the Netherlands. Panel B shows the reduction in infant and childhood mortality and the subsequent increase in mid and later life survival. The brown lines represent the 1850 birth cohort, the blue lines represent the 1900 birth cohort, the green lines represent the 1950. The 1850 and 1900 cohorts are based on the CBS cohort lifetables and the 1950 cohort is based on the prognostic CBS cohort lifetables. The straight lines represent females and the dotted lines represent males.

Due to better knowledge of hygiene and increased food availability people were able to better cope with malnutrition of children, infections with cholera, diarrheal diseases and tuberculosis and as a result, childhood mortality decreased significantly. Next, the increased understanding of (preventive) medicine and medical health care, further improvements in hygiene, and the introduction of public sanitation caused a reduction in mid and later-life mortality[6,22–34]. Besides these developments, multiple generations of long-lived individuals experienced various important events, such as the potato blight of 1847, various outbreaks of smallpox and measles, the first world war, the almost coinciding Spanish flue of 1918, and the second world war[25,35–38]. The increase in human lifespan is illustrated in *Figure 1*. Panel A visualizes the average increase in human survival from 1850 onward in the Netherlands. It also clearly shows that the survival difference between men and

women increased over time (from a two year difference to a 6 year difference) which is speculated to be caused by more unhealthy behavior among men[39–41]. Panel B illustrates the reduction in infant and childhood mortality and the subsequent increase in mid and later life survival.

As all these changes caused a rapid increase in the human life expectancy, it can be expected that many individual non-genetic/social factors contribute to the human survival and longevity. In fact, contemporary research identified many individual factors that associate with lifespan and longevity. Among others, Socio-economic status in terms of income and occupation, educational attainment, sibship size, birth order, but also personality traits, and susceptibility to diseases such as cardio-vascular disease, cancer, and Alzheimer's disease are associated to individual survival and longevity[42–50,42,44,48,51]. These factors are likely to cluster in families as parents and their children share available resources and live in the same physical environment. In fact, research showed that children may inherit personality traits from their parents, either through genetic predispositions or socialization, but also for example their parents' socio-economic status[52–57]. It is however unknown to what extent these factors contribute to the clustering of longevity in members of long-lived families.

The aim of this thesis was to study the familial component of longevity by first establishing a standardized definition of heritable longevity and subsequently investigating its intergenerational transmission characteristics, its life course influence on survival and the interrelation with other, environmental or familial factors. We ultimately aimed to establish a genetically enriched group (cases) and a group which represents the general population (controls) who could be included in future genetic longevity studies.

We used unselected three-generational demographic and mortality data from the Utah Population Database (UPDB, US) and the LINKing System for historical family reconstruction (LINKS, Netherlands) to investigate which survival percentile best isolates the genetic component of longevity. We subsequently determined the importance of long-lived family members for case selection so that those insights can be used in genetic studies to identify novel longevity loci. We applied our knowledge about survival percentiles to the Leiden Longevity Study (LLS) where we investigated 1. a potential sex-specific inheritance pattern of longevity, 2. a potential survival advantage of long-lived sibships as compared to long-lived singletons, and 3. whether the parents of these sibships had a life-long sustained survival advantage. Next, we used the Historical Sample of the Netherlands (HSN) case/control study to establish how many family members of a potential long-lived case should be long-lived in order to avoid phenocopies. Finally, we utilized the insights to identify in the HSN data a novel case and control group for future genetic studies in which we connect deceased ancestors to living family members and compared these to a group of sporadically long-lived persons. Besides these main aims we discuss the survival of spouses marrying into long-lived families, the effect of environmental/social factors on individual as well as familial longevity, the interplay between longevity and family size, and sex specific transmission patterns observed in some of the data.

Discussion of our findings

Main results

In **chapter 2** we reviewed the relevant studies investigating the familial component of longevity. We focused on heritability studies, studies investigating the transmission of lifespan and longevity as well as lifespan and longevity inheritance patterns. We further discussed important environmental/social covariates that affect individual lifespan and longevity and/or potentially affect the transmission of lifespan and longevity between parents and offspring. We emphasized the importance of distinguishing between lifespan and longevity because currently, longevity is often confused with lifespan. Lifespan generally refers to the age at death of a person whereas longevity refers to survival into extreme ages beyond an arbitrarily chosen threshold, such as 80, 90, 100 years, or an extreme survival percentile such as belonging to the top 10%, 5%, of 1% birth cohort specific survivors. We concluded that heritability estimates for lifespan vary slightly but consistently indicate that 12-25% of the variance in lifespan is due to additive genetic effects. The large impact of environmental factors on the average lifespan likely surpasses that of genetic ones and a large range of factors determine early death. In contrast to the number of studies into the heritability of lifespan, studies into the heritability of longevity are scarce and report inconsistent heritability results. Moreover, there are indications that the heritability increases with a more strict cutoff of lifespan towards more extreme ages of survival[1,58–66]. Studies focusing on the transmission pattern of lifespan and longevity are both inconsistent but, for lifespan studies there are indications of a female transmission pattern. We conclude that environmental/social factors, such as socio-economic status, sibship size, maternal age at first and last birth and birth order should be taken into account when investigating the familial component of lifespan and potentially also for longevity. We further conclude that novel research is needed to estimate the heritability of longevity and establish a longevity transmission pattern. Because there is no consensus of how longevity should be defined, we first discuss a strategy to identify a definition of longevity that best represents the heritable component of the trait.

To investigate the familial component of longevity in large genealogical data in the absence of study related selections such as the inclusion of alive persons who survived into old age, we constructed the LINKS data together with the International Institute of Social History (IISH) and the Radboud University (RU). Because of the novel character of the LINKS data, we first set out to validate the life course and family reconstruction quality by comparing the LINKS data with the already existing HSN data. Thus, in **chapter 3**, we compared indicators of fertility, marriage, mortality, and measurements of occupational status of ~400 individuals identified in both databases and concluded that life course and family reconstructions in the HSN and LINKS reflect each other well. As we expected, LINKS provides more complete family information on siblings and parents, whereas the HSN provides more complete life course information, especially for individuals who migrated out of Zeeland. We also observed that the number of children was very similar between the ~400 persons identified in both the HSN an LINKS. This coincides with the very complete life course information in the HSN which accurately captures the births of children. We conclude that life course and family reconstructions based on linked, fragmented observations, such as in LINKS, on individuals constitute a reliable alternative to such reconstructions based on continuous observations from population or parish registers.

After verifying the quality of the LINKS data, we continued to investigate the definition of heritable longevity and the familial clustering of longevity using the LINKS data and the Utah Population Database (UPDB), which combined represent the largest genealogical database with verified (mortality) information in the world. In **chapter 4** we used three-generational mortality data from the UPDB and LINKS, and studied 20,360 families who were unselected for mortality. We focused on 20,360 index persons, their parents (N=40,72), siblings (N=108,122), spouses (N=22,018), and children (N=123,599), comprising a total of 314,819 individuals. We investigated which survival percentile best isolates the heritable component of longevity and we subsequently determined the importance of long-lived family members for case selection so that those insights can be used in genetic studies to identify novel longevity loci. We further studied the non-genetic/social factors, such as socio-economic status, religious denomination, number of children, birth order, and birth cohort, that may explain the intergenerational transmission of longevity. Moreover, we explored the survival of spouses marrying into longevity enriched families as an indicator for shared resources, lifestyles, and potentially socio-economic status during middle and late-life as explaining factors for the familial component of longevity. In addition, it is important to note that we indirectly investigated social and living environmental influences on the familial component of longevity by comparing Utah and Zeeland. Utah and Zeeland distinctly differed in their physical environment, living conditions, and subsequent mortality patterns. Our analyses provided strong evidence that longevity is transmitted as a quantitative genetic trait among survivors up to the top 10% of their birth cohort. We subsequently showed a survival advantage, mounting to 31%, for individuals with top 10% surviving first and second-degree relatives in both databases and across generations, even in the presence of non-long-lived parents. Further results showed that, among others, socio-economic status, sibship size, birth order did not affect the association between parental or sibling longevity and the survival of the index persons. Some factors, such as socio-economic status, birth year, and religious denomination did affect the individual survival of the index person, but as mentioned, independently of the parental and sibling effects. No evidence was observed that spouses marrying into a longevity enriched family also showed a survival benefit. This will be discussed in more detail further on. Interestingly, the Hazard Ratios, reflecting the survival benefit of index persons with 1 or 2 compared to 0 long-lived parents or siblings were remarkably similar between the UPDB and LINKS. This similarity provides a strong indication that the familial component of longevity is very limitedly affected by effects of the physical environment and for example migration patterns. Finally, to guide future genetic studies, we suggest to base case selection on top 10% survivors of their birth cohort with equally long-lived first and second-degree family members.

In **chapter 5**, we applied the new survival percentile threshold based longevity definition to the Leiden Longevity Study (LLS) where we studied the 944 participating long-lived siblings and their relatives to investigate 1. a potential sex-specific inheritance pattern of longevity, 2. a potential survival advantage of long-lived sibships as compared to long-lived singletons and 3. whether the parents of these siblings had a life-long sustained survival advantage. Family longevity scores were estimated to explore whether human longevity is transmitted preferentially through the maternal or paternal line. Standardized mortality ratio's (SMRs) were estimated to investigate whether long-lived siblings have a survival advantage compared to long-lived singletons and we investigated if parents of long-lived siblings harbor a life-long sustained survival advantage compared to the general Dutch population by estimating lifetime SMRs (L-SMRs). We observed that sibships with long-lived mothers and non-long-lived fathers had 0.41 (P=0.024) less observed deaths than sibships with long-lived fathers and non-long-lived mothers and 0.48 (P=0.008) less observed deaths than sibships with both parents non-long-

lived. Participants had 18.6% less deaths compared to matched singletons and parents had a life-long sustained survival advantage (L-SMR=0.510 and 0.688). In conclusion, genetic longevity studies may incorporate the testing of a maternal transmission pattern (further discussed later on) and potential genes involved appeared to beneficially influence the entire life-course of individuals.

In **chapter 4** we addressed the issue of the uncertainty in defining the longevity trait itself and observed that the survival percentile threshold that best reflects the genetic component of longevity is at the top 10% survivors of their birth cohort and beyond. Moreover, we investigated the familial component of longevity and observed that the survival advantage of family members increased with each additional long-lived family member. In **chapter 6** we followed-up on the longevity definition as established in chapter 4. We investigated if longevity is transmitted for multiple generations and whether the longevity effect diminishes over generations. We did this by comparing long-lived cases (died ≥ 80 years) and their descendants to population resembling controls (died between 40 and 59 years) and their descendants. Furthermore, we developed the Longevity Relatives Count (LRC) score to establish how many family members should be long-lived in order to avoid phenocopies. We subsequently investigated how often long-lived parents from a long-lived family pass on their longevity to their children compared to long-lived parents from general population families.

Our analyses included 37,825 persons from 1,326 three-generational families in the HSN case/control study. The analyses in the HSN case/control dataset provide strong evidence that longevity is transmitted for at least 2 subsequent generations if at least 20% of all relatives are long-lived, but preferably 30%. Moreover, the family based cases seem to be at least partially genetically enriched for longevity, as birth year, sibship size, and sex did not affect the transmission of longevity. The evidence for genetic enrichment is strengthened by the fact that their spouses resembled the family based controls as well as the general population in their average survival. Moreover, other studies, as outlined in **chapter 4**, did not obtain evidence that other non-genetic factors, such as religion and socio-economic status could explain the familial component of longevity. Finally, 27% of the F3 descendants showed a survival pattern similar to the general population even though they had at least one long-lived parent. Hence the parents of theses 27% F3 descendants were sporadically long-lived as they did not transmit their longevity. In summary, to select individuals that are enriched for the heritable longevity trait, case should be selected on the basis of being long-lived themselves and having at least 30% long-lived ancestors.

We now have a much more clear understanding about the familial component of longevity. Most importantly, we know that longevity is transmitted as a quantitative genetic trait among survivors up to the top 10% of their birth cohort as long-lived blood relatives independently and additively contribute to the survival advantage of index persons. Long-lived study participants with a family history of longevity have a lifelong sustained survival advantage and their spouses seem to resemble the general population. In line with other studies, we showed that the association between parental longevity and the survival of their offspring is not affected by non-genetic/social factors such as socio-economic status and sibship size. In addition, by using the LRC sore we determined that at least 30% of an individual's relatives should be within the top 10% survivors of their birth cohort before a survival advantage is observed. Moreover, individuals without long-lived relatives represent the general population and may thus be considered as phenocopies even when they do become long-lived. These results provide novel opportunities for future research into longevity.

Secular trends in longevity research

As mentioned earlier, when investigating multiple generations of long-lived persons it is important to take into account that these persons experienced important developments in knowledge about hygiene, healthy lifestyle, medical health care and related technological advancements. It is also important to acknowledge that there were epidemic periods and periods of war which affected mortality. In addition, the life expectancy difference between men and women increased over the last 200 years. In demographic research, these changes over time are known as secular (mortality) trends. A consequence of these secular trends, for longevity research is for example that being 90 years old nowadays is not nearly as special as it was around 1800. Hence, to investigate familial longevity with data spanning more than 200 years, it is important to take these secular trends into account.

An approach to incorporate these secular trends in statistical models is to standardize the measurements for mortality, e.g. age at death, of study participants to that of their birth cohort members who experienced the same developments and epidemic hazards during their life. We used cohort lifetables to calculate birth cohort and sex specific survival percentiles (for example, belonging to the top 10%, 5%, or 1% survivors) so that the mortality of a person is measured relative to his or her birth cohort members who experienced the same secular trends. This approach requires a reference population on which the lifetables are based and has as a main advantage that the survival percentiles are calculated in exactly the same way every time. This ensures a fair comparison between study participants of different birth cohorts and different study populations. For our research, we used both Dutch and Swedish lifetables which are consistent with the lifetables of multiple industrialized societies[67]. Some alternative approaches have been developed[1,59,68–70] which come down to regressing out study population specific environmental effects (that reflect secular trends as good as possible, for example a person's birth cohort) associated with individual mortality and analyzing the residuals as a measure of mortality. There are some clear drawbacks of these study population specific residual methods compared to the lifetable standardization method we used. The residual methods account only for the environmental factors that are included in the statistical model and are thus dependent on what is measured in a specific study and the arbitrary choices of a researcher on how to model these factors. Moreover, because the included environmental factors are specific for a study and depend on what is measured in a study, comparative research between different populations is difficult. The lifetable method provides an advantage because by comparing to birth cohort members it is possible to adjust for all secular trends over time that are shared by members of a specific birth cohort. This includes many environmental factors that are usually not observed in specific studies, for example, relating to the improved living conditions or health care system over time. The lifetable model also allows for fair comparisons across study populations, as illustrated in **chapter 4**. One drawback of comparing different study populations, however, is that the environment factors accounted for in the reference population are unlikely exactly the same in the different study populations. Taking the pros and cons of the different methods into account, we prefer the lifetable method over the residual method.

Spouses seem to resemble the general population

In **chapter 3, 4, and 5** we investigated the survival of spouses marrying with long-lived persons or into longevity enriched families. All our results indicated that the spouses had a survival pattern equal to the general population, except for the spouses investigated in the LINKS study. The LINKS data contains an overrepresentation of persons who stayed in

Zeeland because we could not identify those who migrated to another province or abroad, as is described in **chapter 3**. Moreover, Zeeland had a low number of outmigration[71] which is described in **chapter 4**. The combination between the overrepresentation of stayers and the general pattern of low outmigration potentially caused high levels of relatedness among study participants in the LINKS data and in fact, unpublished results based on the LINKS data confirm this. Hence, it is likely that in Zeeland, spouses and long-lived persons were often (distantly) related to each other and thus shared some of the genetic component contributing to longevity. As a result, the observed survival benefit of spouses marrying into a longevity enriched family in LINKS, is likely caused by their relatedness instead of a shared mid and later-life lifestyle. Other studies showed diverse results regarding a possible survival benefit for spouses of long-lived persons[1,61,65,72,73]. In the Long Life Family study, Pedersen et al. (2017) identified a survival benefit for spouses of longevous siblings. The authors compared the spouses to sex and birth cohort matched controls and suggest assortative mating as an explanation for the observed survival benefit of the spouses[65]. A Quebec study, focused on the spouses of 806 centenarians, also reported a survival benefit[72] and a study of Southern Italy demonstrated that male nonagenarians outlived their spouses, whereas this was not the case for female nonagenarians[73]. In the future, more research is necessary to find out whether spouses marrying to long-lived persons or into a longevity enriched family were already predisposed with a survival advantage, gained a survival advantage, or in fact, resemble the general population.

Sex specific inheritance pattern of longevity

We investigated sex specific longevity transmission effects. Such effects can be divided into 1. Longevity is transmitted stronger via the mother than the father or vice versa, 2. sons or daughters who are more susceptible to parental (either the mother or the father) transmission of longevity, or 3. a combination between the two, for example mothers could transmit longevity more frequently to daughters than to sons. In **chapter 5**, when investigating the LLS, we observed evidence for a maternal transmission pattern with equal distribution to sons and daughters. We discuss two possible mechanisms that may cause this maternal transmission pattern: 1. the transmission of a beneficial genetic component via mitochondrial inheritance, as the mitochondrial DNA is only inherited via mothers and mitochondria play a vital role in many metabolic processes which have been associated to aging and longevity, 2. It might be possible that these long-lived mothers had babies with a high birth weight. In the end of 1800 a high birth weight provided a significant advantage in coping with the often harsh environmental/social circumstances, such as food scarcity and epidemics.

We however did not find evidence for a maternal transmission pattern or a higher level of susceptibility for sons or daughters in the UPDB, LINKS (**chapter 3**), and the HSN case/control study (**chapter 6**). Previous studies focused mainly on a sex specific inheritance pattern of lifespan (**chapter 2**) and had mixed conclusions[74–76]. Several explanations for the mixed results between the LLS and the UPDB, LINKS, and HSN case/control study can be possible. A first explanation concerns the specific time period that we observed in the LLS (1875 - 1941, *Figure 1* of the introduction). The period between 1875 - 1941 was characterized by repeated epidemics such as the smallpox, the measles, the potato blight, and the Spanish flue, creating strong infant and childhood mortality peaks[25,35–38]. This first explanation might be strengthened by the fact that the LLS sample was not random and it might be that many persons were born in places with a high epidemic impact, revealing the maternal transmission effects. Alternatively, the measurement that we used to test for the sex specific transmission pattern in the LLS was binary. This meant that a person was defined as long-lived (top 1% survivor) or alternatively as non-long-lived

(not top 1% survivor). By defining our groups like this, we ignored the survival percentile distance between parents. Consider for example a set of children in family A. If their mother belonged to the top 1% survivors she was considered long-lived. If their father belonged to the top 2% survivors (and thus not to the top 1%) he was not considered long-lived. A father within the 98th percentile of his birth cohort is still very much able to transmit his longevity as we observed in **chapter 4**. The distance between both parents is in this example 1 percentage point. Now consider a set of children in family B. Here their mother also belonged to the top 1% survivors of her birth cohort. However, their father belonged to the top 60% survivors of his birth cohort. The distance here is 59 percentage points. It is not too difficult to imagine that the described distances between fathers and mothers could be unequally distributed in the LLS and may thus have driven the observed maternal transmission pattern. This example points to a more general methodological issue in the literature when analyzing sex specific transmission patterns in longevity, which is generally defined as a binary trait. For future research it would be interesting to use methodology that can incorporate the continuous distribution between the binary longevity cutoffs of parents. In fact, for the LLS we did some preliminary analyses by defining non-long-lived as belonging to the bottom 85% survivors. So far this has not lead to different conclusions about the observed maternal inheritance pattern in the LLS.

In **chapter 2** we addressed fertility measurements, such as the number of children a mother has, or maternal age at first and last birth. We used these measures to investigate their interplay with the intergenerational transmission of longevity. In all our data; the LINKS, UPDB, LLS, and HSN we observed that persons from a smaller sibship had a lower hazard of dying than individuals from a larger sibship. Nonetheless, sibship size did not affect the transmission of longevity to a subsequent generation. In addition, in the LLS we observed a better survival and smaller sibship sizes for children with a long-lived mother and a non-long-lived father than the other way around or for children without any long-lived parents. This implied that long-lived mothers had less children and that these children also lived longer. Replication of these findings in data from less selected study populations would be interesting. We however have not yet explored this exact relation in the other studies.

Identifying long-lived families

One of our main goals was to identify a group of families who were (genetically) enriched for longevity and identify their living descendants who are interesting to include in genetic research. We describe a strategy to identify descendants from such long-lived families using the HSN data in **chapter 6**. The HSN data is interesting to identify individuals from long-lived families because it connects living persons to their deceased ancestors. The HSN however, does not contain very broad pedigrees and misses mortality information for relevant groups in the first and second generation, such as parents, siblings and spouses. Moreover, mortality information is also incomplete for spouses in the other generations and no relatives are included from the spousal family lines. In other words, the pedigrees in the HSN have sufficient depth, in numbers of generations but are very narrow in terms of known relatives, especially for potential inclusion of families into a genetic longevity study.

The use of the Longevity Relatives Count score (LRC) described in **chapter 6** is related to the narrow pedigrees. Based on the HSN case/control study we could build and test the LRC score only for the proband line (compared to the spousal line) with a maximum of 2 generations of deceased relatives (*Figure 1* in **chapter 6**). The observed results look promising but

there is a need to test the score on both the proband and spousal family lines (relatives of both parents). In addition to this, we currently constructed the score based on top 10% surviving relatives and observed that 27% of the F3 descendants showed a survival pattern similar to the general population even though they had at least one long-lived parent. It remains to be investigated if similar levels will be obtained when building the LRC score based on top 5% or 1% survivors, although we expect this to be the case based on the results of **chapter 4**. In **chapter 4** we showed a survival advantage with each additional long-lived relative, when defining long-lived as belonging to either the top 10% and top 5% survivors. Finally, the LRC score would benefit from validation in an independent dataset such as the LINKS data. The LINKS data could be used to estimate a survival difference between individuals with an LRC score of 0 (controls) and those with an LRC score ≥ 0.30 (cases) but will be difficult to use for the identification of potential phenocopies due to the high level of relatedness for individuals in the database and extreme mortality, especially early in life. As a result, SMRs, which can used to estimate whether a group of individuals follow a mortality pattern similar to the general population, cannot be accurately estimated.

In **chapter 4, 5, and 6** we observed a strong familial clustering of longevity within specific families and used non-genetic/social covariates, pedigree information and, genetic assumptions to distinguish between potential genetic and environmental influences to the familial component within the long-lived families. We obtained evidence for potential genetic influences to the familial component of longevity and these genetic influences are illustrated by the transmission of longevity, even if parents themselves did not become long-lived but had long-lived relatives, such as siblings or parents. Likewise, we observed that an additive increase in the number of parents, siblings, or aunts and uncles is associated with an increase in the survival of study participants and the children of study participants. This additive pattern is not necessarily expected if the findings are due to other, non-genetic, factors that cluster within families (for example wealth). This evidence is strengthened by the fact that similar additive associations were identified for study participants and children of study participants without long-lived parents but with long-lived siblings or aunts and uncles (where the latter generally share less environmental influences with the IPs). Further evidence for the transmission of a genetic component was shown by the fact that none of the tested environmental/social factors, including socio-economic status, sibship size, birth year, twin birth, religious denomination, and birth order, affected the associations between parental/sibling longevity and the survival of their children. These findings are in line with other studies using historical pedigree data[2,77,78] and with unpublished results based on the LINKS data (Forthcoming: Mourits et al, 2019). In addition, the fact that we observed very similar results between the different databases used for our analyses, which cover populations with vastly different environmentally related mortality regimes, significantly adds to the generalizability of our observations regarding the genetic component to human longevity.

Nevertheless, it might be possible that 1. the familial component of longevity is explained by other, unobserved, environmental factors, 2. that some of the definitions of historical factors do not capture the same underlying concept as their contemporary counterpart, 3. That adding the environmental/social factors as covariates to the model in order to explain the parent-offspring association does not cover the full extent of the complexity of familial longevity as for example sibling effect may be independent of parental effects. Regarding the third point, the parent-offspring association captures familial longevity components that might be transmitted from parents to their offspring. There may however be other parts to this

familial component, such as sibship effects, that are not captured by the parent-offspring association. An example may be the competition for resources between siblings in times of scarcity. This competition aspect may in fact explain a part of the familial component of longevity that is reflected in siblings and not in the parent-offspring association. Alternatively, the influences of environmental/social covariates may only become present when sufficient contrast is present between long-lived and non-long lived families. For example, socio-economic status may explain the difference between long and short-lived families but might not explain the difference between long-lived and non-long-lived families as the non-long-lived group can still contain very old persons. An example regarding point two concerns socio-economic status, which in historical data is based on HISCO, a profession based measure or a variant to that[79,80], whereas in contemporary data socio-economic status is often measured in terms of a combination between income, educational attainment, and profession[81–83]. Furthermore, regarding point one, factors such as educational attainment, living environment, social networks, but also lifestyle, eating habits, and activity pattern cluster in families[52–57] and thus may explain a part of the familial component of longevity. Moreover, an important study in Science showed that having little money changes the human mindset to a state of short-term thinking, in terms of financial planning, healthy lifestyle, etc.[84]. Hence, growing up in a state of poverty can significantly influence a person's long-term decision making, affecting that person's health, and at least gives a difficult start in life. Future research in existing data may focus on more detailed analyses of the familial component of longevity. In addition, future research in more contemporary populations, supplemented with extensive pedigree information is needed to gain more insight in the role such factors play in explaining the familial component of longevity.

