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

Berg, N.M.A. van den

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).