Title:

Longevity Relatives Count score defines heritable longevity carriers and suggest case improvement in genetic studies.

Authors:

Niels van den Berg[1,2,*] , n.m.a.van_den_berg@lumc.nl

Mar Rodríguez-Girondo[3]

Kees Mandemakers[4]

Angelique A.P.O. Janssens[2]

Marian Beekman[1]

P. Eline Slagboom[1,5]

*: corresponding author

Affiliations:

[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

[5]. Max Planck Institute for Biology of Ageing, Joseph-Stelzmann-Str. 9b, D-50931 Cologne, Germany

***Corresponding author:**

25 Name: Niels van den Berg

26 Email: n.m.a.van_den_berg@lumc.nl

27 Address: Albinusdreef 2, 2333 ZA Leiden the Netherlands

28 Phone: +31 (0) 71 5269738

29

Abstract

Longevity loci represent key mechanisms of a life-long decreased mortality, decreased morbidity, and compression of morbidity towards the end of life. However, identifying such loci has shown to be challenging. One of the most plausible reasons is the uncertainty in defining long-lived cases with a heritable longevity trait amongst long living phenocopies. We observed that longevity is only heritable for persons belonging to the top 10% survivors of their birth cohort, or more extreme, with equally long-lived family members. Here we refine that definition to identify individuals with the heritable longevity trait by using a unique dataset connecting living study participants to their deceased ancestors covering 57,337 persons from 1,326 five-generational families, living between 1788 and 2019. In the first study phase, transmission of longevity to descendants in case and control families was measured by Standard Mortality Ratios. In the second phase we combined all families and compared longevity transmission in groups based on a novel score, summarizing the familial history of longevity (Longevity Relatives Count score, LRC). Using this score, we observed that longevity is transmitted for at least 2 subsequent generations only when at least 20%, and optimally 30%, of all relatives are long-lived. A stronger survival advantage was observed for F3 descendants with ≥30% long-lived ancestors than those with at least one long-lived parent. For future studies, we suggest to include cases with ≥ 30% relatives who belong at least to the top 10% survivors of their birth cohort and are themselves among the 10% longest lived.

Introduction

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). However, genome wide linkage and association studies identified only a few robust loci promoting longevity (22–30). 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 (22–27, 31).

One of the main reasons for the limited success of longevity genetic studies (24–26, 31–33) 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 (34, 35). We explored the definition of heritable longevity in a previous study of Dutch (Province of Zeeland) and Utah datasets, 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 or more extreme. Moreover, the survival advantage of family members increased with each additional long-lived family member, according to their genetic distance to the study participant. This indicates that longevity is transmitted as a quantitative genetic trait (13). However, the majority of genealogical (5–11, 36) and genetic studies (24–26, 31–33) focus only on single, and thus likely sporadic, long-lived individuals

(singletons). Shifting focus to single individuals belonging to the top 10% survivors or more extreme may be an improvement, but has not yet resulted in the identification of robust novel loci (28). While some studies for example investigate the genetic architecture of longevity by using parental age instead of the study participants' age (28, 29) or by focusing on multiple siblings surviving into old age (15, 25), it remains unclear to what extent longevity needs to cluster in families in order to include individuals with the heritable longevity trait, which will increase the power of genetic studies (37).

Here, we aim to establish 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 a 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 (38–40). We primarily identify cases who died beyond 80 years (N=884, on average top 10% (min=top20% and max=top1%) 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. The full data comprises 57,337 persons from 1,326 five-generational families. We subsequently exclude groups with high rates of missing mortality information and where the majority was still alive (Supplementary Figure 4). 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

96 generations of spouses (F2-F3) (Table 1). The dataset is unique in that it covers multiple

97 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.

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) & $SMR_{control\_F3spouses}$=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.

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.

167

168 **Constructing the Longevity Relatives Count score**

169 We now look at the HSN data from a different perspective, the combined approach (Figure
170 1B). In the combined approach we consider the F3 generation as the focal point of the
171 pedigree, instead of the F1 generation, as was the case in the original approach. To identify
172 individuals with the heritable longevity trait, we constructed the LRC score.

173

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

174

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

183

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

185 To determine what proportion of long-lived ancestors could be associated with the survival
186 of F3 descendants, we calculated LRC scores for all F3 descendants and subsequently defined
187 9 mutually exclusive LRC groups (g) of F3 descendants: LRC_g1=0, LRC_g2=[>0 & <0.1],