Future perspectives

Opportunities for longevity research with (big) genealogical data

With the increase of available digitized data, new opportunities have opened up for the analysis of big genealogical data. In the Netherlands, the LINKS project is currently complete for the province of Zeeland and the data is described in **chapter 3 and 4**. The LINKS project is continuing for the entire Netherlands and soon all provinces will be covered in the database, providing researchers with extensive pedigrees and solving the problem of missing migration to other provinces in the LINKS Zeeland version. In addition, worldwide efforts of large genealogy websites such as Geni[85,86] and Ancestry[87] provide a commercial platform with infrastructure for persons to map their family tree and provide a DNA sample to identify unknown relatives. Both the extended LINKS data and the genealogy websites provide novel opportunities to study demographic aspects of longevity and other phenotypes. In fact, the Geni and Ancestry pedigrees have already been used to estimate the heritability of lifespan based on millions of individuals[85,87]. In addition, combining genetic information of living individuals with pedigree data of their deceased and living ancestors has been used to successfully solve cold case murders in the United States[86]. Moreover, combined genetic and pedigree data opens up new research opportunities, as individuals from specific families can be identified based on their family history of a specific trait, such as longevity.

The increase of available digitized data also opens opportunities for linking different data sources. An example, as discussed above concerns combining pedigree and genetic data, but there are also opportunities to link different data sources that are used to reconstruct pedigrees to each other. This allows the cross-checking and improving the quality of pedigree and life course databases, based on a single source. An example of this is discussed in **chapter 3**. In addition, pedigrees in data such as the HSN could be extended by connecting the LINKS data as discussed in **chapter 3**. In the future it may be possible to not only link contemporary genetic data, but also more elaborate data, for example about income, socio-economic status, different health parameters, such as blood pressure, but also social networks, cause of death, lifestyle, and living conditions. By combining all these data sources, rich datasets can be created to provide novel insights into the familial component of longevity and many other traits.

Demographic (longevity) research

In this dissertation we focused mainly on investigating the familial component of longevity by inquiring into the influence of environmental/social factors on individual and familial longevity. Moreover, we studied the definition of the longevity trait itself and used our insights to construct a method, the LRC score, to identify individuals from long-lived families. Now that we have a better understanding about which factors contribute to longevity, how to define longevity and which families to investigate it is possible to use the existing and novel large scale genealogical data to investigate the heritability of longevity in the right families, since this has only been done for lifespan in the general population up to now8[5,87–94]. Following this, it is also possible to test how much of the variance in longevity is due to additive genetic effects, or for example dominant or epistatic effects. Separating the variance of a trait can be done with variance components analysis and was already done with twin studies for lifespan[91,92] but is still open for longevity. In addition to the variance components analysis it is possible to test different transmission patterns of longevity. Testing longevity transmission

patterns can be done with segregation analysis[95–100], but up to now this has not been done for either lifespan or longevity as it requires extensive pedigrees. Segregation analysis allows the estimation of how a phenotype is transmitted to subsequent generations by testing how well the observed transmission fits that of a known genetic pattern. One assumed genetic transmission pattern can be the transmission of a dominant autosomal trait, but many options can be tested, even a non-genetically/socially driven transmission pattern.

Another application of large scale high quality demographic data, such as the LINKS and the UPDB, is to address general demographic research questions. One of the pressing general demographic longevity questions is that of the existence of a decrease in death rates at later ages or the existence of a mortality plateau[101,102]. The existence of a mortality plateau implies that the risk of dying does no longer increase after a certain age. According to some, this implies that the aging process has stopped, since aging goes hand in hand with exponentially increasing death rates during life[103,104]. In animal studies a mortality plateau was already observed[23] but in humans the observations have not led to consistent conclusions[101,102]. A recent study even argues that administrative mistakes in ages at death increase at higher ages and that the accumulation of these mistakes can mimic a mortality leveling-off, or even a mortality plateau in humans[105]. In the Netherlands the civil registration contains a very low level of mistakes as shown in **chapter 3** and in various other studies[106]. Similarly, the mortality data in the UPDB was cross-checked with multiple sources, ensuring high quality mortality data. The low level of mistakes in the Dutch civil registration and the prospect of coverage for the entire Netherlands open up new opportunities to use Dutch data, as well as the UPDB, to investigate the existence of a mortality leveling-off, or mortality plateau at extreme ages in humans.

Using the LRC score, described in **chapter 6**, we were able to identify individuals who likely descent from genetically enriched families for longevity (family cases). Moreover, we identified individuals without any long-lived relatives (family controls) and individuals with up to 20% long-lived relatives but with a mortality pattern that resembles the general population (potential phenocopies) and who are unlikely to be genetically enriched for longevity. For future research it is interesting to connect large scale genealogical data to more contemporary demographic data that contains information on social and economic indicators to investigate what factors cause the familial aggregation of longevity in the potential phenocopy individuals. Moreover, the individuals who become long-lived but have no long-lived relatives at all are interesting to study, as they will likely have acquired their longevity by mean of a healthy lifestyle. We will discuss the potential for genetic research with the family based cases and controls in the next section.

Genetic longevity research

Genetic longevity research focused on the identification of genetic variants with little success[9–16]. Such studies mainly focused on singletons that survived beyond a threshold of for example, 80, 90, or 100 years, and sometimes a survival percentile is used, such as belonging to the top 10%, 5%, or 1% oldest persons of their birth cohort[1,21]. In addition, as in the UK Biobank, middle aged cases are studied based on the longevity of their parents. Loci are sometimes identified in such studies which seem to represent lifespan associated genes, rather than (protective) longevity loci. One of the main reasons for the limited success in genetic longevity studies is the uncertainty in defining the heritable longevity phenotype2 since a large (unknown) part of the population in the last 200 years reaches a high age without representing familial longevity

(phenocopies). It was unknown at what survival percentile longevity becomes heritable in unselected multigenerational datasets (**chapter 4**) and how many family members should be long-lived in order to avoid phenocopies (**chapter 6**). We observed that longevity becomes heritable beyond the top 10% survivors (**chapter 4**) when at least 20%, but ideally 30% of all family members are also within the top 10% survivors (**chapter 6**). We applied this knowledge to construct the LRC score and estimated that 27% of the F3 descendants showed a survival pattern similar to the general population even though they had at least one long-lived parent (**chapter 6**). Hence, in this dissertation, we attempted to improve the inconsistent definition of longevity as an important factor to gain more success in identifying longevity variants. We constructed a longevity definition and identification strategy to select the largest number of persons with a likely genetic enrichment for longevity and we observed that for such definition, the environmental/familial factors seem to play a limited role in reaching the long lived status.

The most consistent evidence has been obtained for genetic variants in APOE and FOXO3A genes[10–15,18], in either genome-wide association studies (GWAS) or candidate gene studies. Such studies use a case-control design, in which allele frequencies among long-lived cases are compared with those in controls. In addition to the misclassification of cases and controls due to the lack of a clear longevity definition, the nature of the used study designs and methods, in GWAS and candidate gene studies only allowed the identification of common genetic variants[107]. The assumed genetic architecture in such studies is that many common variants in the population have small effects on the trait. However, even though longevity studies are likely to often include phenocopies, more genetic longevity variants should have been identify given the very large sample sizes of current meta GWA studies (personal communication on a worldwide longevity GWAS study). Hence, the small proportion of explained heritability by the currently identified loci suggests that rare variants may potentially play a role in the longevity phenotype. Apart from that, the observation that, in comparison to controls, members of long-lived families carry the same numbers of disease risk alleles as other populations of elderly, while showing lower prevalence of age-related disease, may indicate that these families carry protecting factors1[08].

Another interesting possibility concerns the current availability of large genealogical datasets, as described in the "future perspectives" section and combining these data with genetic data. Combining broad genealogic and genetic data provides the possibility to use the distant relatives approach for longevity. In this approach, distantly related long-lived relatives from long-lived families can be examined for the common and rare genetic variants that they share in common (identical by descent, IBD). Distant long-lived relatives in a family are expected to have acquired their longevity by the same genetic variant. For genetic studies it is convenient that such relatives have less DNA in common than for example siblings because it becomes easier to identify the genetic variants responsible for longevity. Moreover, the distant relatives approach has been successfully applied for other complex traits such as oral clefts[109], thoracicaorticaneurysms[110], and osteoarthritis[111]. Second degree nieces and nephews or more distantly related relatives are considered distant relatives and they are illustrated by the fourth (green) generation in *Figure 2* and the "2C" notation in *Figure 3*. These second degree nieces and nephews share on average 238 Centimorgan (CM) DNA strands. More distantly related family members will thus share on average less than 238 CM DNA strands *(Figure 3)*.

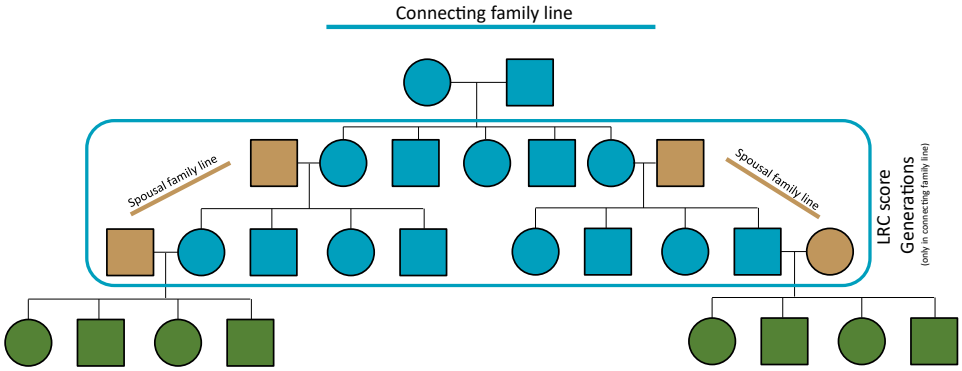


Figure 2: Example pedigree illustrating long-lived families (LRC ≥ 0.30) and distant relatives.

The fourth generation (green color) represents distant relatives (second degree nieces and nephews). A fifth generation (and further) would indicate even more distant relatives. The ancestral (blue and brown colors) generations can be used to identify long-lived families by means of the LRC score.

Here, the intersection between the current availability of big genealogical data, the LRC score to identify the best genetically enriched individuals for longevity, and the need to focus on both common and rare longevity variants using whole genome sequencing data provides novel opportunities to identify the most interesting families for a new genetic study into longevity variants. In addition, current GWA studies focusing primarily on singletons can be extended with familial information in order to rule out misclassified cases who may now obscure the identification of longevity loci in GWAS.

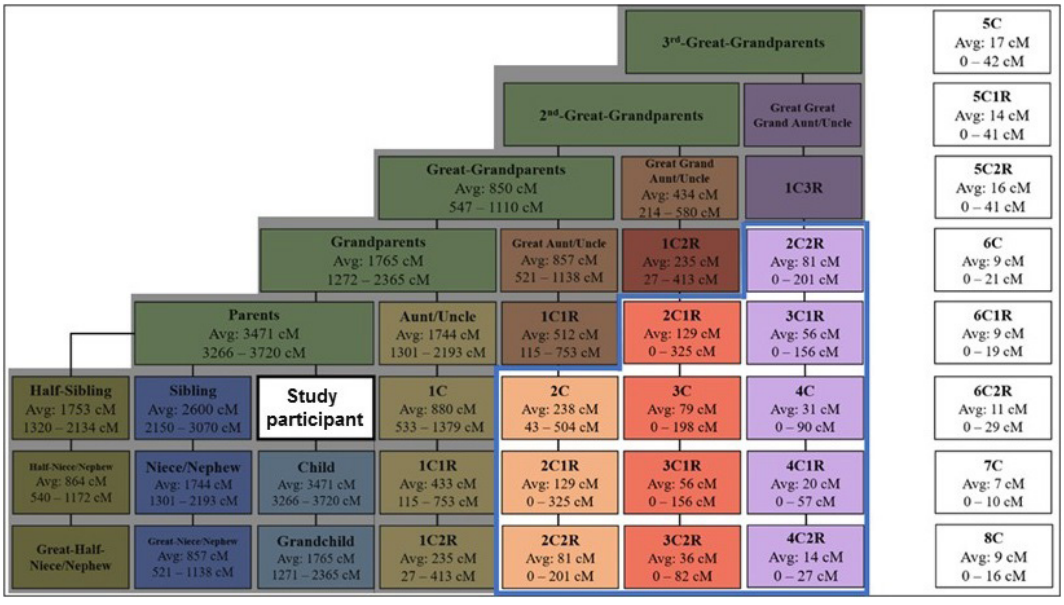


Figure 3: DNA sharing between relatives.

Adjusted from the Genetic Genealogist (https://thegeneticgenealogist.com). G=Great, N=Niece or Nephew, C=Cousin, R=Removed. For example: 3C2R = third cousin 2 times removed (2 generations away).

Final remark

Even though genetic longevity research in humans has been very difficult, the results of this dissertation show that with the selection of cases (and controls) from the proper families new opportunities open up for genetic as well as social research into human longevity.

References

1 Berg, N. van den *et al.* Longevity Around the Turn of the 20th Century: Life-Long Sustained Survival Advantage for Parents of Today's Nonagenarians. *Journals Gerontol. Ser. A* **73**, 1295–1302 (2018).

2 van den Berg, N. *et al.* Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nat. Commun.* **10**, 35 (2019).

3 Westendorp, R. G. J. *et al.* Nonagenarian Siblings and Their Offspring Display Lower Risk of Mortality and Morbidity than Sporadic Nonagenarians: The Leiden Longevity Study. *J. Am. Geriatr. Soc.* **57**, 1634–1637 (2009).

4 Terry, D. F. *et al.* Lower All-Cause, Cardiovascular, and Cancer Mortality in Centenarians' Offspring. *J. Am. Geriatr. Soc.* **52**, 2074–2076 (2004).

5 Dutta, A. *et al.* Longer lived parents: Protective associations with cancer incidence and overall mortality. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **68**, 1409–1418 (2013).

6 Christensen, K. *et al.* Physical and cognitive functioning of people older than 90 years: a comparison of two Danish cohorts born 10 years apart. *Lancet* **382**, 1507–1513 (2013).

7 Christensen, K., McGue, M., Petersen, I., Jeune, B. & Vaupel, J. W. Exceptional longevity does not result in excessive levels of disability. *Proc. Natl. Acad. Sci.* **105**, 13274–13279 (2008).

8 Andersen, S. L., Sebastiani, P., Dworkis, D. a., Feldman, L. & Perls, T. T. Health Span Approximates Life Span Among Many Supercentenarians: Compression of Morbidity at the Approximate Limit of Life Span. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **67A**, 395–405 (2012).

9 Joshi, P. K. *et al.* Genome-wide meta-analysis associates HLA-DQA1/DRB1 and LPA and lifestyle factors with human longevity. *Nat. Commun.* **8**, 1–13 (2017).

10 Eline Slagboom, P., van den Berg, N. & Deelen, J. Phenome and genome based studies into human ageing and longevity: An overview. *Biochim. Biophys. Acta - Mol. Basis Dis.* **1864**, 2742–2751 (2018).

11 Shadyab, A. H. & LaCroix, A. Z. Genetic factors associated with longevity: A review of recent findings. *Ageing Res. Rev.* **19**, 1–7 (2015).

12 Deelen, J. *et al.* Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* **23**, 4420–4432 (2014).

13 Broer, L. *et al.* GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **70**, 110–118 (2015).

14 Willcox, B. J. *et al.* FOXO3A genotype is strongly associated with human longevity. *Proc. Natl. Acad. Sci.* **105**, 13987–13992 (2008).

15 Flachsbart, F. *et al.* Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc. Natl. Acad. Sci.* **106**, 2700–2705 (2009).

16 Pilling, L. C. *et al.* Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany. NY).* **9**, 2504–2520 (2017).

17 Partridge, L., Deelen, J. & Slagboom, P. E. Facing up to the global challenges of ageing. *Nature* **561**, 45–56 (2018).

18 Sebastiani, P. *et al.* Four Genome-Wide Association Studies Identify New Extreme Longevity Variants. *J. Gerontol. A. Biol. Sci. Med. Sci.* **72**, 1453–1464 (2017).

19 Flachsbart, F. *et al.* Immunochip analysis identifies association of the RAD50/IL13 region with human longevity. *Aging Cell* **15**, 585–588 (2016).

20 Zeng, Y. *et al.* Novel loci and pathways significantly associated with longevity. *Sci. Rep.* **6**, 21243 (2016).

21 van den Berg, N., Beekman, M., Smith, K. R., Janssens, A. & Slagboom, P. E. Historical demography and longevity genetics: Back to the future. *Ageing Res. Rev.* **38**, 28–39 (2017).

22 Oeppen, J. & Vaupel, J. W. Broken limits to life expectancy. *Science* **296**, 1029–1031 (2002).

23 Vaupel, J. W. Biodemographic Trajectories of Longevity. *Science (80-.).* **280**, 855–860 (1998).

24 Burger, O., Baudisch, A. & Vaupel, J. W. Human mortality improvement in evolutionary context. *Proc. Natl. Acad. Sci.* **109**, 18210–18214 (2012).

25 Rutten, W. *De vreselijkste aller harpijen: Pokkenepidemieën en pokkenbestrijding in Nederland in de 18e en 19e eeuw.* ('t Goy-Houten: HES Uitgeves B.V., 1997).

26 Walhout, E. C. Is breast best? Evaluating breastfeeding patterns and causes of infant death in a Dutch province in the period 1875–1900. *Hist. Fam.* **15**, 76–90 (2010).

27 Kunst, A. E., Looman, C. W. N. & Mackenbach, J. P. Socio-economic mortality differences in the Netherlands in 1950–1984: A regional study of cause-specific mortality. *Soc. Sci. Med.* **31**, 141–152 (1990).

28 Fogel, R. W. *Aging, Health and Public Policy.* **30**, (2004).

29 Fogel, R. W. & Costa, D. L. A Theory of Technophysio Evolution, With Some Implications for Forecasting Population, Health Care Costs, and Pension Costs. *Demography* **34**, 49 (1997).

30 Wachter, K. W. & Finch, C. E. *Between Zeus and the Salmon: The Biodemography of Longevity. Population (English Edition)* (Washington, DC: The National Academies Press, 1997).

31 Christensen, K., McGue, M., Petersen, I., Jeune, B. & Vaupel, J. W. Exceptional longevity does not result in excessive levels of disability. *Proc. Natl. Acad. Sci.* **105**, 13274–13279 (2008).

32 Zeng, Y., Feng, Q., Hesketh, T., Christensen, K. & Vaupel, J. W. Survival, disabilities in activities of daily living, and physical and cognitive functioning among the oldest-old in China: a cohort study. *Lancet* **389**, 1619–1629 (2017).

33 Pignolo, R. J. Exceptional Human Longevity. *Mayo Clin Proc* **94**, 1–15 (2018).

34 Crimmins, E. M. Lifespan and Healthspan: Past, Present, and Promise. *Gerontologist* **55**, 901–911 (2015).

35 van dijk, I. K. & Mandemakers, K. Like mother, like daughter. Intergenerational transmission of infant mortality clustering in Zeeland, the Netherlands, 1833-1912. *Hist. Life Course Stud.* **1**–26 (2018).

36 Hoogerhuis, O. W. *Baren op Beveland. Vruchtbaarheid en zuigelingensterfte in Goes en omliggende dorpen gedurende de 19e eeuw.* (Wageningen Afd. Agrarische Geschiedenis Wageningen Universiteit 2003, 2003).

37 Hofstee, E. W. *De demografische ontwikkeling van Nederland sinds 1800.* (Haalem: Fibula-Van-Dishoeck, 1974).

38 Ekamper, P., Duin, C. van, Poppel, F. van & Mandemakers, K. Heat Waves and Cold Spells and their Effect on Mortality: an Analysis of Micro-data for the Netherlands in the Nineteenth and Twentieth Centuries. *Ann. Demogr. Hist. (Paris).* **120**, 55 (2010).

39 Janssen, F. & van Poppel, F. The Adoption of Smoking and Its Effect on the Mortality Gender Gap in Netherlands: A Historical Perspective. *Biomed Res. Int.* 2015, 1–12 (2015).

40 Lindahl-Jacobsen, R. *et al.* The male–female health-survival paradox and sex differences in cohort life expectancy in Utah, Denmark, and Sweden 1850–1910. Ann. *Epidemiol.* **23**, 161–166 (2013).

41 Beltrán-Sánchez, H., Finch, C. E. & Crimmins, E. M. Twentieth century surge of excess adult male mortality. *Proc. Natl. Acad. Sci.* **112**, 8993–8998 (2015).

42 Sebastiani, P. *et al.* Families Enriched for Exceptional Longevity also have Increased Health-Span: Findings from the Long Life Family Study. *Front. Public Heal.* **1**, 1–9 (2013).

43 Ailshire, J. A., Beltrán-Sánchez, H. & Crimmins, E. M. Social Characteristics and Health Status of Exceptionally Long-Lived Americans in the Health and Retirement Study. *J. Am. Geriatr. Soc.* **59**, 2241–2248 (2011).

44 Ash, A. S. *et al.* Are Members of Long-Lived Families Healthier Than Their Equally Long-Lived Peers? Evidence From the Long Life Family Study. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **70**, 971–976 (2015).

45 Jopp, D. S., Boerner, K., Ribeiro, O. & Rott, C. Life at Age 100: An International Research Agenda for Centenarian Studies. *J. Aging Soc. Policy* **28**, 133–147 (2016).

46 Boerner, K., Jopp, D. S., Park, M.-K. S. & Rott, C. Whom Do Centenarians Rely on for Support? Findings From the Second Heidelberg Centenarian Study. *J. Aging Soc. Policy* **28**, 165–186 (2016).

47 Seeman, T. E. *et al.* Behavioral and Psychosocial Predictors of Physical Performance: MacArthur Studies of Successful Aging. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **50A**, M177–M183 (1995).

48 Andersen, S. L. *et al.* Personality Factors in the Long Life Family Study. *Journals Gerontol. Ser. B* **68**, 739–749 (2013).

49 Yang, Y. C. *et al.* Social relationships and physiological determinants of longevity across the human life span. *Proc. Natl. Acad. Sci.* **113**, 578–583 (2016).

50 Hensley, B. *et al.* Life Events and Personality Predicting Loneliness Among Centenarians: Findings From the Georgia Centenarian Study. *J. Psychol.* **146**, 173–188 (2012).

51 Yashin, A. I. *et al.* "Predicting" parental longevity from offspring endophenotypes: Data from the Long Life Family Study (LLFS). *Mech. Ageing Dev.* **131**, 215–222 (2010).

52 Heath, A. C. *et al.* Education policy and the heritability of educational attainment. *Nature* **314**, 734–736 (1985).

53 Spera, C., Wentzel, K. R. & Matto, H. C. Parental Aspirations for Their Children's Educational Attainment: Relations to Ethnicity, Parental Education, Children's Academic Performance, and Parental Perceptions of School Climate. *J. Youth Adolesc.* **38**, 1140–1152 (2009).

54 Eccles, J. S. Influences of parents' education on their children's educational attainments: the role of parent and child perceptions. *London Rev. Educ.* **3**, 191–204 (2005).

55 Edvardsson, K. *et al.* Giving offspring a healthy start: parents' experiences of health promotion and lifestyle change during pregnancy and early parenthood. *BMC Public Health* **11**, 936 (2011).

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

57 Thibault, H., Contrand, B., Saubusse, E., Baine, M. & Maurice-Tison, S. Risk factors for overweight and obesity in French adolescents: Physical activity, sedentary behavior and parental characteristics. *Nutrition* **26**, 192–200 (2010).

58 Perls, T. T. *et al.* Life-long sustained mortality advantage of siblings of centenarians. *Proc. Natl. Acad. Sci.* **99**, 8442–8447 (2002).

59 Kerber, R. A., Brien, E. O., Smith, K. R. & Cawthon, R. M. Familial Excess Longevity in Utah Genealogies. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **56**, 130–139 (2001).

60 Gudmundsson, H., Gudbjartsson, D. F. & Kong, A. Inheritance of human longevity in Iceland. *Eur. J. Hum. Genet. EJHG* **8**, 743–749 (2000).

61 Schoenmaker, M. *et al.* Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur. J. Hum. Genet.* **14**, 79–84 (2006).

62 Hjelmborg, J. vB *et al.* Genetic influence on human lifespan and longevity. *Hum. Genet.* **119**, 312–321 (2006).

63 Sebastiani, P., Nussbaum, L., Andersen, S. L., Black, M. J. & Perls, T. T. Increasing Sibling Relative Risk of Survival to Older and Older Ages and the Importance of Precise Definitions of "Aging," "Life Span," and "Longevity". *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **71**, 340–346 (2016).

64 Willcox, B. J., Willcox, D. C., He, Q., Curb, J. D. & Suzuki, M. Siblings of Okinawan centenarians share lifelong mortality advantages. *J. Gerontol. A. Biol. Sci. Med. Sci.* **61**, 345–54 (2006).

65 Pedersen, J. K. *et al.* The Survival of Spouses Marrying Into Longevity-Enriched Families. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **72**, 109–114 (2017).

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

67 Lindahl-Jacobsen, R. *et al.* The male–female health-survival paradox and sex differences in cohort life expectancy in Utah, Denmark, and Sweden 1850–1910. *Ann. Epidemiol.* **23**, 161–166 (2013).

68 Rozing, M. P. *et al.* Familial longevity is associated with decreased thyroid function. *J. Clin. Endocrinol. Metab.* **95**, 4979–4984 (2010).

69 Sebastiani, P. *et al.* A Family Longevity Selection Score: Ranking Sibships by Their Longevity, Size, and Availability for Study. *Am. J. Epidemiol.* **170**, 1555–1562 (2009).

70 Arbeeva, L. S. *et al.* How Well Does the Family Longevity Selection Score Work: A Validation Test Using the Utah Population Database. *Front. Public Heal.* **6**, (2018).

71 van den Berg, N. *et al.* Families in Comparison: An individual-level comparison of life course and family reconstructions between population and vital event registers. *SocArXiv* (2018).

72 Jarry, V., Gagnon, A. & Bourbeau, R. Survival advantage of siblings and spouses of centenarians in 20th-century Quebec. *Can. Stud. Popul.* **39**, 67 (2013).

73 Montesanto, A. *et al.* The genetic component of human longevity: analysis of the survival advantage of parents and siblings of Italian nonagenarians. *Eur. J. Hum. Genet.* **19**, 882–886 (2011).

74 Houde, L., Tremblay, M. & Vézina, H. Intergenerational and Genealogical Approaches for the Study of Longevity in the Saguenay-Lac-St-Jean Population. *Hum. Nat.* **19**, 70–86 (2008).

75 Kemkes-Grottenthaler, A. Parental effects on offspring longevity—evidence from 17th to 19th century reproductive histories. *Ann. Hum. Biol.* **31**, 139–158 (2004).

76 Deluty, J. A., Atzmon, G., Crandall, J., Barzilai, N. & Milman, S. The influence of gender on inheritance of exceptional longevity. *Aging (Albany. NY).* **7**, 412–418 (2015).

77 Danzhen You, Danan Gu & Zeng Yi. Familial Transmission of Human Longevity Among the Oldest-Old in China. *J. Appl. Gerontol.* **29**, 308–332 (2010).

78 Gavrilov, L. A. & Gavrilova, N. S. Predictors of Exceptional Longevity: Effects of Early-Life and Midlife Conditions, and Familial Longevity. *North Am. Actuar. J.* **19**, 174–186 (2015).

79 Leeuwen, M., Maas, H. & Miles, A. HISCO:
Historical international standard classification of occupations. (Leiden Universithy Press, 2002).

80 Mandemakers, K. *et al.* HSN standardized HISCO-coded and classified occupational titles. (2018).

81 Ganzeboom, H. B. G., De Graaf, P. M. & Treiman, D. J. A standard international socio-economic index of occupational status. *Soc. Sci. Res.* **21**, 1–56 (1992).

82 Braveman, P. A. *et al.* Socioeconomic Status in Health Research. *JAMA* **294**, 2879 (2005).

83 Shavers, V. L. Measurement of socioeconomic status in health disparities research.
J. Natl. Med. Assoc. **99**, 1013–23 (2007).

84 Shah, A. K., Mullainathan, S. & Shafir, E. Some Consequences of Having Too Little. *Science (80-.).* **338**, 682–685 (2012).

85 Kaplanis, J. *et al.* Quantitative analysis of population-scale family trees with millions of relatives.
Science **360**, 171–175 (2018).

86 Erlich, Y., Shor, T., Pe'er, I. & Carmi, S. Identity inference of genomic data using long-range familial searches.
Science **4832**, eaau4832 (2018).

87 Ruby, J. G. *et al.* Estimates of the Heritability of Human Longevity Are Substantially Inflated due to Assortative Mating. *Genetics* **210**, 1109–1124 (2018).

88 Skytthe, A. *et al.* Longevity Studies in GenomEUtwin. *Twin Res.* **6**, 448–454 (2003).

89 Hrubec, Z. & Neel, J. V. Familial factors in early deaths: Twins followed 30 years to ages 51?61 in 1978.
Hum. Genet. **59**, 39–46 (1981).

90 Ljungquist, B., Berg, S., Lanke, J., McClearn, G. E. & Pedersen, N. L. The Effect of Genetic Factors for Longevity:
A Comparison of Identical and Fraternal Twins in the Swedish Twin Registry.
Journals Gerontol. Ser. A Biol. Sci. Med. Sci. **53A**, 441–446 (1998).

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

92 Herskind, A. M. *et al.* The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Hum. Genet.* **97**, 319–323 (1996).

93 Wyshak, G. Biodemography and Social Biology Fertility and longevity in twins, sibs , and parents of twins Parents of Twins. *Biodemography Soc. Biol.* **25**, 315–330 (2010).

94 Jarvik, L. F., Falek, A. & Kallmann, F. J. Survival Trends in a Senescent Twin Population. *Am. J. Hum. Genet.* **12**, 170–179 (1960).

95 King, M. C. *et al.* Genetic Epidemiology of Breast Cancer and Associated Cancers in High-Risk Families. I. Segregation Analysis. *JNCI J. Natl. Cancer Inst.* **71**, 463–467 (1983).

96 Williams, W. R., Anderson, D. E. & Rao, D. C. Genetic epidemiology of breast cancer: Segregation analysis of 200 Danish pedigrees. *Genet. Epidemiol.* **1**, 7–20 (1984).

97 Win, A. K. *et al.* Prevalence and Penetrance of Major Genes and Polygenes for Colorectal Cancer. Cancer Epidemiol. *Biomarkers Prev.* **26**, 404–412 (2017).

98 Newman, B., Austin, M. a, Lee, M. & King, M. C. Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 3044–3048 (1988).

99 Carter, B. S., Beaty, T. H., Steinberg, G. D., Childs, B. & Walsh, P. C. Mendelian inheritance of familial prostate cancer. *Proc. Natl. Acad. Sci. U. S. A.* **89**, 3367–71 (1992).

100 Lalouel, J. M. & Morton, N. E. Complex segregation analysis with pointers. *Hum. Hered.* **31**, 312–21 (1981).

101 Barbi, E., Lagona, F., Marsili, M., Vaupel, J. W. & Wachter, K. W. The plateau of human mortality: Demography of longevity pioneers. **360**, 1459–1461 (2018).

102 Dong, X., Milholland, B. & Vijg, J. Evidence for a limit to human lifespan. *Nature* **538**, 257–259 (2016).

103 Weitz, J. S. & Fraser, H. B. Explaining mortality rate plateaus. *Proc. Natl. Acad. Sci.* **98**, 15383–15386 (2001).

104 Chen, H. -y., Zajitschek, F. & Maklakov, A. A. Why ageing stops: heterogeneity explains late-life mortality deceleration in nematodes. *Biol. Lett.* **9**, 20130217–20130217 (2013).

105 Newman, S. J. Errors as a primary cause of late-life mortality deceleration and plateaus. *PLOS Biol.* **16**, e2006776 (2018).

106 Janssens, A. *Family and social change: The household as a process in an industrializing community.* (Cambridge University Press, 2002).

107 Bloss, C. S., Pawlikowska, L. & Schork, N. J. Contemporary human genetic strategies in aging research. *Ageing Res. Rev.* **10**, 191–200 (2011).

108 Beekman, M. *et al.* Genome-wide association study (GWAS)-indentified disease risk alleles do not compromise human longevity. *Proc. Natl. Acad. Sci. USA* **107**, 18046–18049 (2010).

109 Bureau, A. *et al.* Whole Exome Sequencing of Distant Relatives in Multiplex Families Implicates Rare Variants in Candidate Genes for Oral Clefts. *Genetics* **197**, 1039–1044 (2014).

110 Milewicz, D. M., Regalado, E. S., Shendure, J., Nickerson, D. A. & Guo, D. Successes and challenges of using whole exome sequencing to identify novel genes underlying an inherited predisposition for thoracic aortic aneurysms and acute aortic dissections. *Trends Cardiovasc. Med.* **24**, 53–60 (2014).

111 Ramos, Y. F. M. *et al.* A gain of function mutation in TNFRSF11B encoding osteoprotegerin causes osteoarthritis with chondrocalcinosis. *Ann. Rheum. Dis.* **74**, 1756–1762 (2015).

CHAPTER 8

SUPPLEMENTARY MATERIAL

CHAPTER 4 - CHAPTER 5 - CHAPTER 6

Chapter 4

Supplementary Table 1: Survival analysis for IPs by top 5% siblings and top 5% parents

		UPDB			LINKS	
	N (mean)	HR (95% CI)	P-value	N (mean)	HR (95% CI)	P-value
Top 5% parents (F1)						
0 (ref)	8149 (0.79)			8975 (0.88)		
1	1961 (0.19)	0.83 (0.79-0.88)	8.19*10-11	1097 (0.11)	0.79 (0.74-0.84)	2.07*10-12
2	136 (0.02)	0.70 (0.57-0.86)	5.97*10-4	42 (0.01)	0.83 (0.60-1.16)	2.75*10-1
Top 5% sibs (F2)						
0 (ref)	8169 (0.79)			9331 (0.92)		
1	1667 (0.17)	0.82 (0.74-0.91)	1.48*10-4	712 (0.07)	0.79 (0.67-0.93)	4.74*10-3
2+	410 (0.04)	0.75 (0.63-0.89)	9.30*10-4	71 (0.01)	0.65 (0.42-1.02)	6.12*10-2
LDS (F2)						
0 - non-religious(ref)	2753 (0.27)					
1 – baptized	512 (0.05)	0.69 (0.62-0.77)	4.47*10-12	NA	NA	NA
2 - baptized + endowment	6736 (0.66)	0.82 (0.78-0.86)	4.60*10-15	NA	NA	NA
3 - missing	245 (0.02)	0.87 (0.78-1.00)	5.78*10-2	NA	NA	NA
Sibship size (F2)	10246 (6.28)	1.01 (1.00-1.02)	1.04*10-1	10114 (6.34)	1.00 (0.99-1.01)	4.35*10-1
Birth cohort, years (F2)	10246 (1868)	0.99 (>0.99<1.00)	1.28*10-10	10114 (1835)	0.99 (>0.99<1.00)	<1.00*10-15
Sex (F2)						
Man (ref)	5053 (0.49)			4776 (0.48)		
Women	5193 (0.51)	0.69 (0.65-0.74)	<1.00*10-15	5338 (0.52)	1.01 (0.97-1.06)	4.35*10-1
SES – OCC_1950 (F2)						
0 - High (ref)	315 (0.03)			67 (0.01)		
1	1482 (0.14)	1.12 (0.98-1.28)	1.11*10-1	645 (0.06)	0.89 (0.69-1.15)	3.64*10-1
2	400 (0.04)	1.14 (0.97-1.33)	1.20*10-1	536 (0.05)	0.98 (0.76-1.27)	8.66*10-1
3	352 (0.03)	1.22 (1.03-1.44)	1.81*10-2	62 (0.01)	0.77 (0.54-1.10)	1.51*10-1
4	187 (0.02)	1.10 (0.91-1.34)	3.17*10-1	71 (0.01)	0.95 (0.67-1.34)	7.67*10-1
5	891 (0.09)	1.23 (1.07-1.42)	3.59*10-3	733 (0.07)	0.80 (0.62-1.03)	8. 59*10-2
6	668 (0.07)	1.29 (1.12-1.49)	5.11*10-4	311 (0.03)	0.86 (0.66-1.13)	2.81*10-1
7	522 (0.05)	1.25 (1.07-1.45)	4.92*10-3	759 (0.08)	0.82 (0.63-1.06)	2.31*10-1
8	168 (0.02)	1.22 (0.99-1.50)	5.94*10-2	574 (0.06)	0.85 (0.66-1.10)	2.11*10-1
9 – Low	562 (0.05)	1.39 (1.19-1.61)	2.11*10-5	3656 (0.36)	0.83 (0.65-1.07)	1.47*10-1
999 - missing	4699 (0.46)	1.59 (1.40-1.81)	1.74*10-12	2700 (0.26)	0.92 (0.72-1.18)	4.93*10-1

Table corresponds to the CH curves in the top and bottom right panel of supplementary Figure 2. Means represent a mean for a ontinuous variable and a proportion for a categorical variable. Additional covariates are: age mom at birth, birth order, birth intervals (in years), twin birth. When the p-value was lower than 1.00e-15 we indicated the P-value as <1.00e-15. LDS: the church of Jesus Christ of latter-day saints (Mormon church), SES: socio-economic status, OCC: occupational coding scheme of 1950. P-values are estimated with cox regression.

Supplementary Table 2: Frailty survival analysis for Children of IPs by top 5% IP's and aunt and uncles of children

		UPDB			LINKS	
	N (mean)	**HR (95% CI)**	**P-value**	**N (mean)**	**HR (95% CI)**	**P-value**
Top 5% IP (F2)						
0 non LL (ref.)	54607 (0.90)			58196 (0.93)		
1 LL	6191 (0.10)	0.83 (0.80-0.86)	<1.00*10-15	4278 (0.07)	0.83 (0.79-0.88)	5.66*10-13
Top 5% aunts and uncles (F2)						
0 (ref.)	48154 (0.79)			57508 (0.92)		
1	10166 (0.17)	0.95 (0.92-0.98)	3.79*10-4	4465 (0.07)	0.93 (0.89-0.97)	2.40*10-3
2+	2478 (0.04)	0.91 (0.86-0.96)	1.60*10-3	501 (0.01)	0.78 (0.68-0.90)	4.94*10-4
Sibshipsize (F3)	60798 (8.89)	1.02 (1.01-1.02)	<1.00*10-15	62474 (8.52)	1.00 (0.99-1.00)	5.92*10-1
Birth year (F3)	60798 (1892)	0.99 (>0.99<1.00)	<1.00*10-15	62474 (1867)	0.99 (>0.99<1.00)	6.52*10-12
Sex (F3)						
Man (ref.)	31258 (0.51)			32136 (0.52)		
Women	29540 (0.49)	0.62 (0.60-0.63)	<1.00*10-15	30338 (0.48)	0.64 (0.63-0.66)	<1.00*10-15
Famid intercept (variance)	60798 (1.00)	0.34 (0.11)		62474 (1.00)	0.34 (0.11)	
BIC	60798 (1.00)	-23798.10		62474 (1.00)	-21555.34	

Additional covariates are: birth order, birth intervals (years), age mom at birth. Religion, Socio-economic status, twin birth have been stratified. When the p-value was lower than 1.00e-15 we indicated the P-value as <1.00e-15. BIC: Bayesian Information Criterion, Famid: family identifier. P-values are estimated with cox regression.

Supplementary Table 3: Survival analysis for IP's by top 5% siblings for IPs without top 5% parents

		UPDB non-longevous parents	
	N (mean)	**HR (CI)**	**P-value**
Top 5% sibs of RP			
0 (ref.)	6665 (0.82)		
1	1219 (0.15)	0.79 (0.70-0.90)	<0.0001
2+	165 (0.02)	0.77 (0.62-0.95)	0.0168
		LINKS non-longevous parents	
	N (mean)	**HR (CI)**	**P-value**
Top 5% sibs of RP			
0 (ref.)	8354 (0.93)		
1	567 (0.06)	0.76 (0.63-0.92)	0.0045
2+	54 (0.01)	0.61 (0.37-0.99)	0.0500

Means represent a mean for a continuous variable and a proportion for a categorical variable. Additional covariates are: religion, sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, and twin birth. Here P-values were rounded to 4 digits. P-values are estimated with cox regression.

Supplementary Table 4:
frailty survival analysis for Children of IP's by top 5% aunt and uncles of children without top 5% parents

		UPDB non-longevous RP + non longevous spouse	
	N (mean)	**HR (CI)**	**P-value**
Top 5% aunts and uncles			
0 (ref.)	39338 (80)		
1	7934 (16)	0.93 (0.90-0.96)	<0.0001
2+	1816 (4)	0.91 (0.85-0.97)	0.0047
		LINKS non-longevous RP + non longevous spouse	
	N (mean)	**HR (CI)**	**P-value**
Top 5% aunts and uncles			
0 (ref.)	50165 (92)		
1	3737 (7)	0.93 (0.88-0.98)	0.0037
2+	345 (1)	0.73 (0.62-0.86)	<0.0001

Means represent a mean for a continuous variable and a proportion for a categorical variable. Additional covariates are: religion, sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, and twin birth. Here P-values were rounded to 4 digits. P-values are estimated with cox regression.

Supplementary Table 5: Survival analysis for IP's by top 5% fathers and mothers

			UPDB			LINKS	
		N (mean)	**HR (95% CI)**	**P-value**	**N (mean)**	**HR (95% CI)**	**P-value**
F1-F2 (IP)	Top 5% parents						
	0 Both parents NL	7798 (0.79)			8927 (0.88)		
	1 Pa LL / Ma NL	1037 (0.10)	0.83 (0.77-0.89)	<0.0001	495 (0.5)	0.789(0.77-0.86)	<0.0001
	2 Ma LL / Pa NL	876 (0.9)	0.83 (0.78-0.89)	<0.0001	601 (0.6)	0.77 (0.70-0.84)	<0.0001
	3 Both parents LL	136 (0.2)	0.69 (0.58-0.82)	<0.0001	42 (0.1)	0.86 (0.63-1.16)	0.3368
F2-F3 (full)	Top 5% parents						
	0 Both parents NL	47960 (0.80)			53422 (0.86)		
	1 Pa LL / Ma NL	6176 (0.10)	0.86 (0.83-0.89)	<0.0001	4668 (0.07)	0.85 (0.81-0.89)	<0.0001
	2 Ma LL / Pa NL	5352 (0.09)	0.83 (0.79-0.86)	<0.0001	3783 (0.06)	0.82 (0.78-0.86)	<0.0001
	3 Both parents LL	754 (0.01)	0.68 (0.61-0.76)	<0.0001	438 (0.01)	0.63 (0.54-0.73)	<0.0001

Means represent a mean for a continuous variable and a proportion for a categorical variable. Additional covariates are: religion, sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, and twin birth. PA=father, MA=mother, NL=non-longevous, LL=longevous, NL=non-longevous, longevous was defined as belonging to the top 10% of a persons' birth cohort. Here P-values were rounded to 4 digits. P-values are estimated with cox regression.

Supplementary Table 6 - construction of final statistical models UPDB - 10 percent

	Model 1		Model 2		Model 3		Model 4	
	N (mean)	HR (95% CI)	N (mean)	HR (95% CI)	N (mean)	HR (95% CI)	N (mean	HR (95% CI)
Top 10% parents (F1)								
0 (ref)	6640 (0.65)				7861 (0.78)		6640 (0.65)	
1	3167 (0.31)	0.89 (0.84-0.93)			2096 (0.20)	0.90 (0.86-0.95)	3167 (0.31)	0.88 (0.83-0.92)
2	439 (0.4)	0.74 (0.66-0.83)			184 (0.2)	0.76 (0.67-0.85)	439 (0.4)	0.73 (0.65-0.83)
Top 10% sibs (F2)								
0 (ref)			6720 (0.66)		8644 (0.85)		6720 (0.66)	
1			2495 (0.24)	0.82 (0.75-0.89)	1256 (0.13)	0.83 (0.76-0.90)	2495 (0.24)	0.82 (0.76-0.90)
2+			1031 (0.10)	0.74 (0.66-0.82)	214 (0.2)	0.76 (0.68-0.84)	1031 (0.10)	0.74 (0.66-0.82)
LDS (F2)								
0 - non-religious(ref)							2753 (0.27)	
1 – baptized							512 (0.05)	NA
2 - baptized + endowment							6736 (0.66)	NA
3 - missing							245 (0.02)	NA
Sibship size (F2)							10246 (6.28)	1.01 (1.00-1.02)
Birth cohort, years (F2)							10246 (1868)	0.99 (>0.99<1.00)
Sex (F2)								
Man (ref)							5053 (0.49)	
Women							5193 (0.51)	0.71 (0.67-0.76)
SES – OCC_1950 (F2)								
0 - High (ref)							315 (0.03)	
1							1482 (0.14)	1.16 (1.01-1.34)
2							400 (0.04)	1.19 (1.00-1.40)
3							352 (0.03)	1.24 (1.05-1.48)
4							187 (0.02)	1.14 (0.93-1.40)
5							891 (0.09)	1.31 (1.13-1.52)
6							668 (0.07)	1.34 (1.15-1.56)
7							522 (0.05)	1.27 (1.08-1.50)
8							168 (0.02)	1.21 (0.97-1.50)
9 - Low							562 (0.05)	1.48 (1.26-1.73)
999 - missing							4699 (0.46)	1.61 (1.40-1.84)

Means represent a mean for a continuous variable and a proportion for a categorical variable. Additional covariates are: age mom at birth, birth order, birth intervals (in years), and twin birth. LDS: the church of Jesus Christ of latter-day saints (Mormon church), SES: socio-economic status, OCC: occupational coding scheme of 1950. P-values are estimated with cox regression.

Supplementary Table 7 - construction of final statistical models LINKS - 10 percent

	Model 1		Model 2		Model 3		Model 4	
	N (mean)	HR (95% CI)	N (mean)	HR (95% CI)	N (mean)	HR (95% CI)	N (mean	HR (95% CI)
Top 10% parents (F1)								
0 (ref)	7861 (0.78)				7861 (0.78)		7861 (0.78)	
1	2096 (0.20)	0.81 (0.77-0.86)			2096 (0.20)	0.82 (0.78-0.86)	2096 (0.20)	0.82 (0.78-0.86)
2	184 (0.2)	0.68 (0.58-0.81)			184 (0.2)	0.69 (0.58-0.82)	184 (0.2)	0.69 (0.58-0.82)
Top 10% sibs (F2)								
0 (ref)			8644 (0.85)		8644 (0.85)		8644 (0.85)	
1			1256 (0.13)	0.82 (0.72-0.92)	1256 (0.13)	0.83 (0.73-0.94)	1256 (0.13)	0.82 (0.73-0.93)
2+			214 (0.2)	0.72 (0.56-0.92)	214 (0.2)	0.76 (0.59-0.97)	214 (0.2)	0.75 (0.58-0.96)
LDS (F2)								
0 - non-religious(ref)								
1 – baptized							NA	NA
2 - baptized + endowment							NA	NA
3 - missing							NA	NA
Sibship size (F2)							10114 (6.34)	1.01 (1.00-1.02)
Birth cohort, years (F2)							10114 (1835)	0.99 (>0.99<1.00)
Sex (F2)								
Man (ref)							4776 (0.48)	
Women							5338 (0.52)	1.01 (0.96-1.06)
SES – OCC_1950 (F2)								
0 - High (ref)							67 (0.01)	
1							645 (0.06)	0.88 (0.68-1.14)
2							536 (0.05)	0.97 (0.75-1.27)
3							62 (0.01)	0.76 (0.53-1.10)
4							71 (0.01)	0.99 (0.70-1.40)
5							733 (0.07)	0.80 (0.62-1.04)
6							311 (0.03)	0.86 (0.65-1.13)
7							759 (0.08)	0.84 (0.65-1.10)
8							574 (0.06)	0.85 (0.65-1.11)
9 - Low							3656 (0.36)	0.83 (0.65-1.07)
999 - missing							2700 (0.26)	0.93 (0.72-1.20)

Means represent a mean for a continuous variable and a proportion for a categorical variable. Additional covariates are: age mom at birth, birth order, birth intervals (in years), twin birth. LDS: the church of Jesus Christ of latter-day saints (Mormon church), SES: socio-economic status, OCC: occupational coding scheme of 1950. P-values are estimated with cox regression.

Supplementary Table 8: Survival analysis for IP's by top 10% siblings among IPs without top 10% parents

		UPDB non-longevous parents	
	N (mean)	**HR (CI)**	**P-value**
Top 10% sibs of RP			
0 (ref.)	4639 (0.70)		
1	1473 (0.22)	0.85 (0.79-0.91)	<0.0001
2+	528 (0.8)	0.78 (0.67-0.90)	<0.0001
		LINKS non-longevous parents	
	N (mean)	**HR (CI)**	**P-value**
Top 10% sibs of RP			
0 (ref.)	6867 (0.87)		
1	886 (0.11)	0.78 (0.72-0.85)	<0.0001
2+	108 (0.2)	0.72 (0.53-0.99)	0.0429

Means represent a mean for a continuous variable and a proportion for a categorical variable. Additional covariates are: religion, sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, and twin birth. Here P-values were rounded to 4 digits. P-values are estimated with cox regression.

Supplementary Table 9: Survival analysis for IP's by top 10% fathers and mothers

			UPDB			LINKS	
		N (mean)	**HR (95% CI)**	**P-value**	**N (mean)**	**HR (95% CI)**	**P-value**
F1-F2 (IP)	Top 10% parents						
	0 Both parents NL	6334 (0.64)			7817 (0.77)		
	1 Pa LL / Ma NL	1653 (0.17)	0.88 (0.84-0.93)	<0.0001	1124 (0.11)	0.83 (0.78-0.89)	<0.0001
	2 Ma LL / Pa NL	1421 (0.15)	0.84 (0.79-0.89)	<0.0001	940 (0.09)	0.80 (0.75-0.86)	<0.0001
	3 Both parents LL	439 (0.4)	0.73 (0.66-0.80)	<0.0001	184 (0.03)	0.72 (0.62-0.83)	<0.0001
F2-F3 (full)	Top 10% parents						
	0 Both parents NL	38423 (0.64)			45644 (0.73)		
	1 Pa LL / Ma NL	10522 (0.17)	0.89 (0.86-0.92)	<0.0001	8643 (0.14)	0.86 (0.83-0.89)	<0.0001
	2 Ma LL / Pa NL	8969 (0.15)	0.86 (0.83-0.89)	<0.0001	6360 (0.10)	0.82 (0.79-0.86)	<0.0001
	3 Both parents LL	2328 (0.4)	0.72 (0.68-0.76)	<0.0001	1664 (0.03)	0.73 (0.67-0.79)	<0.0001

Means represent a mean for a continuous variable and a proportion for a categorical variable. Additional covariates are: religion, sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, and twin birth. PA=father, MA=mother, NL=non-longevous, LL=longevous, NL=non-longevous, longevous was defined as belonging to the top 10% of a persons' birth cohort. Here P-values were rounded to 4 digits. P-values are estimated with cox regression.