188 LRC_g3=[≥0.1 & <0.2], LRC_g4=[≥0.2 & <0.3], LRC_g5=[≥0.3 & <0.4], LRC_g6=[≥0.4 & <0.5],
189 LRC_g7=[≥0.5 & <0.6], LRC_g8=[≥0.6 & <0.7], LRC_g9=[≥0.7 & ≥1.0]. For each group of F3
190 descendants we explored whether they have a survival benefit compared to the general
191 population by estimating SMRs (Figure 2). F3 descendants without any long-lived ancestors
192 (LRC score of 0) had a survival pattern that resembled the general population (SMR=0.97
193 (95%CI=0.93-1.01)). Similarly, we observed a survival pattern that resembled the general
194 population for F3 descendants with up to 20% long-lived ancestors (group 2 and 3,
195 SMR=0.97 (95%CI=0.91-1.04) and SMR=0.95 (95%CI=0.91-1.00) respectively). This shows
196 that the long-lived ancestors of group 2 and 3 F3 descendants were likely phenocopies
197 instead of genetically enriched long-lived persons. We observed a pattern of excess survival
198 for F3 descendants with more than 20% long-lived ancestors. The weakest significant effect
199 was observed for group 3, with an SMR of 0.84 (95%CI=0.80-0.89) which is comparable to
200 the excess survival of the F3 descendants of the singleton F1 cases in the original approach
201 (first part of the results). The strongest significant effect was observed for group 8, with an
202 SMR of 0.56 (95%CI=0.45-0.69). Hence, the higher the degree of long-lived ancestors, the
203 lower the SMR. This indicates that the more long-lived ancestors an F3 descendant has, the
204 higher the level of excess survival of these F3 descendants is compared to the general
205 population, and the more likely that genetic effects drive the transmission of longevity.

206

207 Using the LRC score family method we defined a new case and control group in the F3
208 generation, which is based on the presence or absence of longevity among the ancestors of
209 the F3 generation and potential excess survival or mortality in the F3 generation itself
210 (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.

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.

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.

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, 36, 41, 7–12, 15, 17). This knowledge about the familial clustering of longevity was utilized to construct longevity ranking scores such as

303 the Family Mortality History Score (FMHS) (42), the est(SE) which subsequently was
304 developed into the FLOSS (43, 44), the Longevity Family Score (LFS) which is an adaptation to
305 the est(SE) and the FMHS (12), and finally a method was developed to rank individuals by the
306 survival of their ancestors, the Familial Excess Longevity (FEL) score (45). The FMHS, FLOSS,
307 and LFS all resemble excess survival of a family (FMHS focus on parents and FLOSS and LFS
308 focus on siblings) compared to the general population. The FEL score focuses on excess
309 survival, defined as the difference between a person's attained and expected age, derived
310 from an accelerated failure time model. This excess survival was estimated for ancestors and
311 from this a score was created for individuals. Although these scores all resemble a
312 continuous familial estimate of a lifespan advantage and not necessarily longevity, they
313 might be used as an inclusion tool for cases in genetic (association) studies (43). However,
314 these scores are not based on a clear longevity definition that represents the heritable
315 longevity trait and they always require an arbitrary and difficult to interpret decision to
316 make a cutoff in the scores so that they resemble longevity. In addition, the majority of the
317 scores are not based on ancestors and thus do not capture the full family history of
318 longevity. As such, the scores are not suitable to establish the proportion of family members
319 that should be long-lived in order to properly define long-lived cases with a heritable
320 longevity trait and thus, increase the power of genetic longevity studies.
321
322 To overcome these issues, we developed a novel tool based on mapping the longevity of a
323 person's ancestors, the LRC score. The LRC score can be used to select carriers of the
324 heritable longevity trait (cases) and controls who resemble the general population. Another
325 interesting group, which we did not address in this article, is composed of persons without
326 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 (37, 46–54). Moreover, it was shown that the power to identify genetic variants decreases at an equal rate to the level of misclassification (37). 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 versa (37). Interestingly, when known, methods exist to adjust for the level of phenotypic misclassification (47–49, 51, 55), 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

375 offspring survival (13). Similarly, previous studies showed only a minor (56) or no (13, 57)
376 influence of early and mid-life environmental covariates, such as farm ownership, parental
377 literacy, parental and own occupation, and birth intervals, on the association between
378 parental longevity and offspring survival. We, however, cannot completely rule out that
379 other, unobserved non-genetic familial effects may affect our results. The observed excess
380 survival of F2 case and control group spouses in the original approach seem to be an
381 exception, as we observed a survival advantage for both groups. This is likely a form of
382 ascertainment bias because mortality data for this group was difficult to obtain in the Dutch
383 Personal Records Database, leading to an overrepresentation of high ages at death. These
384 observations add to the mixed results about whether spouses married to a long-lived person
385 have a survival advantage themselves (8, 12–15, 58).
386
387 Our results have two important implications. First, existing studies based on living study
388 participants who have not yet reached the ages to express longevity, but have ancestral
389 survival data, such as UK Biobank, can now better distinguish cases by incorporating a
390 liability based on the LRC score. Second, new studies would obtain a maximum power to
391 identify loci that promote survival to the highest ages in the population when cases are
392 included with at least 30% (LRC≥0.30) ancestors who belong at least to the top 10% survivors
393 of their birth cohort and are themselves among the 10% longest lived. More extreme
394 selections can be made on the survival percentile by for example focusing on the top 5% or
395 1% survivors, and/or on the proportion of long-lived family members, for example 50%.
396 However, this is not strictly necessary and might unnecessarily lead to limited sample sizes
397 (13). In addition, controls without any ancestors living to the top 10% survivors of their birth
398 cohort should be included, as their mortality pattern resembles that of the general