Supplementary Table 10: Sex specific survival analysis for IP's by top 10% parents

		UPDB			LINKS		
		N (mean)	HR (95% CI)	P-value	N (mean)	HR (95% CI)	P-value
F1-F2 (IP)	Top 10% parents (F1)						
	0	6640 (0.65)			7861 (0.78)		
	1	3167 (0.31)	0.85 (0.75-0.96)	0.0067	2096 (0.20)	0.83 (0.78-0.91)	<0.0001
	2	439 (0.4)	0.68 (0.59-0.80)	<0.0001	184 (0.2)	0.61 (0.48-0.76)	<0.0001
	Sex						
	Man (ref)	5053 (0.49)			4776 (0.48)		
	Women	5193 (0.51)	0.73 (0.68-0.78)	<0.0001	5338 (0.52)	1.02 (0.97-1.08)	0.3600
	Sex* Top 10% parents (F1)						
	0 Top 10 parents | women (ref)	6640 (0.65)			7861 (0.78)		
	1 Top 10 parents | women	3167 (0.31)	0.92 (0.78-1.10)	0.3440	2096 (0.20)	0.96 (0.86-1.10)	0.4041
	2 Top 10 parents | women	439 (0.4)	1.10 (0.89-1.35)	0.4010	184 (0.2)	1.36 (0.97-1.91)	0.0748
F2-F3 (Child of IP - one LL parent)	Top 10% parents (F1)						
	0	48619 (0.80)			53378 (0.85)		
	1	12179 (0.20)	0.83 (0.77-0.89)	<0.0001	9096 (0.15)	0.86 (0.78-0.94)	<0.0017
	Sex						
	Man (ref)	31258 (0.51)			32136 (0.52)		
	Women	29540 (0.49)	0.56 (0.55-0.58)	<0.0001	30338 (0.48)	0.64 (0.63-0.66)	<0.0001
	Sex* Top 10% parents (F1)						
	0 Top 10 parents | women (ref)	48619 (0.80)			53378 (0.85)		
	1 Top 10 parents | women	12179 (0.20)	1.03 (0.98-1.08)	0.2390	9096 (0.15)	1.00 (0.94-1.05)	0.1863
F2-F3 (Child of IP - two LL parents)	Top 10% parents (F1)						
	0	58672 (0.96)			60962 (0.97)		
	2	2126 (0.04)	0.68 (0.57-0.80)	<0.0001	1512 (0.03)	0.78 (0.63-0.96)	0.0138
	Sex						
	Man (ref)	31258 (0.51)			32136 (0.52)		
	Women	29540 (0.49)	0.56 (0.55-0.58)	<0.0001	30338 (0.48)	0.64 (0.63-0.66)	<0.0001
	Sex* Top 10% parents (F1)						
	0 Top 10 parents | women (ref)	58672 (0.96)			60962 (0.97)		
	2 Top 10 parents | women	2126 (0.04)	1.09 (0.99-1.21)	0.0812	1512 (0.03)	1.00 (0.87-1.13)	0.1947

Means represent a mean for a continuous variable and a proportion for a categorical variable. Additional covariates are: religion, sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, and twin birth. PA=father, MA=mother, NL=non-longevous, LL=longevous, NL=non-longevous, longevous was defined as belonging to the top 10% of a persons' birth cohort. Here P-values were rounded to 4 digits. P-values are estimated with cox regression.

Supplementary Table 11:

Frailty survival analysis for Children of IP's By top 10% aunt and uncles of children without top 10% parents

		UPDB non-longevous RP + non-longevous spouse	
	N (mean)	**HR (CI)**	**P-value**
Top 10% aunts and uncles			
0 (ref.)	26475 (0.67)		
1	9374 (0.24)	0.96 (0.93-0.99)	0.0170
2+	3516 (0.8)	0.90 (0.86-0.95)	<0.0001
		LINKS **non-longevous RP + non-longevous spouse**	
	N (mean)	**HR (CI)**	**P-value**
Top 10% aunts and uncles			
0 (ref.)	40031 (0.86)		
1	5651 (0.12)	0.95 (0.91-0.99)	0.0134
2+	890 (0.2)	0.81 (0.73-0.90)	<0.0001

Means represent a mean for a continuous variable and a proportion for a categorical variable. Additional covariates are: religion, sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, and twin birth. Here P-values were rounded to 4 digits. P-values are estimated with cox regression.

Supplementary Table 12: Selection criteria for the random sampling of F2 IPs

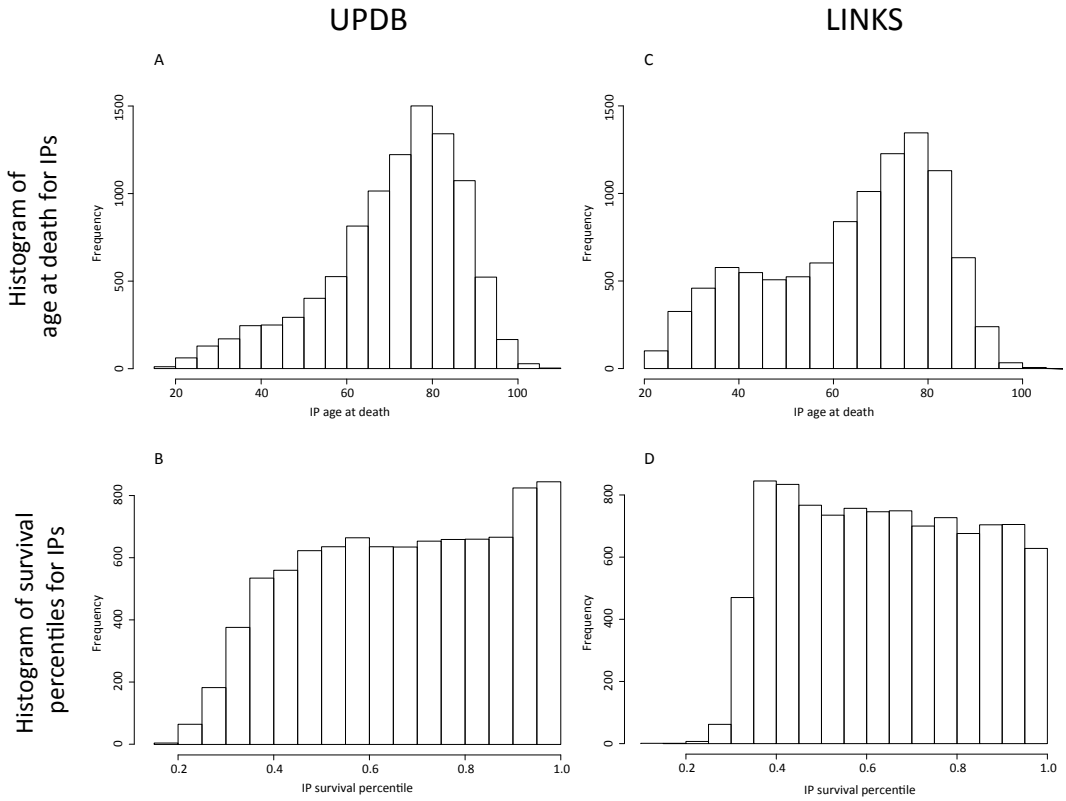
Selection	**Motivation**
At least 1 identifiable child	**Part of the study focused on children, hence children needed to be available for the index persons**
At least 1 identifiable sibling	To ensure that the influence of siblings could be analyzed we excluded families with only a single child. In addition this was a method to make sure that dummy families were excluded
A spouse should be identified	Part of the study focused on spouses who married to longevous index persons. For this, and because index persons needed to have a child, a spouse was required to be available
A known sex	To be able to distinguish between males and females, the sex of at least the index persons needed to be available
Availability of a birth date	To be able to study the survival of the complete group of index persons with the best possible data, a date of birth and a date of death needed to be available. In addition, in the LINKS data selecting on an available birth and death date was a quality check that made sure that index person was indeed part of the identified family.
Availability of a death	
In the UPDB: should be identified on a genealogy record	All genealogy records are verified. Hence, this was a double check to make sure that the index person was indeed part of the identified family.

UPDB: Utah Population Database, LINKS: LINKing System for historical family reconstruction.

Supplementary Table 13: Demographic spread for included UPDB persons

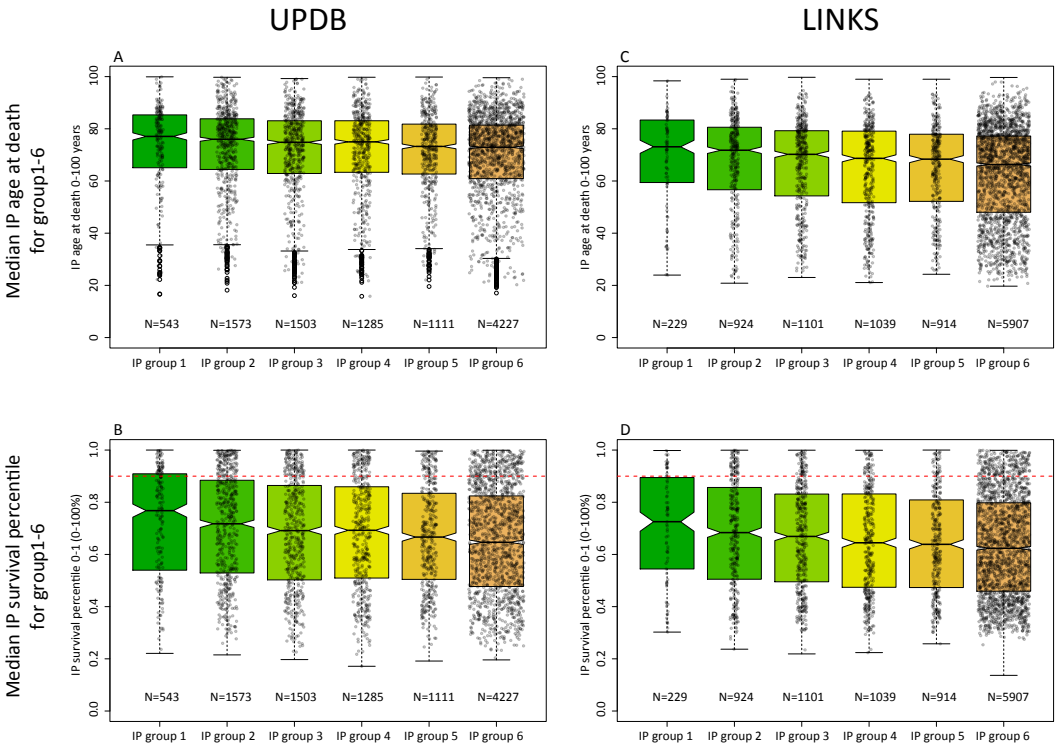
Role	**Fathers**	**Mothers**	**IPs**	**Siblings**	**Spouses**	**Children**
Lived in Utah (%)	7600 (80)	8478 (87)	9901 (97)	32026 (70)	9998 (97)	49839 (92)
Not lived in Utah (%)	1884 (20)	1229 (13)	345 (3)	13674 (30)	356 (3)	4236 (8)
Birth continent	**America**	**Europe**	**Asia**	**Africa**	**Australia**	**Other**
Lived in Utah	85141 (91)	11490 (73)	35 (65)	4 (100)	195 (73)	171 (34)
Not lived in Utah	8310 (9)	4363 (27)	19 (45)	0 (0)	73 (27)	329 (66)

Numbers are based on uncensored individuals. Numbers are based on the UPDB only.



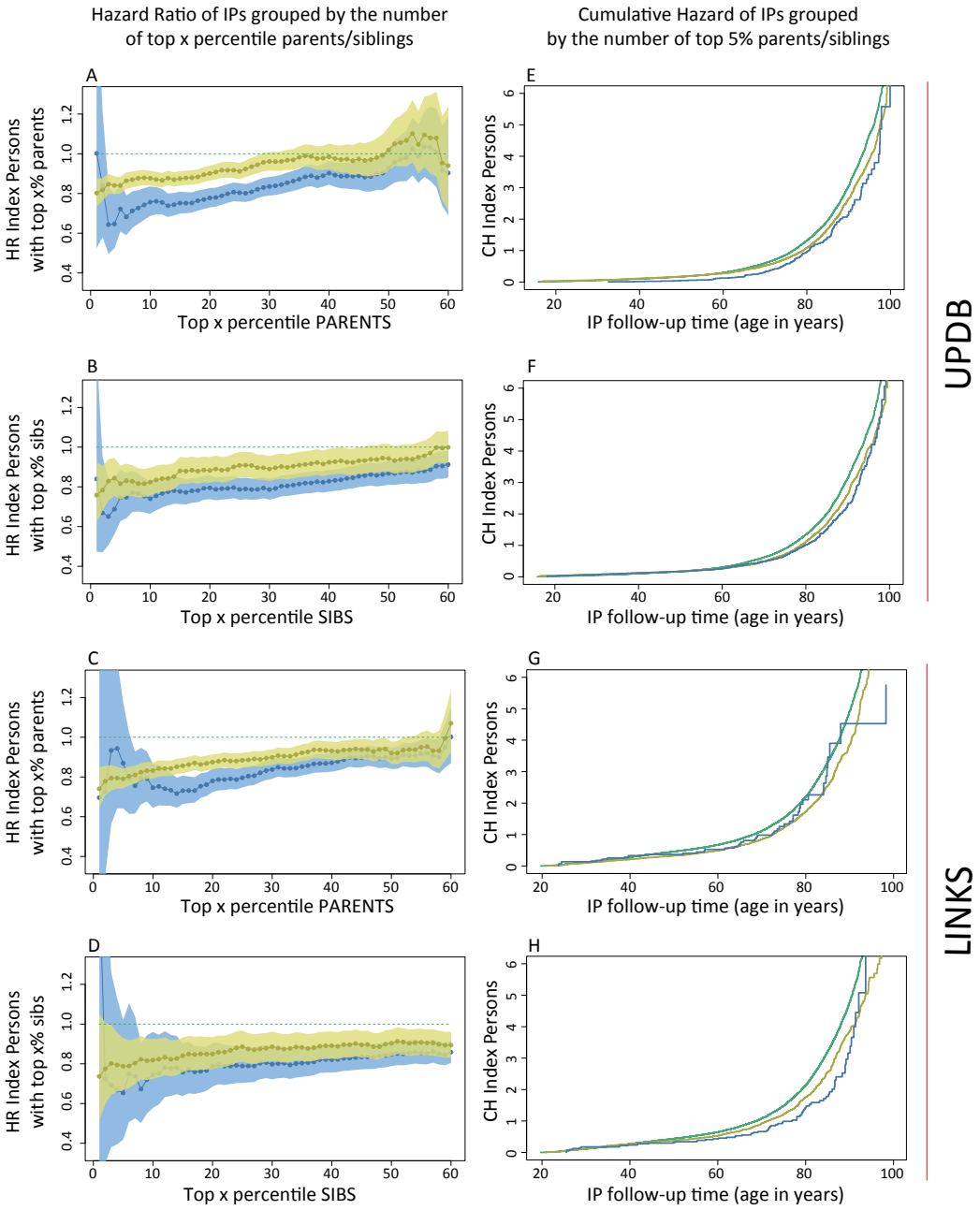
Supplementary Figure 1: Histogram of IP age at death and IP survival percentiles

Panel A and C depict a histogram for the IPs ages at death with the UPDB and LINKS data respectively. Panel B and D depict a histogram for the IPs survival percentiles with the UPDB and LINKS data respectively. IP=Index Person.



Supplementary Figure 2: Median age at death and survival percentile for IPs grouped by their parental and sibling survival in mutual exclusive groups

This figure relates to main Figure 3 and shows the median + quantiles and variation for IPs' age at death on the top row (panel A and C). The bottom row (panel B and D) shows the median + quantiles and variation for IPs' survival percentiles. Nodes are based on 1/6th of the total sample size for illustrative purposes. The red lines on the bottom row represent the cut-off for the top 10 percent surviving IPs for the different groups. Similar to the decrease in HR for the different groups illustrated in main Figure 2 and the increase in age at death or survival percentile for the different groups illustrated in this figure, there is an increase in top 10% surviving IPs. For the UPDB data 17% of the total number of IPs in group 6 belongs to the top 10% survivors, this is 19% for group 5, 23% for group 4, 26% for group 3, 29% for group 2, and 37% for group 1. For the LINKS data the numbers, in similar order, are 13%, 14%, 16%, 18%, 23% and 32%.

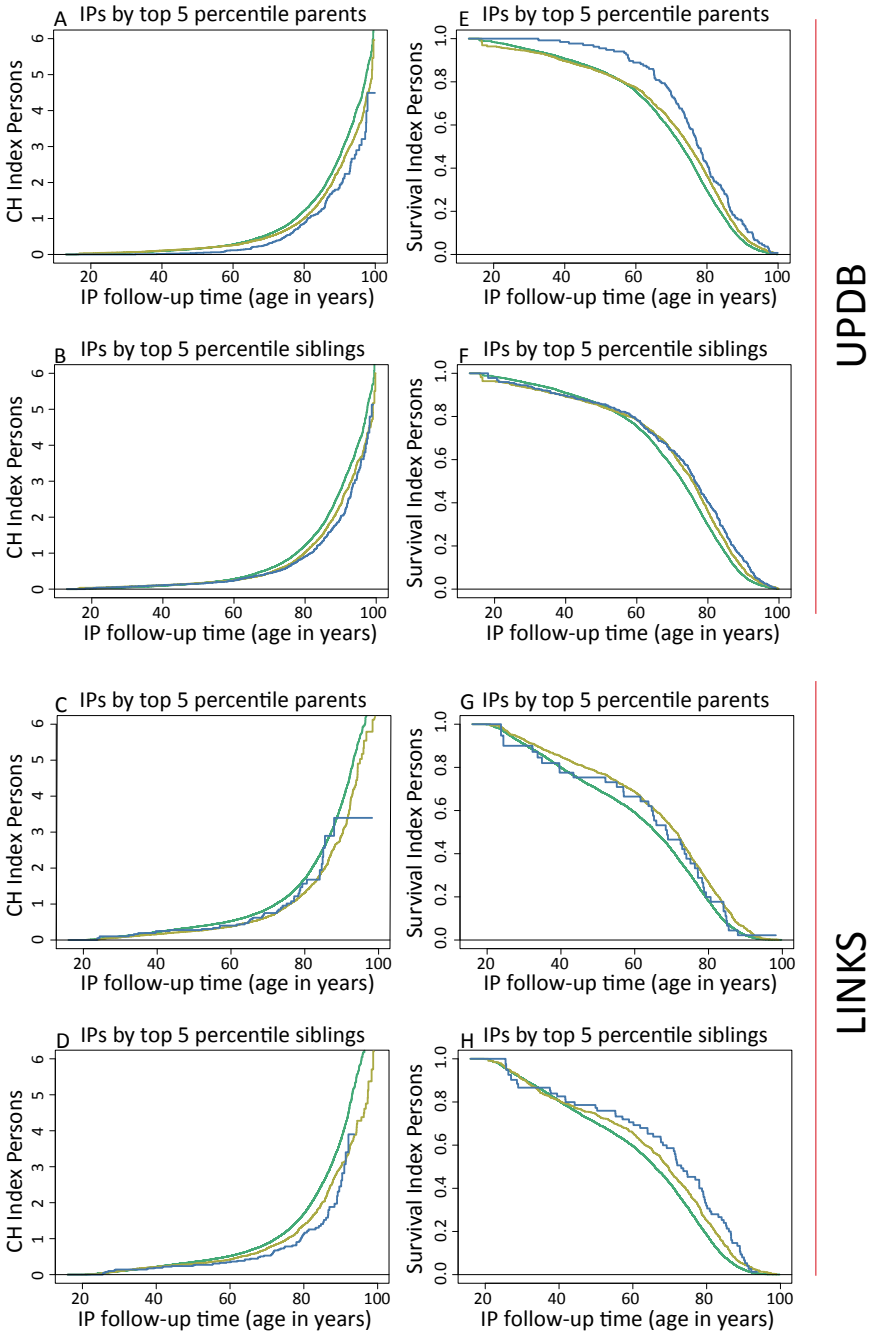


Supplementary Figure 3:

Survival of IPs with parents and siblings belonging to the 1st until 60th percentile survivors of their birth cohort

This figure depicts the Hazard Ratio (HR) for IPs (left column, panel A-D) with 1 and 2 parents or 1 and 2+ siblings belonging to the top x percentile (x = 1,2,3, ..., 60) of survivors of their birth cohort. The percentile groups (x-axis) are mutually inclusive, meaning that a first-degree family member who belonged to the top 1% also belonged to the top 5% etc. The figure also depicts the Cumulative Hazard (CH) for index persons (IPs, right column,

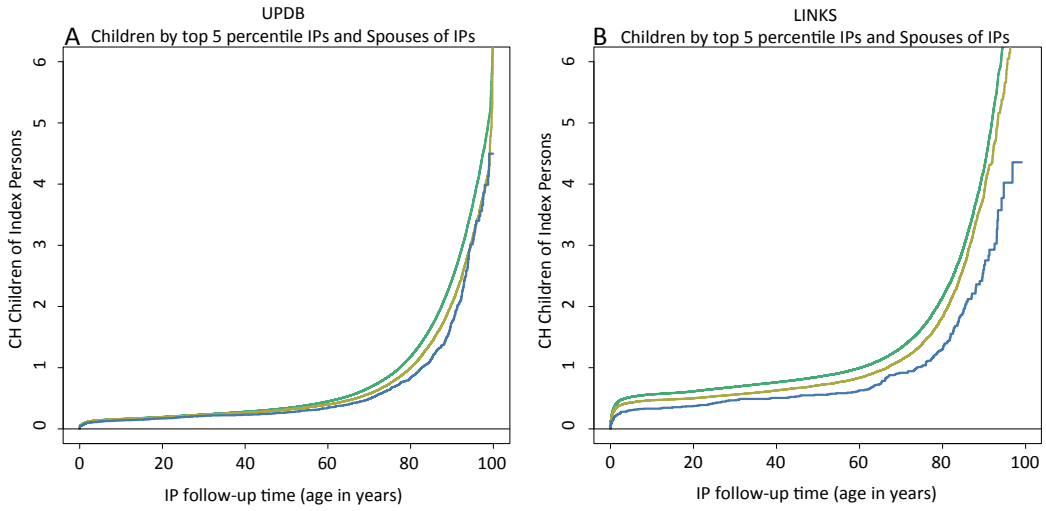
panel E-H) with 1 and 2 parents or 1 and 2+ siblings who belong to the top 5%. Green (dotted) lines present the reference group of 0 top x percentile parents or siblings, yellow lines represent 1 top x percentile parents or siblings, blue lines represent 2 or 2+ top x percentile siblings. Left column: x-axes represent the top x birth cohort based survival percentile, the y-axes represent the hazard ratio (HR) of dying for IPs having 1 and 2 or 2+ top x percentile parents or siblings compared to having 0 top x percentile parents or siblings. Right column: x-axes represent IP years of survival, y-axes represent the IPs' cumulative hazard of dying while having 1 and 2 or 2+ top 5th percentile parents or siblings compared to having 0 top 5th percentile parents or siblings. All estimates are adjusted for religion (UPDB only), sibship size, birth cohort, sex, socio-economic status, mother's age at birth, birth order, birth intervals, twin birth, and number of top 5% parents or number of top 5% siblings for the sibling and parent analyses respectively. Error bars represent confidence intervals.



Supplementary Figure 4:

Kaplan-Meier and Nelson-Aalen plots for IPs by the longevity of their parents and siblings at the top 5%

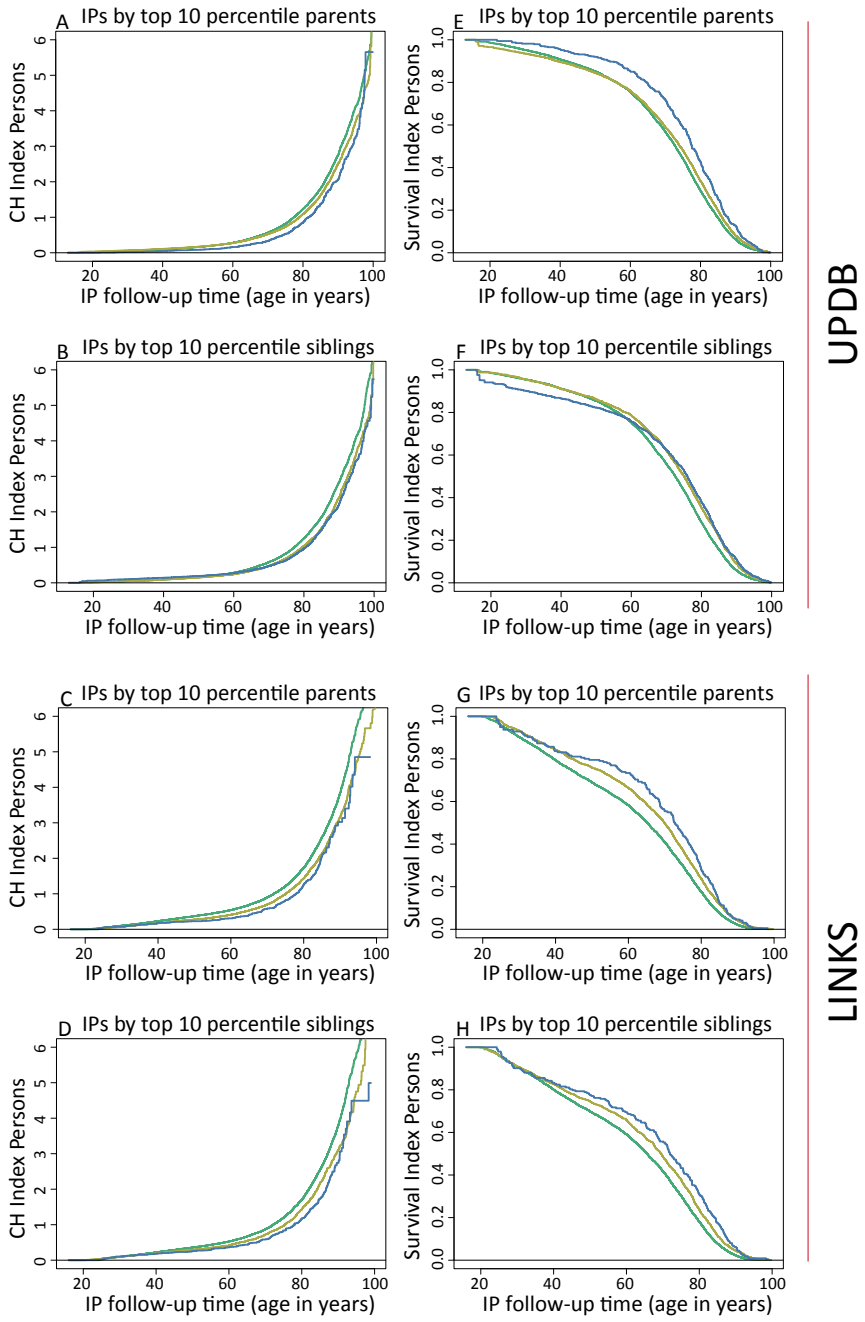
Green lines present 0 top x percentile parents or siblings, yellow lines represent 1 top x percentile parents or siblings, blue lines represent 2 or 2+ longevous parents or siblings. CH=Cumulative Hazard, IP=Index Person. Left column depicts the Kaplan-Meier curves, right column depicts the Nelson-Aalen curves. Panel A, B, E, and F represent UPDB IPs, panel C, D, G, and H represent LINKS IPs. CH=Cumulative Hazard, IP=Index Person.



Supplementary Figure 5:

Nelson-Aalen plots for children of IPs by the longevity of their parents (IPs and spouses of IPs) at the top 5%

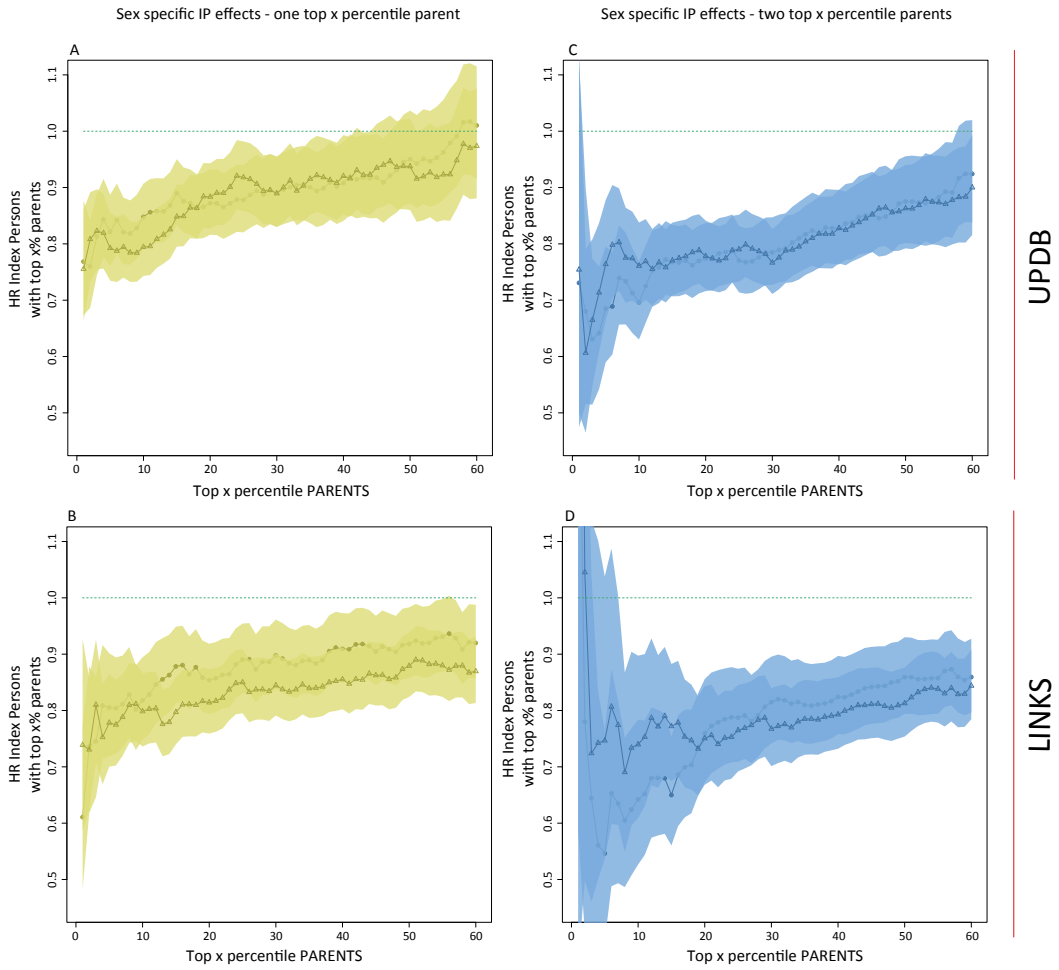
Green lines present 0 top x percentile parents or siblings, yellow lines represent 1 top x percentile parents or siblings, blue lines represent 2 or 2+ longevous parents or siblings. Panel A represents UPDB children of IPs and panel B represents LINKS children of IPs. CH=Cumulative Hazard, IP=Index Person.



Supplementary Figure 6:

Kaplan Meier and Nelson-Aalen plots for IPs by the longevity of their parents and siblings at the top 10%

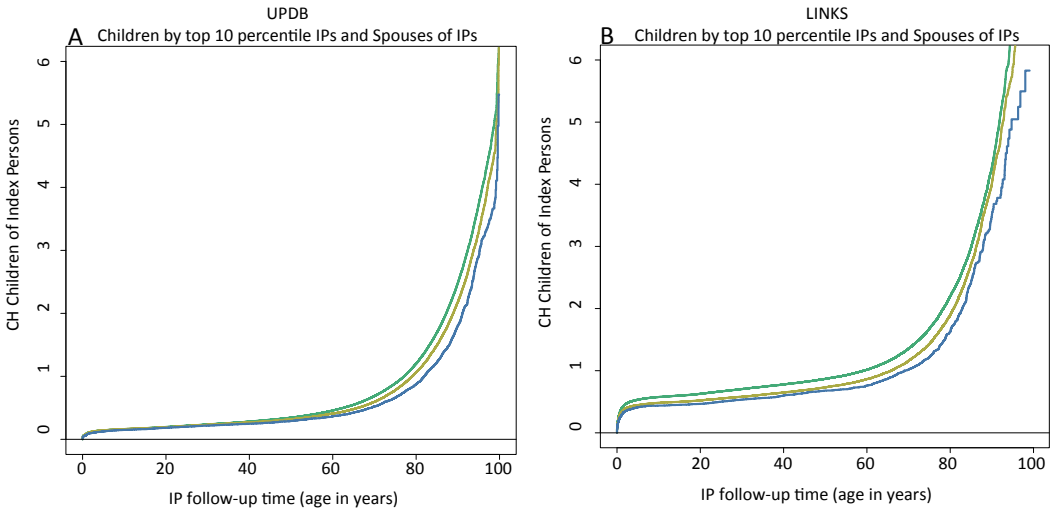
Green lines present 0 top x percentile parents or siblings, yellow lines represent 1 top x percentile parents or siblings, blue lines represent 2 or 2+ longevous parents or siblings. CH=Cumulative Hazard, IP=Index Person. Left column depicts the Kaplan-Meier curves, right column depicts the Nelson-Aalen curves. Panel A, B, E, and F represent UPDB IPs, panel C, D, G, and H represent LINKS IPs. CH=Cumulative Hazard, IP=Index Person.



Supplementary Figure 7:

Sex specific survival of IPs with parents belonging to the 1st until 60th percentile survivors of their birth cohort

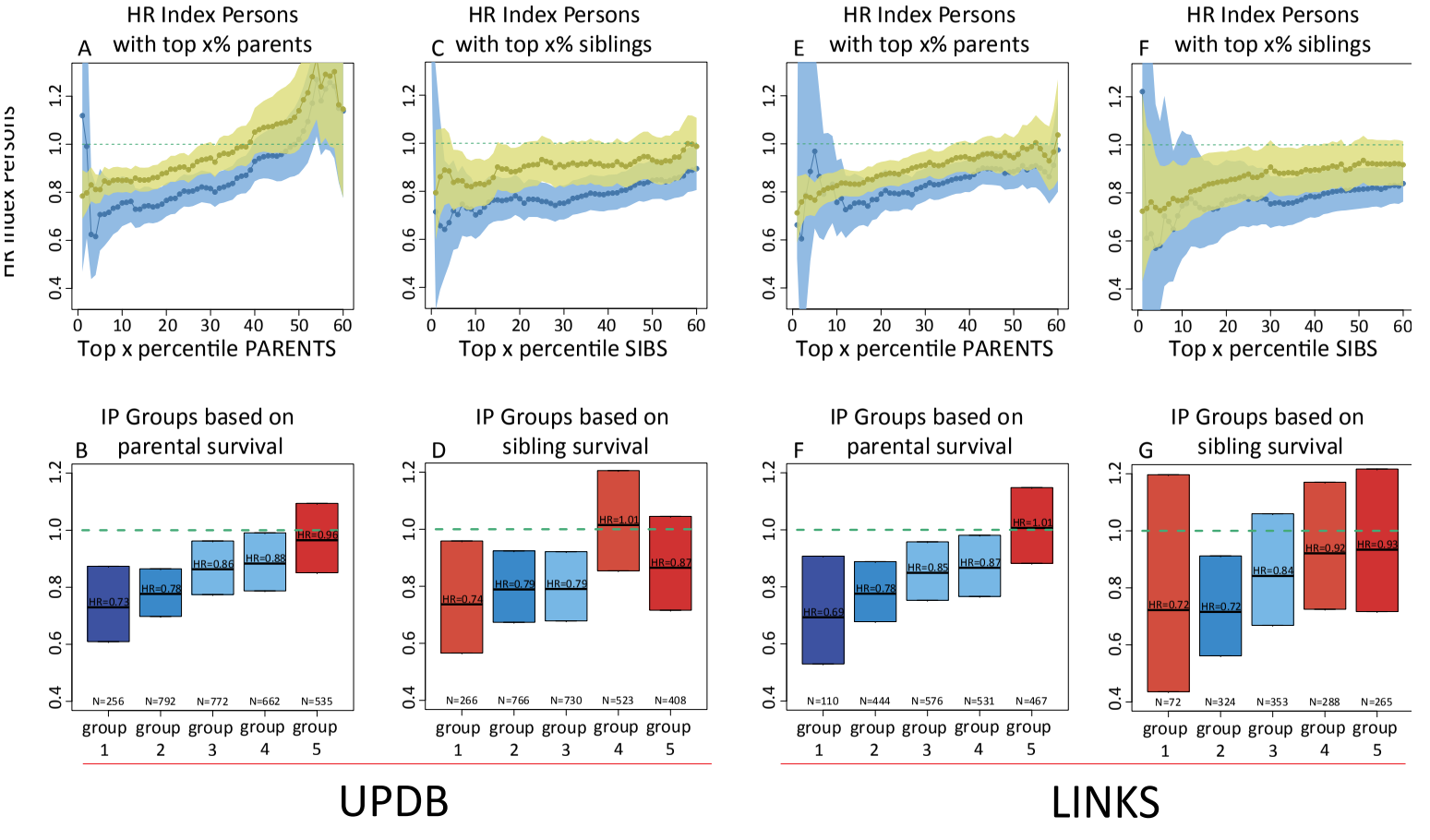
Green lines present 0 top x percentile parents, yellow lines represent 1 top x percentile parents, blue lines represent 2 or 2+ longevous parents. Round nodes=males, triangle nodes=females. Top row (panel A and C) represent UPDB IPs, bottom row (panel B and D) represents LINKS IPs. HR=Hazard Ratio, IP=Index Person. Error bars represent confidence intervals.



Supplementary Figure 8:

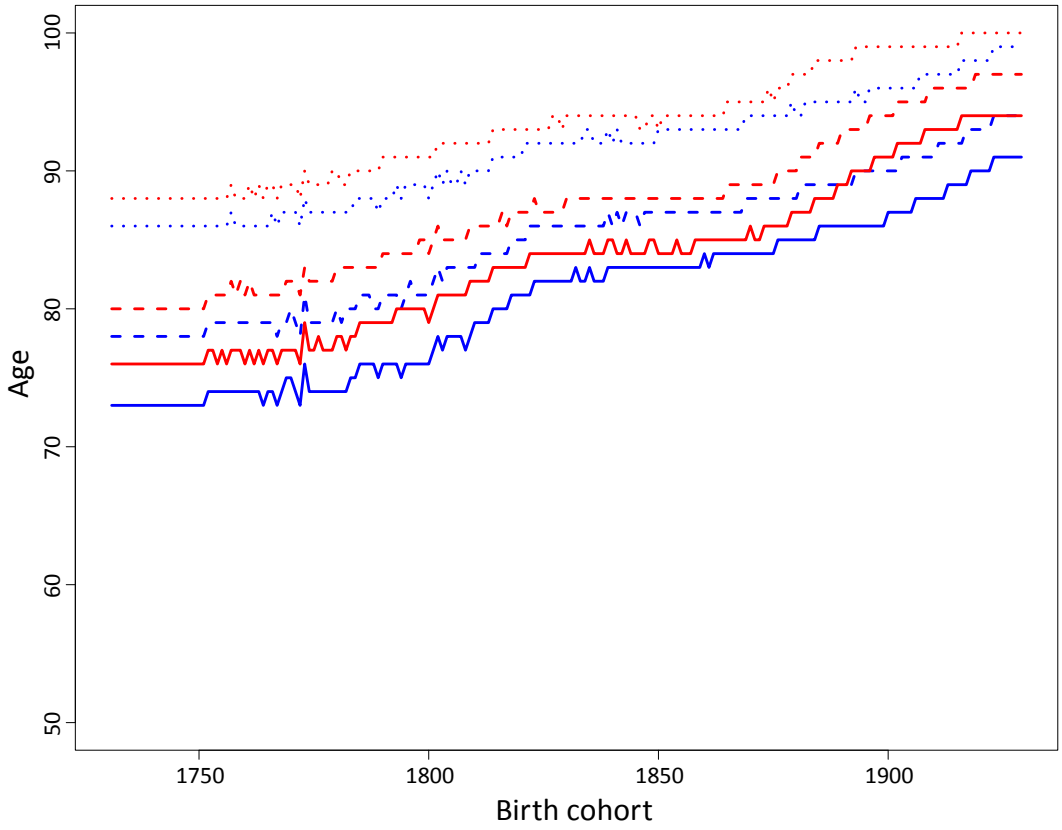
Nelson-Aalen plots for children of IPs by the longevity of their parents (IPs and spouses of IPs) at the top 10%

Green lines present 0 top x percentile parents or siblings, yellow lines represent 1 top x percentile parents or siblings, blue lines represent 2 or 2+ longevous parents or siblings. Panel A represents UPDB children of IPs and panel B represents LINKS children of IPs. CH=Cumulative Hazard, IP=Index Person.



Supplementary Figure 9: Results represented by main Figure 2 and 3 with the IP numbers split in half

Figure illustrates results similar to main Figure 2 and 3 with the number of IPs cut in half. Top row: green lines present 0 top x percentile parents or siblings, yellow lines represent 1 top x percentile parents or siblings, blue lines represent 2 or 2+ longevous parents or siblings. Bottom row: green lines represent the reference category, which is group 6. Top column (Panel A, C, D, and E) represents Figure 2, bottom column (Panel B, D, F, and G) represents Figure 3. HR=Hazard Ratio, IP=Index Person. Error bars represent confidence intervals.



Supplementary Figure 10: UPDB and LINKS birth cohorts mapping of age by top 1, 5, and 10th percentile

This figure represents the the percentile-age pairings from the Swedish lifetables used to calculate survival percentiles in both the UPDB and LINKS datasets. Line colors: Blue: men, Red: women. Line patterns: Dotted lines represent the top 1% survivors of the specific birth cohorts. Broken lines represent the top 5% survivors of the specific birth cohorts. Unbroken lines represent the top 10% survivors of the specific birth cohorts.

Chapter 5

Table A1: Lifetable conditioning by group

Group	Conditioning (age)	Reason
Parents (F0)	**First death in the group, 28 years**	**Parents are indirectly selected to have reached an age to be able to have at least 2 offspring that could be enrolled as participants**
Participants (F1)	Age at inclusion, differs per individual	Participants are directly selected to be alive until the age of inclusion
Siblings (F1)	No conditioning was applied	Siblings are not directly or indirectly selected
Spouses (F1)	First death in the group, 28 years	Spouses are indirectly selected to be old enough to marry

Table A2: Exact SMR values, CIs and N connected to Figure 3

Age	Parents	N parents	Siblings	N siblings	Spouses	N spouses
1			**0.69 (0.65-0.73)**	**2269**		
2			0.66 (0.64-0.7)	2075		
3			0.67 (0.64-0.7)	2058		
4			0.67 (0.64-0.7)	2044		
5			0.67 (0.64-0.7)	2037		
6			0.67 (0.64-0.7)	2035		
7			0.67 (0.64-0.7)	2028		
8			0.67 (0.64-0.7)	2028		
9			0.67 (0.64-0.7)	2025		
10			0.67 (0.64-0.7)	2022		
11			0.67 (0.64-0.7)	2021		
12			0.67 (0.64-0.7)	2016		
13			0.67 (0.64-0.7)	2016		
14			0.67 (0.64-0.7)	2014		
15			0.67 (0.64-0.7)	2012		
16			0.67 (0.64-0.7)	2010		
17			0.67 (0.64-0.7)	2009		
18			0.67 (0.64-0.7)	2004		
19			0.67 (0.64-0.7)	2000		
20			0.67 (0.64-0.7)	1998		
21			0.67 (0.64-0.7)	1992		
22			0.67 (0.64-0.7)	1987		
23			0.67 (0.64-0.7)	1984		
24			0.67 (0.64-0.7)	1981		

Age	Parents	N parents	Siblings	N siblings	Spouses	N spouses
25			0.67 (0.64-0.7)	1977		
26			0.67 (0.64-0.7)	1972		
27			0.66 (0.63-0.69)	1965		
28	0.68 (0.64-0.72)	820	0.66 (0.63-0.69)	1960	1.02 (0-1.08)	690
29	0.68 (0.64-0.72)	819	0.66 (0.63-0.69)	1956	1.02 (0-1.09)	685
30	0.68 (0.65-0.72)	819	0.66 (0.63-0.69)	1953	1.02 (0.95-1.09)	685
31	0.69 (0.65-0.73)	819	0.66 (0.63-0.69)	1950	1.02 (0.95-1.09)	685
32	0.69 (0.65-0.73)	819	0.66 (0.63-0.69)	1947	1.02 (0.95-1.09)	683
33	0.69 (0.65-0.73)	818	0.66 (0.63-0.69)	1945	1.03 (0.96-1.1)	681
34	0.69 (0.65-0.73)	817	0.66 (0.63-0.69)	1941	1.03 (0.96-1.1)	681
35	0.69 (0.65-0.73)	815	0.66 (0.63-0.69)	1937	1.03 (0.97-1.1)	679
36	0.69 (0.65-0.73)	813	0.66 (0.63-0.69)	1933	1.03 (0.97-1.1)	676
37	0.69 (0.66-0.74)	811	0.66 (0.63-0.69)	1930	1.03 (0.97-1.1)	675
38	0.69 (0.66-0.74)	809	0.66 (0.63-0.69)	1927	1.03 (0.97-1.1)	675
39	0.7 (0.66-0.74)	808	0.66 (0.63-0.69)	1923	1.03 (0.97-1.1)	674
40	0.7 (0.66-0.74)	807	0.66 (0.63-0.69)	1919	1.03 (0.98-1.11)	673
41	0.7 (0.66-0.74)	804	0.66 (0.63-0.69)	1915	1.04 (0.98-1.11)	670
42	0.7 (0.66-0.74)	800	0.66 (0.63-0.69)	1909	1.04 (0.98-1.11)	669
43	0.7 (0.66-0.74)	798	0.66 (0.63-0.69)	1906	1.04 (0.98-1.11)	667
44	0.7 (0.66-0.74)	792	0.66 (0.63-0.69)	1901	1.04 (0.98-1.11)	666
45	0.69 (0.66-0.74)	788	0.66 (0.63-0.69)	1898	1.04 (0.98-1.11)	663
46	0.69 (0.66-0.74)	783	0.66 (0.63-0.69)	1891	1.04 (0.98-1.11)	660
47	0.69 (0.66-0.74)	780	0.66 (0.63-0.69)	1885	1.04 (0.98-1.11)	660
48	0.69 (0.66-0.74)	778	0.66 (0.63-0.69)	1884	1.04 (0.98-1.11)	659
49	0.69 (0.66-0.74)	774	0.66 (0.63-0.69)	1880	1.04 (0.98-1.11)	655
50	0.69 (0.66-0.74)	769	0.66 (0.63-0.69)	1873	1.05 (0.99-1.12)	651
51	0.69 (0.65-0.74)	763	0.66 (0.63-0.69)	1866	1.04 (0.99-1.12)	648
52	0.69 (0.65-0.74)	760	0.66 (0.63-0.69)	1860	1.04 (0.98-1.12)	642
53	0.69 (0.65-0.74)	755	0.66 (0.63-0.69)	1856	1.05 (0.98-1.12)	640
54	0.69 (0.66-0.74)	752	0.66 (0.63-0.69)	1848	1.04 (0.98-1.12)	633
55	0.7 (0.66-0.74)	748	0.66 (0.63-0.69)	1843	1.05 (0.98-1.12)	627
56	0.7 (0.66-0.74)	743	0.66 (0.63-0.69)	1835	1.05 (0.98-1.12)	623
57	0.7 (0.66-0.74)	739	0.66 (0.63-0.69)	1828	1.05 (0.98-1.12)	619
58	0.7 (0.66-0.74)	734	0.66 (0.63-0.69)	1820	1.05 (0.99-1.13)	614
59	0.7 (0.66-0.75)	728	0.66 (0.63-0.69)	1809	1.06 (0.99-1.13)	609
60	0.7 (0.66-0.75)	722	0.67 (0.64-0.7)	1804	1.06 (1-1.14)	593
61	0.7 (0.66-0.75)	713	0.66 (0.63-0.7)	1786	1.07 (1-1.14)	584

Age	Parents	N parents	Siblings	N siblings	Spouses	N spouses
62	0.7 (0.66-0.75)	706	0.67 (0.64-0.7)	1776	1.06 (0.99-1.14)	571
63	0.7 (0.66-0.75)	695	0.67 (0.64-0.7)	1762	1.06 (0.99-1.14)	563
64	0.7 (0.66-0.74)	683	0.67 (0.64-0.7)	1744	1.06 (0.99-1.13)	551
65	0.7 (0.66-0.74)	672	0.67 (0.64-0.7)	1722	1.06 (0.99-1.14)	542
66	0.69 (0.65-0.74)	654	0.67 (0.63-0.7)	1699	1.06 (0.99-1.15)	530
67	0.69 (0.65-0.74)	642	0.67 (0.63-0.7)	1674	1.07 (1-1.16)	512
68	0.69 (0.65-0.74)	628	0.67 (0.64-0.71)	1654	1.08 (1.01-1.16)	494
69	0.68 (0.65-0.73)	608	0.67 (0.64-0.7)	1624	1.08 (1-1.16)	483
70	0.69 (0.65-0.74)	597	0.67 (0.64-0.71)	1601	1.07 (1-1.16)	462
71	0.69 (0.64-0.74)	581	0.67 (0.64-0.71)	1571	1.09 (1.02-1.17)	440
72	0.68 (0.64-0.73)	562	0.68 (0.64-0.71)	1544	1.08 (1.01-1.17)	425
73	0.68 (0.64-0.74)	547	0.68 (0.64-0.71)	1506	1.08 (1-1.16)	405
74	0.68 (0.64-0.73)	526	0.68 (0.65-0.72)	1471	1.09 (1.01-1.17)	390
75	0.69 (0.65-0.74)	512	0.68 (0.65-0.72)	1425	1.09 (1.02-1.18)	372
76	0.68 (0.64-0.73)	486	0.69 (0.65-0.72)	1382	1.12 (1.03-1.21)	356
77	0.69 (0.64-0.74)	469	0.69 (0.65-0.73)	1330	1.13 (1.05-1.24)	331
78	0.7 (0.65-0.76)	452	0.69 (0.65-0.73)	1273	1.17 (1.07-1.27)	300
79	0.7 (0.65-0.76)	428	0.7 (0.66-0.74)	1214	1.17 (1.08-1.29)	279
80	0.7 (0.65-0.76)	404	0.7 (0.66-0.74)	1147	1.17 (1.07-1.29)	249
81	0.71 (0.66-0.78)	383	0.7 (0.67-0.74)	1076	1.2 (1.1-1.34)	223
82	0.73 (0.67-0.79)	363	0.71 (0.67-0.76)	1001	1.2 (1.09-1.34)	192
83	0.72 (0.67-0.79)	330	0.73 (0.69-0.78)	940	1.22 (1.1-1.38)	169
84	0.73 (0.67-0.8)	301	0.73 (0.68-0.77)	837	1.22 (1.09-1.37)	135
85	0.75 (0.68-0.83)	279	0.72 (0.68-0.77)	740	1.27 (1.13-1.45)	112
86	0.75 (0.68-0.83)	249	0.74 (0.69-0.8)	667	1.21 (1.07-1.39)	87
87	0.78 (0.71-0.88)	228	0.78 (0.71-0.84)	589	1.26 (1.12-1.47)	69
88	0.8 (0.72-0.9)	202	0.8 (0.73-0.86)	508	1.24 (1.07-1.44)	49
89	0.77 (0.69-0.88)	165	0.78 (0.71-0.86)	409	1.31 (1.12-1.57)	32
90	0.73 (0.63-0.83)	129	0.83 (0.75-0.92)	350	1.32 (1.09-1.6)	22
91	0.74 (0.63-0.85)	108	0.86 (0.78-0.96)	292	1.25 (0.99-1.55)	10
92	0.75 (0.64-0.87)	90	0.8 (0.71-0.9)	221	1.3 (0.97-1.73)	5
93	0.74 (0.61-0.87)	71	0.82 (0.71-0.94)	175	1.23 (0.78-1.86)	NA
94	0.73 (0.59-0.88)	55	0.75 (0.64-0.88)	125	1.23 (0.63-2.05)	NA
95	0.82 (0.66-1.03)	48	0.78 (0.65-0.93)	93		NA
96	0.94 (0.72-1.21)	40	0.75 (0.6-0.91)	65		NA
97	0.83 (0.54-1.14)	25	0.72 (0.5-0.97)	44		NA
98	0.8 (0.4-1.22)	17	0.82 (0.52-1.19)	34		NA
99	0.44 (0-1.04)	10	0.42 (0.03-0.8)	18		NA

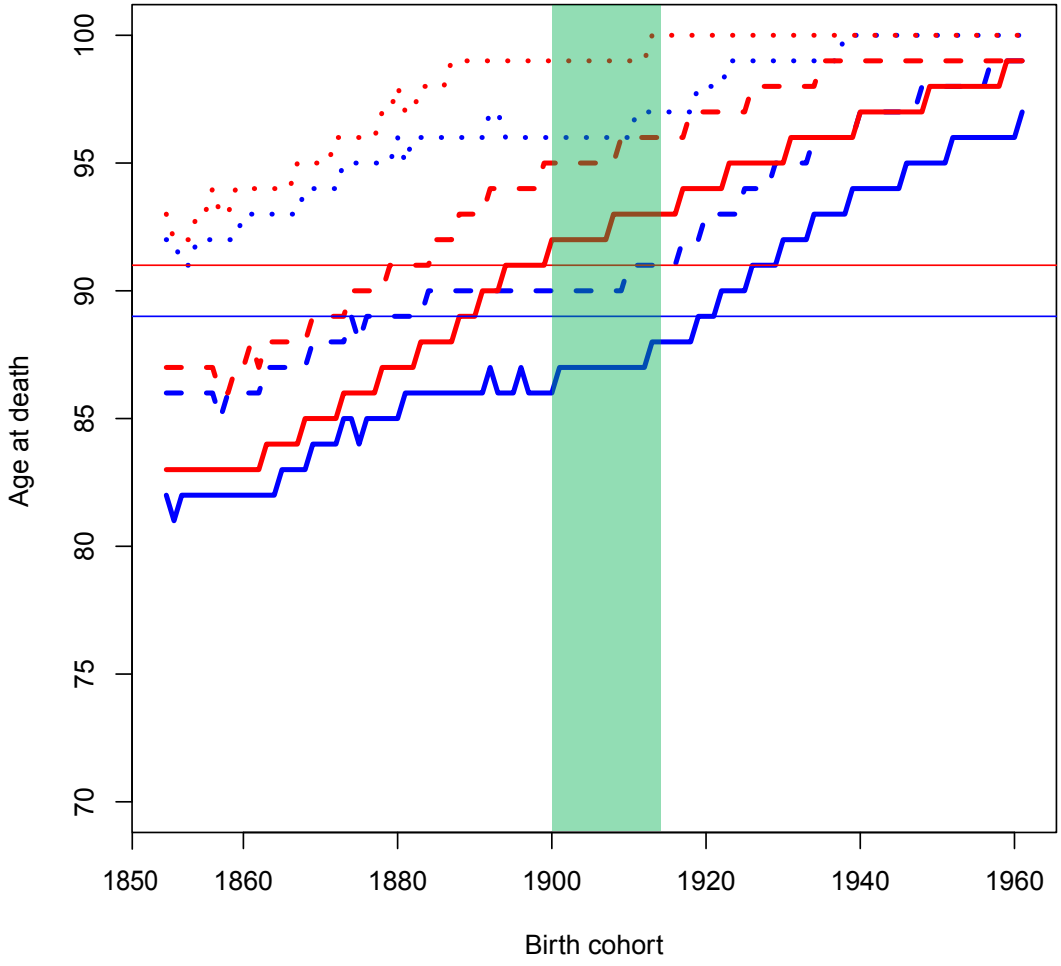


Figure A1: LLS birth cohorts top 10, 5, 1 percentile by participants

The horizontal lines represent the minimal inclusion criteria for the LLS participants (men >=89 and women >=91). The green rectangle represents the birth cohorts for the LLS participants (1900 - 1916). Line colors: Blue: men, Red: women. Line patterns: Dotted lines represent the top 1% survivors of the specific birth cohorts, broken lines represent the top 5% survivors of the specific birth cohorts, unbroken lines represent the top 10% survivors of the specific birth cohorts.

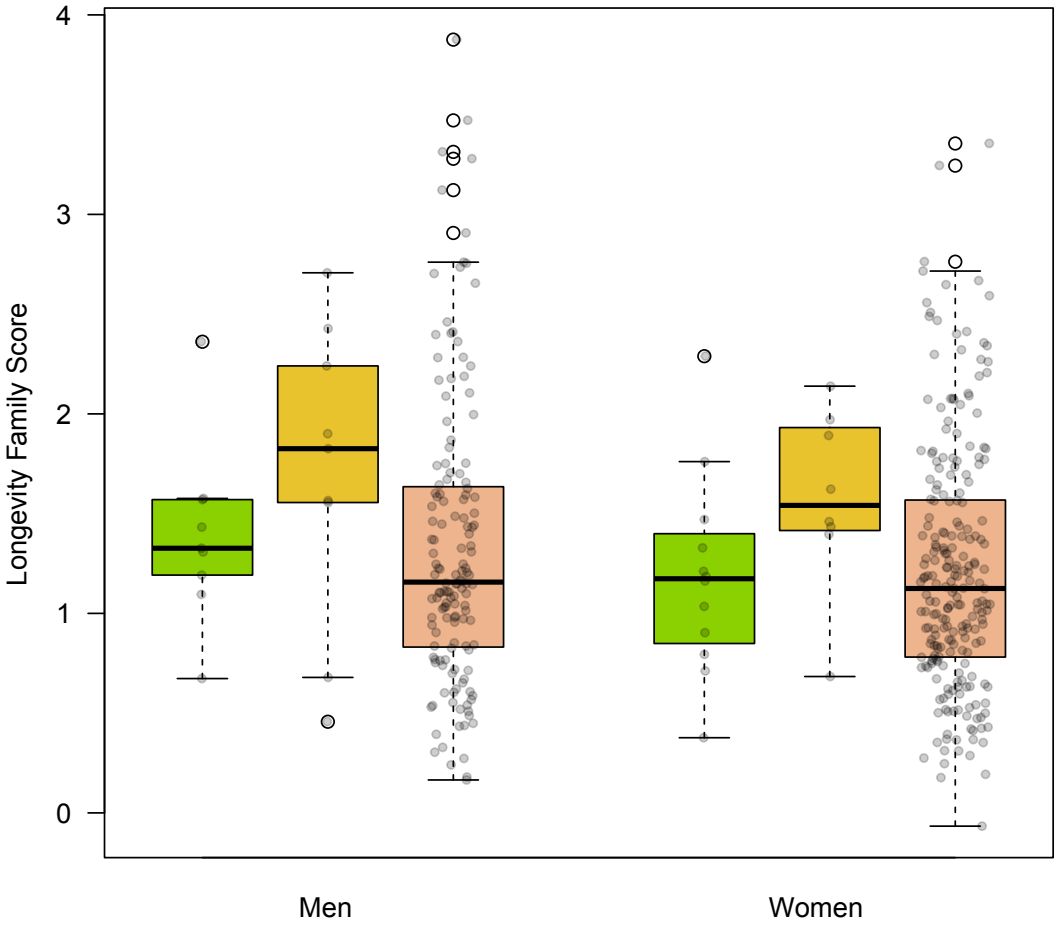


Figure A2: Median sex specific Longevity family score per sibship with one or none long-lived parent

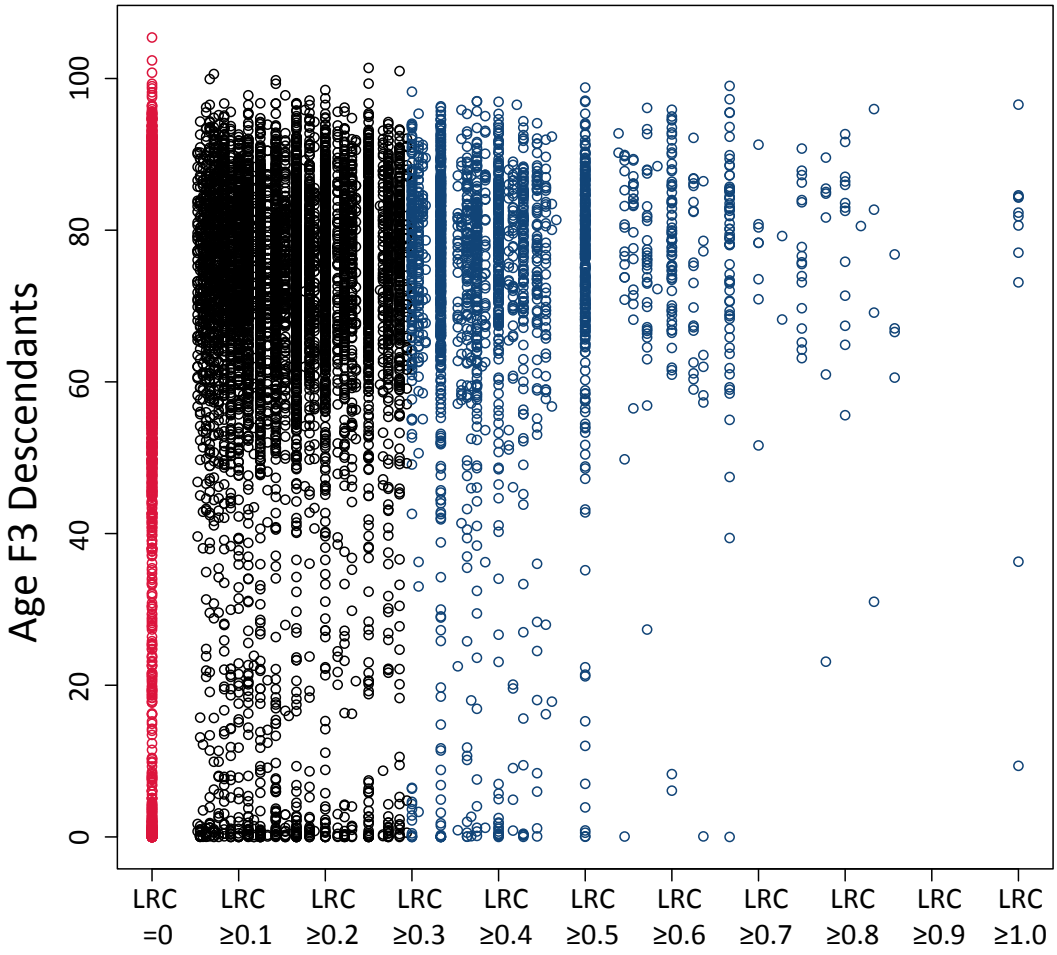
Each gray dot represents a complete sibship. Green boxplot represents the group of sibships with long-lived father and a non-long-lived mother (N=21). Orange boxplot represents the group of sibships with a long-lived mother and a non-long-lived father (N=17). Light brown boxplot represents the group of sibships with both parents not long-lived (N=371).

Chapter 6

Supplementary Table 1: Increase in excess survival with an increase in parental survival percentile

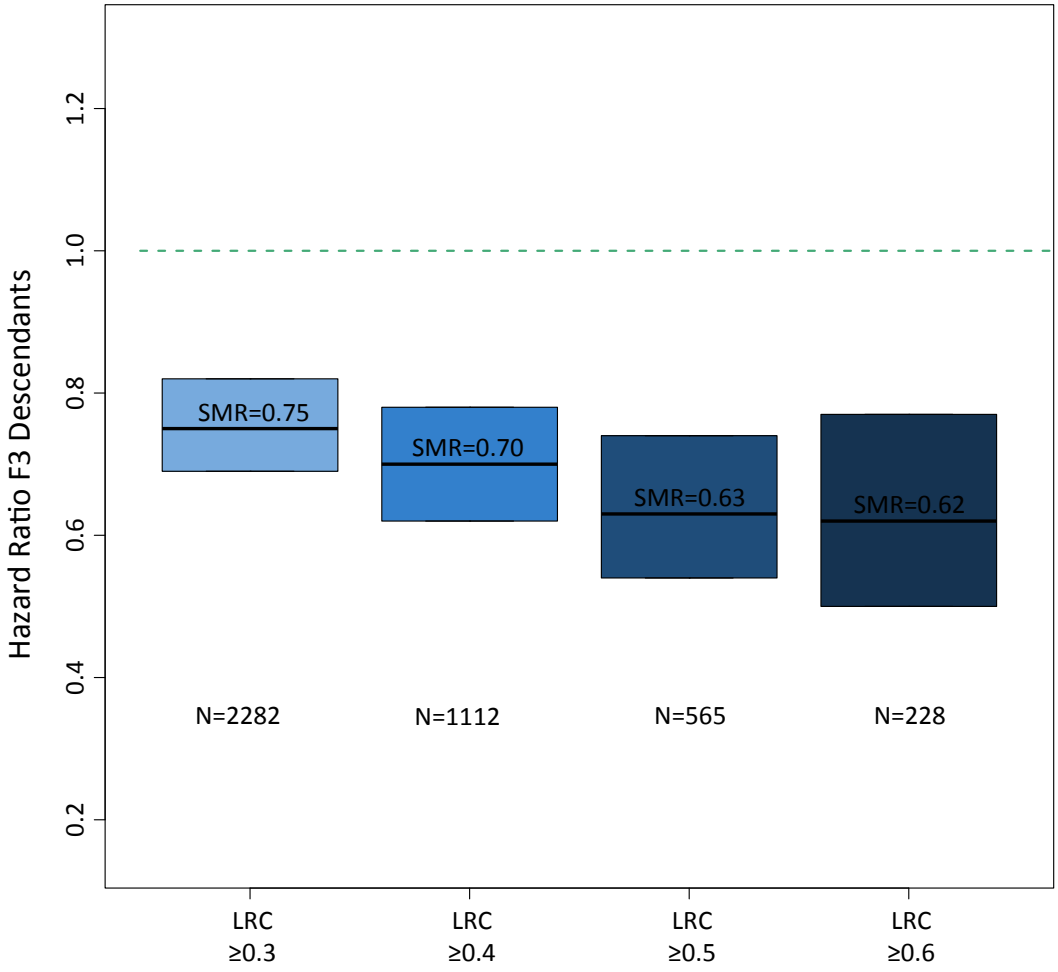
Percentile threshold	SMR (CI)	Number (N)
F2 children (original design - case)		
No parental selection	0.87 (0.84-0.89)	4416
≥85% surviving parent	0.85 (0.83-0.88)	3770
≥90% surviving parent	0.85 (0.82-0.88)	2565
≥95% surviving parent	0.86 (0.81-0.9)	1336
≥99% surviving parent	0.80 (0.7-0.9)	239
≥99.5% surviving parent	0.77 (0.65-0.9)	130
F3 children (original design - case)		
No parental selection	0.86 (0.84-0.89)	9010
≥50% surviving parent	0.84 (0.82-0.87)	7476
≥60% surviving parent	0.83 (0.81-0.86)	6357
≥70% surviving parent	0.81 (0.78-0.84)	5144
≥80% surviving parent	0.77 (0.74-0.81)	3434
≥85% surviving parent	0.76 (0.72-0.8)	2639
≥90% surviving parent	0.71 (0.67-0.76)	1813
≥95% surviving parent	0.69 (0.63-0.76)	798
≥99% surviving parent	0.65 (0.49-0.85)	78
F3 children (original design - control)		
No selection	0.96 (0.93-1)	4353
≥50% surviving parent	0.95 (0.91-0.99)	3425
≥60% surviving parent	0.92 (0.87-0.96)	2828
≥70% surviving parent	0.89 (0.84-0.93)	2168
≥80% surviving parent	0.85 (0.79-0.91)	1479
≥85% surviving parent	0.84 (0.78-0.91)	1107
≥90% surviving parent	0.84 (0.76-0.92)	772
≥95% surviving parent	0.77 (0.67-0.89)	317
≥99% surviving parent	0.94 (0.62-1.38)	41

Surviving parent refers to the proband IP for the F2 children. It refers to having at least 1 parent belonging to the specified survival percentile threshold for the F3 children. Estimates are only based on the proband line and not on the spousal line.



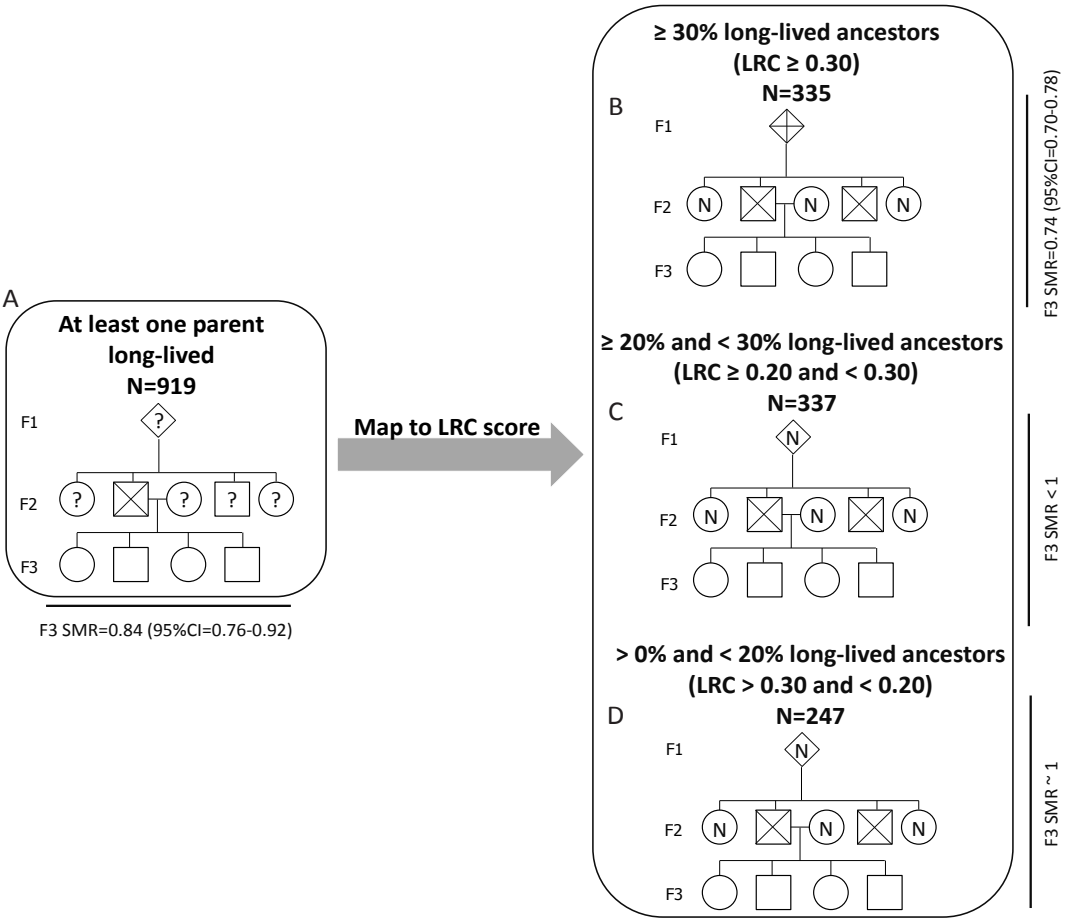
Supplementary Figure 1: Family cases and controls distributed by their reached age

The x-axis represents the Longevity Relatives Count (LRC) scores of the F3 descendants. The y-axis represents the attained age of the F3 descendants. The attained age can either be an age at death or an age at last observation. Around 50% of the F3 descendants is still alive (Table 1). The red color shows the family controls (those without any long-lived ancestors) and the blue color shows the family cases (those with an LRC ≥0.30).



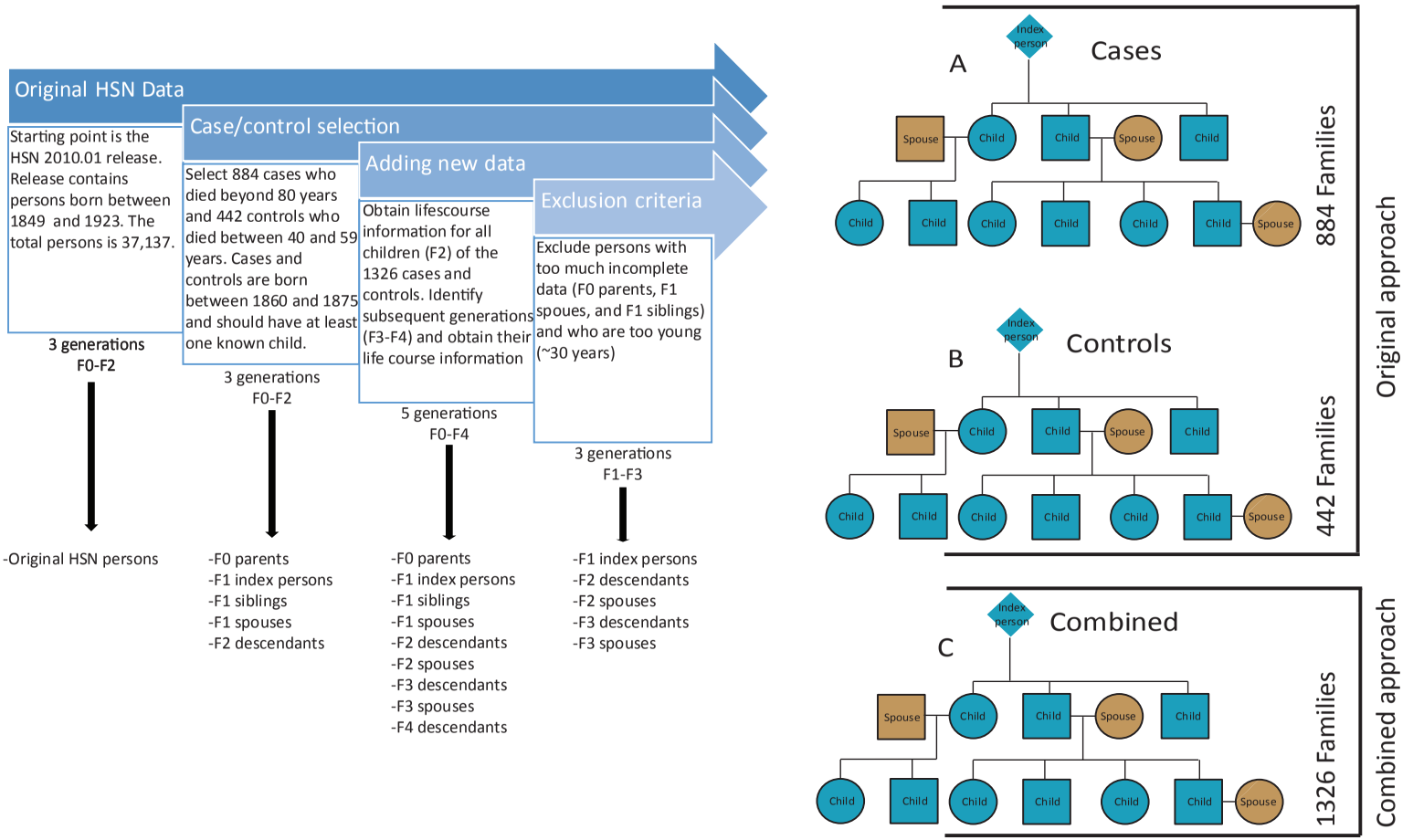
Supplementary Figure 2: SMR of F3 descendants with an increasing proportion of long-lived ancestors

The x-axis represents the Longevity Relatives Count (LRC) score for the F3 descendants. The y-axis shows the Hazard Ratios (HRs) for the F3 descendants. The stronger the effect (lower HR) the darker the blue color.



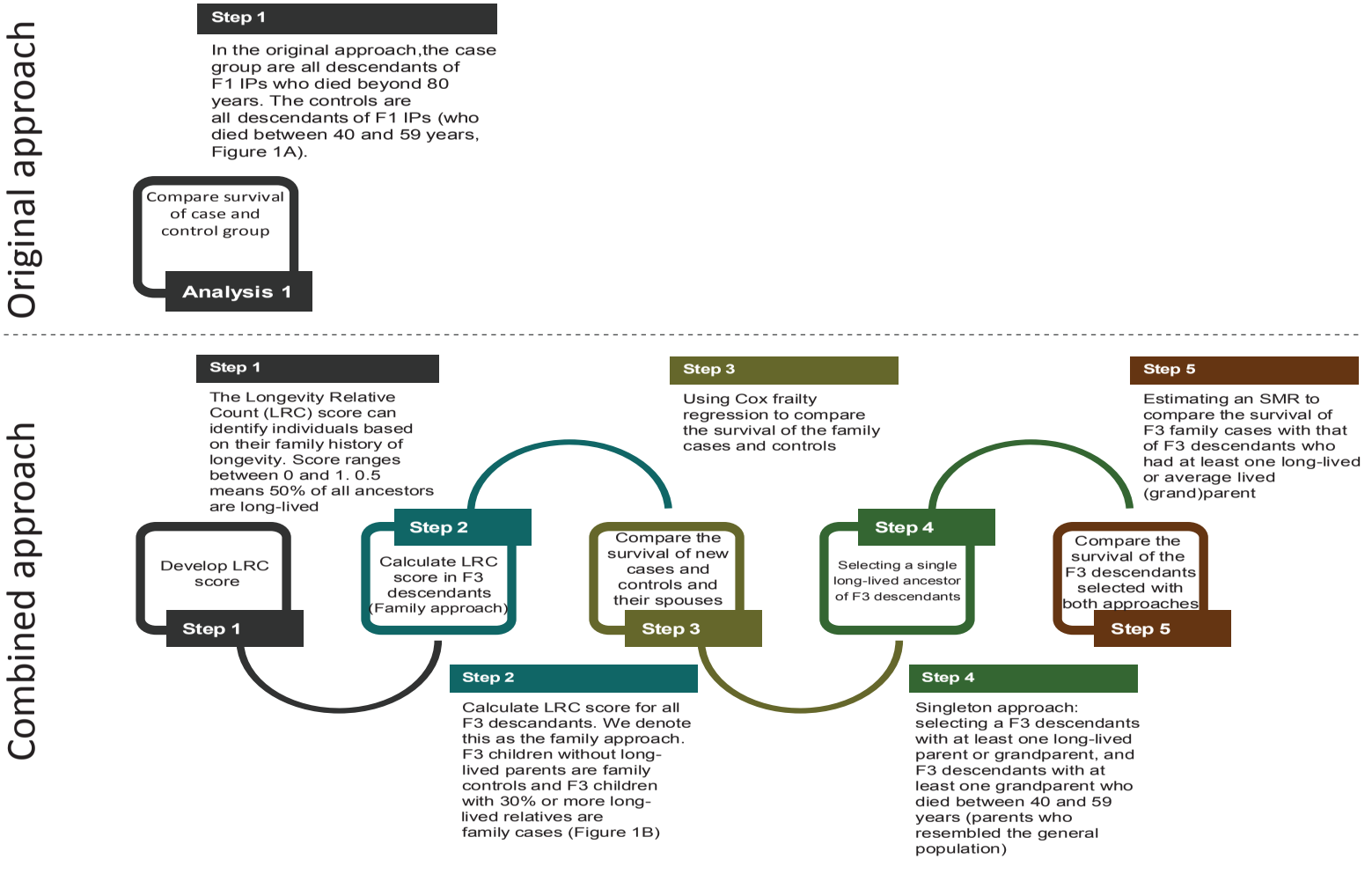
Supplementary Figure 3: Selection of F3 descendants based on the LRC and on at least one long-lived parent

The ? sign shows that the survival of that specific ancestor was unknown. The N sign shows that the ancestor was nog long-lived (top 10% survivor). The X sign shows that the ancestor was long-lived. Panel A shows the F3 descendants with at least one long-lived parent. As illustrated, at least one means that we actively selected F3 descendants with one long-lived parent. That means that the other ancestors could also be long-lived but we did not take that information into account. This resembles the selection procedure of genetic longevity studies which focus on singletons. Panel B shows the ancestors with 30% long-lived ancestors or more and the corresponding standardized mortality ratio (SMR) observed for that group of F3 descendants. Panel C shows the F3 descendants who had between 20% and 30% long-lived ancestors and the corresponding SMR observed for that group. Panel D shows the F3 descendants with more than 0 and less than 20% long-lived ancestors and the corresponding SMR to that group.



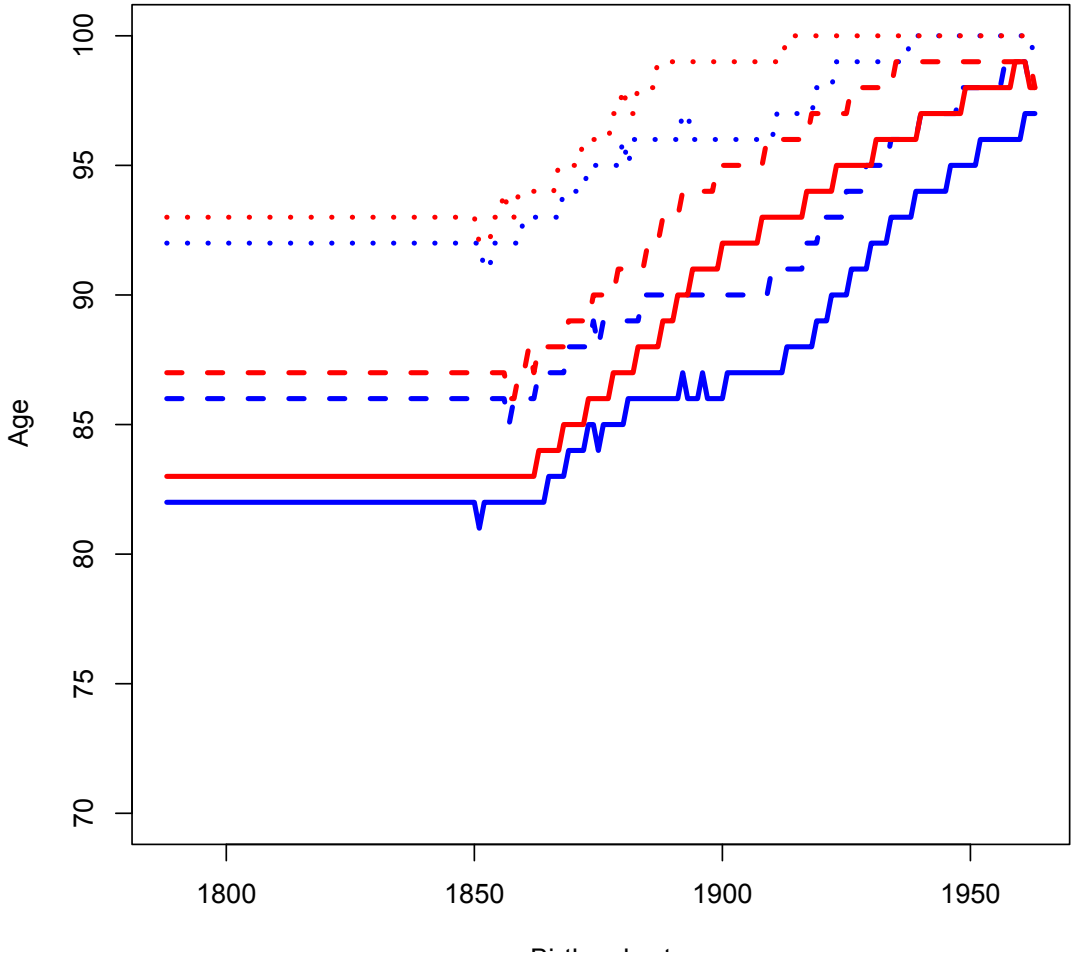
Supplementary Figure 4: Data cleaning procedure into the original and combined approach

Figure provides an overview of the specific sample selection for this study. It starts with the original Historical Sample of the Netherlands data and works towards the specific study persons for this study and finally to the two different approaches used in this study.



Supplementary Figure 5: Subsequent analyses steps in the original and combined approach

This figure shows the different steps used for the analyses in the original and the combined approach



Supplementary Figure 6: HSN birth cohorts mapping of age by top 1, 5, and 10th percentile

This figure represents the percentile-age pairings from the Dutch lifetables used to calculate survival percentiles in HSN case/control datasets. Line colors: Blue: men, Red: women. Line patterns: Dotted lines represent the top 1% survivors of the specific birth cohorts. Broken lines represent the top 5% survivors of the specific birth cohorts. Unbroken lines represent the top 10% survivors of the specific birth cohorts.

CHAPTER 9

SUMMARY

Identifying genes that code for proteins associated to longevity is an important aspect of aging research. These longevity genes likely represent key mechanisms of a life-long decreased mortality and a decreased probability of age-related disease causing persons to be free of disease until old age (compression of morbidity towards the end of life). However, the identification of longevity genes has been challenging and only a handful of genetic variants have been shown to associate with longevity across multiple independent genome-wide linkage and association studies. The most compelling evidence was obtained for variants in the APOE and FOXO3A genes as they have been consistently identified with either genome-wide association studies or candidate gene studies.

One of the main reasons for the limited success of genetic longevity studies is the uncertainty in defining the heritable longevity trait itself. The increased life expectancy of the past 200 years due to non-genetic/social factors such as improved hygiene, nutrition and medication. As a result, there are likely many phenocopies (a person who shows the characteristics of a genotype (survival into extreme ages) but where the underlying mechanism lies in non-genetic factors) among the long-lived persons selected for our genetic studies. To illustrate this, the number of centenarians (survivors to 100+ years) increased from 1 in 10,000 in 1994 to 1 in 5,000 in 2012. In this thesis I show that the solution to identifying longevity genes may lie in the familial clustering of longevity and the inclusion of persons with the heritable longevity trait (persons descending from a long-lived family) in future longevity research.

In **chapter 2** we reviewed the relevant studies investigating the familial component of longevity. We focused on heritability studies, studies investigating the transmission of lifespan and longevity as well as lifespan and longevity inheritance patterns. We further discussed important environmental/social covariates that affect individual lifespan and longevity and/or potentially affect the transmission of lifespan and longevity between parents and offspring. We emphasized the importance of distinguishing between lifespan and longevity because currently, longevity is often confused with lifespan. Lifespan generally refers to the age at death of a person whereas longevity refers to survival into extreme ages. These extremes can be defined as passing away after 80, 90, 100 years, or an extreme survival percentile such as belonging to the top 10%, 5%, of 1% birth cohort specific survivors. Estimates for lifespan vary slightly between 12% and 25% which means that this percentage of the differences in lifespan is due to genetic factors. The impact of environmental factors on the average lifespan thus likely surpasses the genetic impact (max 25% genetic and 75 non-genetic) and the interaction between both might potentially be important. In contrast to the number of studies into the heritability of lifespan, studies into the heritability of longevity are scarce and report inconsistent heritability results. Moreover, there are indications that the heritability increases with a more strict cutoff of lifespan towards more extreme ages of survival. Studies focusing on the transmission pattern of lifespan and longevity are both inconsistent but, for lifespan studies there are indications of a female transmission pattern. We conclude that environmental/social factors, such as socio-economic status, sibship size, maternal age at first and last birth and birth order should be taken into account when investigating the familial component of lifespan and longevity. We further conclude that novel research is needed to estimate the heritability of longevity and establish a longevity transmission pattern. Because there is no consensus of how longevity should be defined, we first discuss a strategy to identify a definition of longevity that best represents the heritable component of the trait. In this new strategy we emphasize the importance of big family tree databases that do not contain study related selections (such as a requirement that study participants have to be older than 90 years to participate in a study) and allow the testing of

multiple longevity definitions. We further emphasize the importance of using survival percentiles (for example, belonging to the top 10% survivors of you ow birth cohort) to correct for changes over time such as the increasing knowledge of good hygiene and better health care. Hence, big family tree data can help to develop a new definition of longevity which can be applied in aging research.

We wanted to investigate the familial component of longevity in large family tree data in the absence of study related selections such as the inclusion of alive persons who survived into old age. Hence, we constructed the LINKing System for historical family reconstruction (LINKS) data together with the International Institute of Social History (IISH) and the Radboud University (RU). The LINKS data consists of birth, marriage, and death certificates that were linked together on the basis of name combinations of the persons mentioned on the certificates. The result is a family tree database (with multiple generations of reconstructed families and persons in these families) and the life courses of the persons in the family tree. Currently, LINKS is available for the province of Zeeland and in the future this will be extended to the entire Netherlands. Life course reconstruction refers to constructing all important events that happened in the life of a person, such as a birth, marriage, moving and passing away. Family reconstruction refers to connecting kin so that family ties become visible. An example is linking children to parents and siblings. Because of the novel character of the LINKS data, we first set out to validate the life course and family reconstruction quality by comparing the LINKS data with the already existing HSN data. Thus, in **chapter 3**, we compared indicators of fertility, marriage, mortality, and measurements of occupational status of around 400 individuals who were present both in the HSN and LINKS data. We concluded that life course and family reconstructions in the HSN and LINKS reflect each other well. As we expected on the basis of differences in the data sources underlying both databases, LINKS provides more complete family information on siblings and parents, whereas the HSN provides more complete life course information, especially for individuals who migrated out of Zeeland. We also observed that the number of children was very similar between the 400 research persons. This coincides with the very complete life course information in the HSN which accurately captures the births of children. We conclude that life course and family reconstructions based on civil certificates, such as in LINKS, in linked persons constitute a reliable alternative for reconstructions based on population registers, such as in the HSN. After verifying the quality of the LINKS data we were able to investigate the definition of heritable longevity and the familial clustering of longevity using the LINKS data and the Utah Population Database (UPDB). The UPDB is a large family tree database that started with family cards supplied by the Mormon Church situated in Utah, US. These cards provided the life course and family relations of the person that was central on the card. The database exists for decades and is currently extended with all persons from Utah. The data are verified by linking them to birth, marriages, and death records as well as medical records, driver licence records and censuses, ensuring a high data quality. Besides that, the connection between data from living persons and their deceased ancestors is unique in the world. Combined, the data represent the largest family tree database with verified (mortality) information in the world. In **chapter 4** we used three-generational mortality data from the UPDB and LINKS, and studied 20,360 families who were unselected for mortality. We focused on 20,360 index persons, their parents (N=40,72), siblings (N=108,122), spouses (N=22,018), and children (N=123,599), comprising a total of 314,819 individuals. We investigated which survival percentile best isolates the heritable component of longevity and we subsequently determined the importance of long-lived family members for case selection so that those insights can be used in genetic studies to identify novel longevity genes. We further studied the non-genetic/social factors, such as socio-economic status, religious

denomination, number of children, birth order, and birth cohort, that may explain the intergenerational transmission of longevity. Moreover, we explored the survival of spouses marrying into longevity enriched families as an indicator for shared resources, lifestyles, and potentially socio-economic status during middle and late-life as explaining factors for the familial component of longevity. It is important to note that we indirectly investigated social and living environmental influences on the familial component of longevity by comparing Utah and Zeeland. Utah and Zeeland distinctly differed in their physical environment, living conditions, and subsequent mortality patterns. For example, in Zeeland there was a lack of clean drinking water whereas this was barely a problem in Utah. Our results indicated a survival advantage, mounting to 31%, for individuals with an increasing number of top 10% surviving first and second-degree relatives in both databases and across generations, even in the presence of non-long-lived parents. As such, our analyses provided strong evidence that longevity is transmitted as a quantitative genetic trait among survivors up to the top 10% of their birth cohort.

In chapter 4 we did not obtain evidence that factors such as socio-economic status, sibship size, birth order affected the association between parental or sibling longevity and the survival of the index persons. Some factors, such as socio-economic status, birth year, and religious denomination did affect the individual survival of the index person themselves, but as mentioned, independently of the parental and sibling effects. No evidence was observed that spouses marrying into a longevity enriched family also showed a survival benefit. This could, however, be expected when families live long due to socio-economic benefits or because persons find a partner in the same social environment. Interestingly, the results between the UPDB and LINKS were almost identical. These similarities provide strong evidence that the familial component of longevity is only to a small extent influenced by the living environment and for example migration patterns. A possible limitation is that not all information of socio-economic factors were measured in historical family tree databases, that some information was not that extensively measured compared to the current standard, or that the role of certain socio-economic influences changed over time. Finally, to guide future genetic studies, we suggest to select persons belonging to the top 10% survivors with first and/or second degree relatives who also belong to the top 10% survivors.

In **chapter 5**, we applied the new survival percentile threshold based longevity definition to the Leiden Longevity Study (LLS) where we studied the 944 participating long-lived brothers and sisters of 89 and 91 years and older and their relatives. We investigated 1. a potential sex-specific inheritance pattern of longevity, 2. a potential survival advantage of long-lived sibships as compared to long-lived singletons and 3. whether the parents of these siblings had a life-long sustained survival advantage. Family longevity scores were estimated to explore whether human longevity is transmitted preferentially through the maternal or paternal line. Standardized mortality ratio's (SMRs) were estimated to investigate whether long-lived siblings have a survival advantage compared to matched long-lived singletons (Dutch individuals from the same birth year and sex). In addition, we investigated if parents of long-lived siblings harbor a life-long sustained survival advantage compared to the general Dutch population by estimating lifetime SMRs (L-SMRs). We observed that sibships with long-lived mothers and non-long-lived fathers had a lower hazard of dying than sibships with long-lived fathers and non-long-lived mothers and also had a lower hazard of dying than sibships with both parents non-long-lived. Participating siblings had 18.6% less deaths compared to matched singletons and parents had a life-long sustained survival advantage. In conclusion, genetic longevity studies may incorporate the testing of a maternal transmission pattern and potential genes involved appeared to beneficially influence the entire life-course of individuals.

In chapter 4 we addressed the issue of the uncertainty in defining the longevity trait itself and observed that the survival percentile threshold that best reflects the heritable component of longevity is at the top 10% survivors (or more extreme, e.g. top 5% or top 1% survivors) of their birth cohort. Besides that, we observed that the survival advantage of family members increased with each additional long-lived family member. In **chapter 6** we followed-up on the longevity definition as established in chapter 4. We investigated if longevity is transmitted for multiple generations and whether the longevity effect diminishes over generations. To answer these questions we extended the HSN data by identifying 1326 persons, born between 1860 and 1875 in the Dutch population registers. Out of these 1,326 persons, 844 died at 80 years or older (cases) and 442 died between 40 and 59 years. We subsequently identified the children of the children until reaching the living descendants. We refer to this study as the HSN Long Lives. In the HSN Long Lives we compared long-lived cases and their descendants to population resembling controls and their descendants. We developed the Longevity Relatives Count (LRC) score to establish how many family members should be long-lived in order to avoid phenocopies among the cases in a genetic study. We subsequently investigated how often long-lived parents from a long-lived family pass on their longevity to their children compared to long-lived parents from general population families. Our analyses included 37,825 persons from 1,326 three-generational families in the HSN Long Lives study. The analyses in the HSN Long Lives dataset provide strong evidence that longevity is transmitted for at least 2 subsequent generations if at least 20% of all relatives are long-lived, but preferably 30%. Moreover, the cases with 30% long-lived family members seem to be at least partially genetically enriched for longevity, as factors such as birth year, sibship size, and sex did not affect the transmission of longevity. The evidence is strengthened by the fact that their spouses resembled the controls (third generation descendants without any long-lived family members) as well as the general population in their average survival. Moreover, it is known from other historical demographic research that a large variation of factors, such as religion, and socio-economic status do not influence the association between parental and offspring longevity. Finally, 27% of the third generation descendants showed a survival pattern similar to the general population even though they had at least one long-lived parent. Hence, it appears that they did not have the heritable longevity trait and thus did not transmit their longevity. In summary we conclude that to select individuals who are enriched for the heritable longevity trait, cases should be selected on the basis of being long-lived themselves and having at least 30% long-lived ancestors.

The insights obtained in this dissertation can be used to aid current genetic longevity studies. For example, the case and control definition in genetic association studies can be sharpened by including mortality information from family members such as parents which is likely to increase the study's power. The insights can also be used to design novel genetic longevity studies, for example by using the LRC score to select the families most enriched for longevity in large family tree databases. Combined with whole genome sequencing (mapping the entire gnome) data and research techniques making optimal use of the available family information, not only common but also rare genetic variants may be obtained. With the current insights new efforts can be made to estimate the heritability of longevity, to separate genetic and non-genetic contributions to the familial component of longevity, and to identify a longevity inheritance pattern. In addition, further investigation of the social mechanisms that underlie the familial clustering of longevity might be studied. This can partially be done in current family tree data but also requires novel data in which living members who descend from long-lived families are studied and followed-up, for example to study their social network, lifestyle and genetic profile. Finally, the large historical family tree data can be used to enrich contemporary databases, such as the LLS. These large data can also be used for more general

aging research questions such as the exitance of a mortality plateau in humans (the decline or stagnation of the increase in the hazard of dying).

CHAPTER 10

NEDERLANDSE SAMENVATTING

Een belangrijk aspect van verouderingsonderzoek is het identificeren van genen die coderen voor eiwitten die aan langlevendheid bijdragen. Deze langlevendheid genen kunnen indicatief zijn voor de mechanismen die ten grondslag liggen aan verlaagde sterftekansen gedurende het hele leven en een verlaagde kans op onderdomsziektes waardoor mensen tot op hoge leeftijd vrij van ziekte blijven (compressie van verouderingsziekten tot aan het einde van het leven). De identificatie van langlevendheid genen is echter uitdagend gebleken en slechts een handvol genetische varianten is gevonden in meerdere onafhankelijke genoom wijde linkage en associatie studies. Het meest robuuste bewijs is aanwezig voor varianten van de APOE en FOXO3A genen omdat deze consistent geïdentificeerd zijn in kandidaat gen of genoom wijde associatie studies.

Een van de belangrijkste redenen voor het beperkte succes van genetisch langlevendheid onderzoek is de onzekerheid rond het definiëren van de erfelijk bepaalde langlevendheid. De toename in levensverwachting in de afgelopen 200 jaar is vooral veroorzaakt door niet genetische/sociale factoren zoals betere hygiëne, voeding en medicatie. Daardoor bevinden zich onder de langlevende personen die geïncludeerd worden in genetisch onderzoek waarschijnlijk veel fenokopieën (een persoon die de karakteristieken van een bepaald genotype vertoont (extreem hoge leeftijd) maar waarbij de oorzaak ligt in niet genetische factoren. Ter illustratie, het aantal eeuwelingen (100+ers) is gestegen van 1 op de 10.000 in 1994 naar 1 op de 5.000 in 2012. In deze dissertatie laat ik zien dat de oplossing voor het identificeren van langlevendheid genen gezocht kan worden in de familiaire clustering van langlevendheid en de inclusie van personen met een erfelijke vorm van langlevendheid (personen die uit een lang levende familie komen) in toekomstig onderzoek.

In **hoofdstuk 2** worden studies samengevat waarin onderzoek naar de familiaire component van langlevendheid beschreven werd. Hierbij richtten we ons op erfelijkheidsstudies, studies naar de overdracht van levensduur en langlevendheid (transmissie) van ouders naar kinderen en de overervingspatronen waarmee dat gebeurt (van moeders op dochters of kleindochters bijvoorbeeld). Daarnaast bediscussiëren we belangrijke omgevings/sociale factoren die de levensduur en langlevendheid van individuen en/of de transmissie van levensduur en langlevendheid beïnvloeden. We benadrukken het belang van het maken van onderscheid tussen levensduur en langlevendheid omdat deze vaak door elkaar gehaald worden in onderzoek. Levensduur refereert naar iemands leeftijd van overlijden terwijl langlevendheid refereert naar personen die extreme leeftijden bereiken. Die extremen kun je definiëren als overlijden boven de 80, 90, of 100 jaar, of door een extreem overlevingspercentiel zoals alle personen behorende tot de top 10%, 5% of 1% langstlevende personen van hun geboortecohort. Schattingen van erfelijkheid van levensduur variëren tussen de 12% en 25%, wat betekent dat zoveel procent van verschillen in levensduur door genetische factoren wordt verklaard. De impact van omgevingsfactoren op de gemiddelde levensduur is dus waarschijnlijk veel groter dan de genetische impact (maximaal 25% genetisch en 75% niet genetisch) en wellicht speelt de interactie tussen de twee ook een belangrijke rol. In tegenstelling tot het aantal studies naar de erfelijkheid van levensduur zijn er bijna geen studies gedaan naar de erfelijkheid van langlevendheid (dus bij de extremen) en de studies die gedaan zijn laten inconsistente resultaten zien. Daarnaast zijn er indicaties dat de erfelijkheid toeneemt met een striktere definitie van levensduur naar langlevendheid. Studies die zich richten op de transmissie patronen van levensduur en langlevendheid leveren beide inconsistente resultaten op maar levensduur studies lijken een indicatie voor een maternaal overervingspatroon te geven. Wij concluderen dat omgevings/sociale factoren zoals sociaal economische status, aantal broers en zussen, de leeftijd waarop een vrouw haar eerste en laatste kind kreeg en geboortevolgorde in ogenschouw genomen moeten worden in onderzoek naar de familiaire component van zowel

levensduur als langlevendheid. We concluderen verder dat nieuw onderzoek nodig is om de erfelijkheid van langlevendheid te schatten, net als een overervingspatroon voor langlevendheid. Omdat er nog geen consensus is over hoe langlevendheid gedefinieerd moet worden stellen we een nieuwe strategie voor om tot een onderbouwde definitie te komen die de erfelijke component weerspiegelt. In deze nieuwe strategie benadrukken wij het belang van grote stamboom databases waarin geen studie gerelateerde selecties (zoals dat personen bijvoorbeeld 90 jaar of ouder moesten zijn om mee te kunnen doen aan een studie) zijn gemaakt en waarin dus veel definities van langleveneheid onderzocht kunnen worden. Ook benadrukken wij het belang van het gebruik van overlevingspercentielen (bijvoorbeeld behoren tot de top 10% overlevers van je eigen geboortecohort) om te corrigeren voor veranderingen over tijd, zoals de toenemende kennis over goede hygiëne en de steeds beter wordende gezondheidszorg. Grote stamboom databases kunnen dus helpen om tot een nieuwe definitie van langlevendheid te komen die goed onderbouwd is en vervolgens toegepast kan worden in verouderingsonderzoek.

We wilden graag de familiaire component van langlevendheid onderzoeken in grote stamboom data zonder studie gerelateerde selecties, zoals de inclusie alleen van levende personen die een hoge leeftijd hebben bereikt. We hebben daarom de "LINKing System for historical family reconstruction" (LINKS) data geconstrueerd samen met het Internationaal Instituut voor Sociale Geschiedenis (IISG) en de Radboud Universiteit (RU). De LINKS data bestaat uit geboorte, huwelijks en overlijdensaktes die aan elkaar gelinkt zijn op basis van naam combinaties van de personen die op de aktes vermeld staan. Het resultaat is een stamboom database (met meerdere generaties aan gereconstrueerde families en personen in die families) en de levenslopen van de personen in de stamboom. Momenteel is LINKS beschikbaar voor de provincie Zeeland en in de toekomst wordt dit uitgebreid tot heel Nederland. Levensloop reconstructie verwijst naar het reconstrueren van alle belangrijke gebeurtenissen die gebeurd zijn in het leven van een persoon, zoals een geboorte, huwelijk, verhuizing en overlijden. Familie reconstructie verwijst naar het aan elkaar koppelen van verwantschappen zodat familieverbanden zichtbaar worden, zoals het koppelen van kinderen aan ouders en broers en zussen. Omdat de LINKS data nieuw was hebben we de levensloop en familie reconstructies eerst gevalideerd met de al bestaande Historische Steekproef van Nederland (HSN). Wij hebben daarom in **hoofdstuk 3** indicatoren van fertiliteit, huwelijk, mortaliteit en beroepsstatus van ongeveer 400 personen die zowel in de HSN als de LINKS data aanwezig waren met elkaar vergeleken. We konden concluderen dat levensloop en familie reconstructies in beide databases elkaar goed weerspiegelden. Zoals verwacht, op basis van verschillen in de bron en opbouw in de database, was er meer complete familie informatie met betrekking tot broers, zussen en ouders in de LINKS data aanwezig, terwijl er in de HSN data meer complete levensloop informatie aanwezig was. Dit was in het bijzonder zichtbaar voor personen die emigreerden uit de provincie Zeeland. Daarnaast observeerden we dat het aantal kinderen geregistreerd in LINKS en HSN vrijwel identiek was tussen de 400 onderzoekspersonen. Dit hangt samen met de zeer complete levensloop informatie in de HSN waardoor geboortes van kinderen accuraat zijn vastgelegd. We concluderen dat levensloop en familie reconstructies gebaseerd op de burgerlijke stand, zoals in de LINKS, in gelinkte personen een betrouwbaar alternatief kunnen bieden voor reconstructies gebaseerd op de bevolkingsregisters, zoals in de HSN.

Na het verifiëren van de kwaliteit van de LINKS data konden we de definitie van erfelijke langlevendheid en de familiare clustering van langlevendheid in de LINKS data en Utah Populatie Database (UPDB) onderzoeken. De UPDB is een grote stamboom database die begonnen is met de familie kaarten die werden aangeleverd door de Mormoonse Kerk gevestigd in Utah, VS. Deze kaarten geven voor de persoon die centraal staat op de kaart de levensloop en familie relaties weer. De

database bestaat al tientallen jaren en is momenteel uitgebreid met mensen uit heel Utah. Alle gegevens zijn geverifieerd door koppeling aan geboorte, huwelijks en overlijdens aktes. Ook is de data gekoppeld aan medische dossiers rijbewijs databases en volkstellingen waardoor de data van extreem hoge kwaliteit is. Daarnaast is de koppeling tussen gegevens van levende mensen en hun overleden voorouders uniek in de wereld. Samen vormen deze databases de grootste genealogische data verzameling met geverifieerde (mortaliteits) gegevens. In **hoofdstuk 4** hebben wij drie generationele mortaliteits data uit de UPDB en LINKS gebruik om 20.360 families te bestuderen die ongeselecteerd waren voor mortaliteit. Meer specifiek hebben we 20.360 index personen, hun ouders (N=40,72), broers en zussen (N=108,122), partners (N=22,018) en kinderen (N=123,599) onderzocht met een totaal van 314,819 personen. We hebben onderzocht welk overlevingspercentiel het beste de erfelijke component van langlevendheid weerspiegelt en daaropvolgend het belang van lang levende familieleden onderzocht voor het selecteren van personen in genetisch onderzoek. De verworven inzichten kunnen gebruikt worden om nieuwe langlevendheid genen te identificeren in een studie gebaseerd op goed gefundeerde definities van langlevende 'cases' en controle personen. Daarnaast hebben we niet genetische/sociale factoren zoals sociaal economische status, religieuze denominatie, aantal kinderen, geboortevolgorde en geboortecohort onderzocht, omdat deze mogelijk de intergenerationele transmissie van langlevendheid verklaren. Ook hebben wij de overleving van partners die huwen in een lang levende familie onderzocht. Deze partners kunnen indicatief zijn voor gedeelde middelen, leefstijl en sociaal economische status gedurende middelbare en latere leeftijden en zo de familiaire component van langlevendheid verklaren. Het is belangrijk op te merken dat we sociale en leefomgevingsinvloeden ook indirect hebben kunnen onderzoeken door Utah en Zeeland met elkaar te vergelijken. Utah en Zeeland verschilden onderling enorm in leefomgeving, levensstandaard en sterftepatronen die daaruit voortvloeiden. Zo was er bijvoorbeeld in Zeeland een gebrek aan schoon drinkwater terwijl dit in Utah amper een probleem was. We observeerden een overlevingsvoordeel, dat opliep tot 31%, bij personen met een toenemend aantal top 10% eerste en tweedegraads verwanten in beide databases en generaties, zelfs in de afwezigheid van lang levende ouders. Onze analyses hebben sterk bewijs geleverd dat langlevendheid wordt overgedragen als een kwantitatief genetische eigenschap bij personen die behoren tot minimaal de top 10% overlevers van hun eigen geboortecohort.

Wij konden in de analyses van hoofdstuk 4 geen bewijs vinden dat factoren zoals sociaal economische status, aantal broers en zussen en geboortevolgorde de associatie tussen lang levende ouders of broers en zussen en de overleving van index personen beïnvloeden. Een aantal factoren zoals sociaal economische status, geboortejaar en religieuze denominatie beïnvloeden echter wel de overleving van de index personen zelf. Er is geen bewijs gevonden dat partners die huwen in een lang levende familie zelf ook een overlevingsvoordeel hebben, dat zou je wel kunnen verwachten als die families lang leven door sociaal economische voordelen of doordat mensen hun partner zoeken in dezelfde sociale omgeving. Het was interessant om te zien dat de resultaten vrijwel identiek waren tussen de UPDB en LINKS. Deze gelijkenis geeft sterk bewijs dat de familiaire component van langlevendheid maar beperkt beïnvloed wordt door de leefomgeving en bijvoorbeeld migratie patronen. Een mogelijke kanttekening is echter dat niet alle gegevens over sociale en economische factoren gemeten zijn in historische databases, dat sommige gegevens minder volledig gemeten zijn dan tegenwoordig of dat de rol van bepaalde sociaal of economische invloeden veranderd over tijd. Om toekomstig onderzoek richting te geven adviseren wij om personen te selecteren die behoren tot de top 10% overlevers met eerste en/of tweedegraads verwanten die ook tot de top 10% overlevers behoren.

In **hoofdstuk 5** hebben wij de nieuwe overlevingspercentiel drempel toegepast als definitie voor langlevendheid in de Leiden Langleven Studie (LLS) waar we 944 deelnemende broers en zussen van 89 en 91 jaar of ouder en hun verwanten bestudeerden. We onderzochten 1. Een potentieel sekse specifiek overervingspatroon voor langlevendheid, 2. Een potentieel overlevingsvoordeel van lang levende broers en zussen in vergelijking tot lang levende individuen en 3. Of de ouders van de deelnemende broers en zussen een levenslang verkleinde sterftekans hadden. We hebben familie scores berekend om te onderzoeken of langlevendheid preferentieel via de mannelijke of vrouwelijke lijn wordt doorgegeven. "Standardized mortality ratio's" (SMRs) zijn geschat om te onderzoeken of lang levende broers en zussen een overlevingsvoordeel hebben ten opzichte van gematchte controle individuen (Nederlandse individuen uit een zelfde geboortejaar en met hetzelfde geslacht). Daarnaast hebben we "lifetime SMRs" (L-SMRs) gebruikt om te onderzoeken of de ouders van deze lang levende broers en zussen een levenslang overlevingsvoordeel hebben ten opzichte van de algemene Nederlandse bevolking. We observeerden een kleinere sterftekans op ieder moment in het leven bij broers en zussen met lang levende moeders en niet lang levende vaders dan bij broers en zussen met lang levende vaders en niet lang levende moeders, evenals voor broers en zussen waarbij beide ouders niet langlevend waren. Deelnemende broers en zussen hadden 18,6% minder geobserveerde overlijdens dan gematchte controle individuen en ouders hadden een levenslang overlevingsvoordeel. Onze eindconclusie is dat genetisch onderzoek naar langlevendheid het testen van maternale transmissie patronen zouden kunnen incorporeren. Daarnaast lijkt het er op dat potentiele langlevendheid genen een gunstige invloed hebben gedurende het hele leven.

In hoofdstuk 4 hebben we onderzoek gedaan naar de onzekerheid in de definitie van langlevendheid en observeerden wij dat de overlevingspercentiel drempel, die de erfelijke component van langlevendheid het beste reflecteert, ligt bij de top 10% overlevers (of extremer) van hun eigen geboortecohort. We observeerden daarnaast dat het overlevingsvoordeel van familieleden toenam met ieder extra lang langlevend familielid. In **hoofdstuk 6** bouwen we voort op de definitie van langlevendheid zoals vastgesteld in hoofdstuk 4. We hebben onderzocht of langlevendheid overgedragen wordt op meerdere generaties en of het effect uitdooft over generaties. Om deze vragen te kunnen beantwoorden hebben wij de HSN data uitgebreid door van 1326 personen, geboren tussen 1860 en 1875, in de HSN alle kinderen op te zoeken in het Nederlandse bevolkingsregister. Van de 1326 personen waren er 884 overleden op 80 jarige leeftijd of ouder ("cases") en 442 overleden tussen hun 40e en 59e levensjaar (controles). Vervolgens hebben wij de kinderen van de kinderen opgezocht totdat we bij de levende nabestaanden kwamen. Deze studie heet de "HSN Long Lives". Wij hebben langlevende "cases" en hun nazaten vergeleken met controles, die een sterftepatroon hadden gelijk aan de algemene populatie, en hun nazaten. We hebben de "Longevity Relatives Count" (LRC) score ontwikkeld om vast te stellen hoeveel familieleden langlevend moeten zijn om fenokopieën te vermijden onder de "cases" in een genetische studie. Daaropvolgend hebben we onderzocht hoe vaak langlevende ouders uit langlevende families hun langlevendheid doorgeven aan hun kinderen in vergelijking tot langlevende ouders uit controle families die het sterfte patroon uit de algemene populatie volgden. Onze analyses zijn gebaseerd op 37.825 personen uit 1.326 drie generatie families in de HSN Long Lives studie. De analyses in de HSN Long Lives dataset leveren sterk bewijs dat langlevendheid in ieder geval 2 generaties wordt overgedragen als in ieder geval 20% van alle verwanten langlevend zijn, maar idealiter 30%. Bovendien lijken de "cases" met 30% langlevende familieleden in ieder geval deels genetisch verreikt voor langlevendheid aangezien factoren als geboortejaar, geslacht en aantal broers en zussen de transmissie van langlevendheid niet beïnvloedde. Dit bewijs wordt versterkt door het feit dat partners een sterftepatroon hadden dat gelijk was aan de controles (derde generatie nakomelingen zonder langlevende familieleden) en

de algemene Nederlandse bevolking. Het is bovendien bekend uit ander historisch demografisch onderzoek dat een grote variatie aan factoren, zoals religie en sociaal economische status het effect tussen langlevendheid van ouders en kinderen niet beïnvloed. Ten slotte concluderen we dat 27% van de derde generatie nakomelingen een sterftepatroon laat zien dat gelijk is aan de algemene Nederlandse bevolking, ondanks dat ze in ieder geval één langlevende ouder hadden. Het lijkt er dus op dat de ouders van deze nakomelingen niet de erfelijke vorm van langlevendheid hadden en daarom hun langlevendheid niet overgedragen hebben. Samenvattend concluderen we dat wanneer "cases" met erfelijke langlevendheid geselecteerd moeten worden, het noodzakelijk is om lang levende personen met in ieder geval 30% langlevende voorouders te selecteren.

De inzichten die verkregen zijn in deze dissertatie kunnen gebruikt worden om genetisch onderzoek naar langlevendheid een boost te geven. Zo kan bijvoorbeeld de "case" en control definitie in genetische associatie studies aangescherpt worden door sterfte informatie van familieleden zoals ouders mee te nemen om zodoende de analysekracht van deze onderzoeken te vergroten. Ook kunnen de inzichten gebruikt worden om nieuwe genetische studies naar langlevendheid op te zetten, door bijvoorbeeld de LRC score te gebruiken om de families die het meest verrijkt zijn voor langlevendheid te identificeren in grote stamboom databases. Gecombineerd met "whole genome sequencing" (het in kaart brengen van het hele genoom) data en onderzoekstechnieken die optimaal gebruik maken van beschikbare familie informatie, kunnen wellicht niet alleen veelvoorkomende maar ook zeldzame genetische varianten geïdentificeerd worden. Met de huidige inzichten kunnen nieuwe pogingen ondernomen worden om de erfelijkheid van langlevendheid te schatten, om genetische en niet genetische factoren die bijdragen aan familiaire langlevendheid te onderscheiden en om een overervingspatroon te identificeren. Daarnaast is het mogelijk om de sociale mechanismes die onderliggend zijn aan de familiaire clustering van langlevendheid verder te onderzoeken. Dit kan gedeeltelijk worden gedaan in grote stamboom datasets, zoals gebruikt in deze dissertatie, maar vereist ook nieuwe data waarin levende mensen die afkomstig zijn uit een lang levende familie bestudeerd en gevolgd kunnen worden, bijvoorbeeld op het gebied van hun sociale netwerk, leefstijl en genetisch profiel. Tot slot is het mogelijk om de grote historische stamboom databases te gebruiken om bestaande hedendaagse databases, zoals de LLS te verrijken, evenals het verrichten van meer algemene verouderingsanalyses, bijvoorbeeld naar het bestaan van een mortaliteitsplateau (het afnemen of volledig stagneren van de toename in sterftekansen).

CHAPTER 11

APPENDIX

PUBLICATIONS - CV - DANKWOORD

Publications

Scientific publications

van den Berg, N., Rodríguez-Girondo, M., de Craen, A. J. M., Houwing-Duistermaat, J. J., Beekman, M., & Slagboom, P. E. (2018). Longevity Around the Turn of the 20th Century: Life-Long Sustained Survival Advantage for Parents of Today's Nonagenarians. The Journals of Gerontology: Series A, 73(10), 1295–1302.
https://doi.org/10.1093/gerona/gly049

Eline Slagboom, P., **van den Berg, N.**, & Deelen, J. (2018). Phenome and genome based studies into human ageing and longevity: An overview. Biochimica et Biophysica Acta (BBA) - Molecular Basis of Disease, 1864(9), 2742–2751.
https://doi.org/10.1016/j.bbadis.2017.09.017

Mourits, R. J., **van den Berg, N.**, Rodriguez-Girondo, M., Mandemakers, K., Slagboom, P. E., Beekman, M., & Janssens, A. A. P. O. (2019). Intergenerational transmission of longevity is not affected by other familial factors: Evidence from 16,905 Dutch families from Zeeland, 1812-1962. BioRxiv, 1–40.
https://doi.org/10.1101/781500

van den Berg, N., Beekman, M., Smith, K. R., Janssens, A., & Slagboom, P. E. (2017). Historical demography and longevity genetics: Back to the future. Ageing Research Reviews, 38, 28–39.
https://doi.org/10.1016/j.arr.2017.06.005

van den Berg, N., Rodriguez-Girondo, M., Mandemakers, K., Janssens, A. A. P. O., Beekman, M., & Slagboom, P. E. (2019). Longevity Relatives Count score defines heritable longevity carriers and suggest case improvement in genetic studies. BioRxiv.
https://doi.org/10.1101/609891

van den Berg, N., Rodríguez-Girondo, M., van Dijk, I. K., Mourits, R. J., Mandemakers, K., Janssens, A. A. P. O., … Slagboom, P. E. (2019). Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. Nature Communications, 10(1), 35.
https://doi.org/10.1038/s41467-018-07925-0

van den Berg, N., van Dijk, I. K., Mourits, R. J., Janssens, A., Slagboom, P. E., & Mandemakers, K. (2018). Families in Comparison: An individual-level comparison of life course and family reconstructions between population and vital event registers. Population Studies

Press publications (selection)

New Scientist: Bij de meeste mensen is levensverwachting helemaal niet zo erfelijk (28-01-2019),
https://newscientist.nl/nieuws/bij-de-meeste-mensen-is-levensverwachting-helemaal-niet-zo-erfelijk/

NOS: Lang leven vooral door lang levende moeder (27-03-2018),
https://nos.nl/artikel/2224657-lang-leven-vooral-door-lang-levende-moeder.html

RTL nieuws: Extreem oude moeder dan heb je zelf ook een grotere kans om heel oud te Worden (28-03-2018),
https://www.rtlnieuws.nl/lifestyle/gezondheid/artikel/4129131/extreem-oude-moeder-dan-heb-je-zelf-ook-een-grotere-kans-om

Health 24: Look to your aunts, uncles and parents for clues to your longevity (27-01-2019),
https://www.health24.com/Lifestyle/Ageing-well/News/look-to-your-aunts-uncles-and-parents-for-clues-to-your-longevity-20190113

Business standard: Your uncle's genes may decide your longevity (14-01-2019),
https://www.businessstandard.com/article/news-ians/your-uncle-s-genes-may-decide-your-longevity-study-119011401117_1.html

De Standaard: Lang leven is erfelijk, maar genen zijn nog spoorloos (08-01-2019),
https://www.standaard.be/cnt/dmf20190107_04085922

Omroep Zeeland: Over het geheim van oud worden (14-01-2019),
https://soundcloud.com/omroepzeeland/zeeuwse-kamer-over-het-geheim-van-oud-worden-1412019?in=omroepzeeland/sets/zeeuwse-kamer-2019

NRC: Meer dan 100? Dat is te danken aan de moeder (02-04-2018),
https://www.nrc.nl/nieuws/2018/04/02/levensduur-meer-dan-100-dat-is-te-danken-aan-de-moeder-a1597882

AD: Stokoud worden heb je vooral te danken aan je moeder (28-03-2018),
https://www.ad.nl/gezond/stokoud-worden-heb-je-vooral-te-danken-aan-je-moeder~a4647fff/

Volkskrant: Heeft de leeftijd waarop opa overlijdt invloed op de levensverwachting van de kleinkinderen? (08-02-2019),
https://www.volkskrant.nl/wetenschap/heeft-de-leeftijd-waarop-opa-overlijdt-invloed-op-de-levensverwachting-van-de-kleinkinderen~b3732ceb/

AD: Zeeland levert het bewijs: lang leven is erfelijk (08-01-2019),
https://www.ad.nl/zeeland/zeeland-levert-het-bewijs-lang-leven-is-erfelijk~a6f8b097/

Curriculum Vitae

Niels van den Berg was born on the 10th of May 1987 in Haps, the Netherlands. After graduating from "het Merletcollege" in Cuijk, he studied Information Technology, with a specialization towards system administration and databases. After that, Niels switched fields to study social sciences, and obtained a bachelor's degree in teaching and science at the Fontys University of applied Sciences and the University of Tilburg in 2012. Next, he combined the knowledge of both fields in a master on Quantitative Sociology at the University of Tilburg. He focused on statistics, methodology, and family sociology, and graduated in 2014 with distinction. During his masters Niels developed a passion for studying families from an interdisciplinary perspective. In 2014 he followed this passion into his PhD, which was at the intersection of the life sciences and social sciences. He was embedded at the Radboud University in prof. Angelique Janssens' group (Social and Demographic history) where he developed his expertise on the historical and social aspects of longevity and at the Leiden University Medical Center in prof. Eline Slagboom's group (Molecular Epidemiology) where he learned about the genetic and molecular mechanisms of aging and longevity, and data analytics in this field. Moreover, as Slagboom's group is embedded in the department of Biomedical Data Sciences, Niels could expand his interest in statistics. During his PhD he studied the familial clustering and transmission of human longevity, analyzing multiple big historical genealogical databases, such as the Utah Population Database (UPDB), LINKS, Historical Sample of the Netherlands (HSN), and Leiden Longevity Study (LLS) which cover millions of individuals. Currently, Niels is working as a researcher in prof. Slagboom's group to implement the results obtained in his PhD research.

Dankwoord

Nu ik aan het dankwoord begin realiseer ik mij dat de afgelopen jaren enorm snel zijn gegaan en dat ik er enorm van genoten heb.

Allereest wil ik graag Angelique en Eline bedanken dat ik mijn promotieonderzoek heb mogen doen op jullie afdelingen en zo veel heb mogen leren van alles wat daar gebeurde. Eline, jou bedank ik graag in het bijzonder voor de goede wetenschappelijke discussies die wij hebben gehad maar ook voor de gezelligheid, bijvoorbeeld tijdens borrels. Jij leek altijd mijn enthousiasme te delen, zelfs wanneer ik daardoor soms iets te hard begon te praten, en stimuleerde mij om het hoogst haalbare te bereiken. Marian en Mar, eigenlijk kan ik over jullie hetzelfde zeggen. Marian, ik voeg daar graag aan toe dat jij mij als dagelijks begeleider structuur hebt bijgebracht, onder andere in het schrijven van papers, het maken van presentaties en het nadenken over wat precies mijn boodschap was in stukken tekst. Mar, ik heb de solide basis van mijn statistische analyses aan jou te danken. Jullie stonden altijd klaar als ik weer eens wilde brainstormen over een (nieuw) idee.

Rick en Ingrid, mijn projectgenoten. Ik wil jullie graag bedanken voor de mooie tijd tijdens maar ook buiten het werk. We zijn tot mooie gezamenlijke publicaties gekomen en hebben een goede tijd gehad in Utah en op onze trip naar Yellowstone.

ESDG, Molepi en Medische Statistiek collega's, ik wil jullie graag bedanken voor de fijne samenwerking en een mooie tijd in Nijmegen en Leiden. Jeanine, Kees en Jan, bedankt voor alle overleggen over het project en de data constructie. Kees, jou bedank ik graag in het bijzonder omdat zonder jou er geen data was geweest. Jij bleef rustig wanneer ik het heet onder de voeten kreeg. Daarnaast bedank ik graag iedereen met wie ik in de afgelopen jaren kamer S5-08 heb mogen delen. Het was altijd een erg gezellige kamer waar goed over wetenshap gediscussieerd kon worden. I would also like to thank Ken for hosting my stay in Utah. I felt very welcome, thoroughly enjoyed my stay, and was happy to see our work turn into great science.

Een promotie vindt niet alleen op het werk plaats, daarom wil ik graag al mijn "benny lava" vrienden bedanken. Daarnaast wil ik graag mijn beachvolleybal vrienden, studievrienden en oud huisgenoten bedanken.

Mijn dank gaat ook uit naar mijn ouders, die mij alle kansen hebben gegeven en mij altijd gesteund hebben in de keuzes die ik maakte. Natuurlijk kan ik mijn twee zusjes Manon en Muriel en broertje Ward niet vergeten. Ward, ik heb tijdens mijn master 2 jaar bij jou gelogeerd in Tilburg. Ik heb daar enorm van genoten en ben blij dat jij mijn paranymf bent. Chiel, ook jij bent paranymf. We kennen elkaar al vanaf mijn geboorte en hebben vele mooie en interessante momenten gedeeld.

Als laatste gaat mijn dank naar Manon, mijn steun, toeverlaat en op zijn tijd ook vraagbaak. Tijdens mijn promotie hebben wij samen een huis gebouwd, zijn wij getrouwd en dragen nu de zorg voor onze dochter Féline. Op naar een mooie toekomst samen.