399 population. Finally, for future research it may be interesting to study the environmental
400 factors causing the longevity in those individuals who were long-lived but had no long-lived
401 ancestors. If our proposed method is consistently applied across studies, the comparative
402 nature of longevity studies may improve and facilitate the discovery of novel genetic
403 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 (38–40). These 37,137 persons were subsequently identified in the Dutch population registers and followed in the registers throughout their entire life course (39, 40, 59). 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 (40, 59). 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

426 Mandemakers and Cor Munnik, Historical Sample of the Netherlands. Project Genes, Germs
427 and Resources. Dataset LongLives. Release 2016.01.
428
429 **Extending the HSN study**
430 For this study we extended the pedigrees until we identified the living descendants for all
431 1326 families. From the population registers we know the names of all F2 descendants and
432 we subsequently identified the F2 descendants on personal cards (PCs) and personal lists
433 (PLs) which were obtained from the Dutch central bureau of genealogy (CBG). These PLs and
434 PCs were respectively introduced in 1939 and 1994 as the individualized and subsequently,
435 digitized form of the population register (40). The cards contain similar information to the
436 population registers and because of privacy legislation could only be obtained for deceased
437 persons, one year after they passed away (https://cbg.nl/bronnen/cbg-
438 verzamelingen/persoons kaarten-en-lijsten). Hence, from these cards we obtained similar
439 life course and mortality information for the F2 descendants as for the F1 IPs and we
440 obtained the names of their descendants (F3). We repeated this procedure until no cards
441 could be obtained anymore, which was at the F3 generation. Thus the F4 generation was not
442 identified on the PCs of PLs anymore. In conclusion, we identified and obtained information
443 for the F2 descendants, F2 spouses, F3 descendants, F3 spouses, and F4 descendants
444 (Figure1A and Table1). We will refer to this database as the HSN case/control database.
445
446 **Obtaining information for the living descendants**
447 In a final step we obtained as much mortality information as possible for the relatives of the
448 identified persons and we obtained addresses, as contact information for the living
449 descendants. This information was obtained through the Personal Records Database (PRD)

450 which is managed by Dutch governmental service for identity information.
451 https://www.government.nl/topics/personal-data/personal-records-database-brp. The PRD
452 contains PL information on all Dutch citizens (alive and death) and PC information is
453 continuously added. We were granted permission (permission number: 2016-0000364875)
454 to obtain the date of death, date of last observation, current living address, and identifying
455 information such as names of a person's father and mother to double check if the person
456 identified in the PRD was identical to the person in our HSN case/control database. Using the
457 PRD we were able to obtain addresses for F3 and F4 descendants and additional mortality
458 information for F2 descendants, F2 spouses, F3 descendants, F3 spouses, and F4
459 descendants (Figure1A and Table1). The final database covers 57,337 persons from 1,326
460 five-generational families (F0-F4) and contains F1 index persons (IPs), their parents (F0),
461 siblings (F1), spouses (F1), and 3 consecutive generations of descendants (F2-F4) and
462 spouses (F2-F4), connecting deceased persons to their living descendants.
463
464 **Exclusion criteria and study population**
465 Due to the nature of the source data there is a high rate of missing mortality information for
466 F0 parents, F1 spouses and F1 siblings, which we therefore excluded from analyses. We
467 further excluded F4 descendants because 92% is still alive (Table 1 and Figure 1B). The final
468 study population covers 37,825 persons from 1,326 three-generational families (F1-F3) and
469 contains F1 index persons (IPs), 2 consecutive generations of descendants (F2-F3) and 2
470 generations of spouses (F2-F3).
471
472 **Statistical analyses**

Statistical analyses were conducted using R version 3.4.1 (60). 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 (61, 62). 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

497 deaths are given by the sum of all individual cumulative hazards based on the birth cohort
498 and sex specific lifetables of the Dutch population. An SMR between 1 and 0 indicates excess
499 survival, an SMR of 1 indicates that the study population shows a similar survival to the
500 reference population, and an SMR above 1 indicates excess mortality. The SMR can be
501 estimated conditional on the specific age at which an individual starts to be observed in the
502 study (correction for left truncation). This was necessary to avoid selection bias if individuals
503 in a study population were not at risk of dying before a specific age of entry.
504

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

505

506 Where d_i=dead status (1=dead, 0=alive), H_{t0i}=sex and birth year specific cumulative hazard
507 based on lifetable, t_i=timing, referring to age at death or last observation, t_{0i}=liftable age
508 conditioning, for example from birth (t_{0i}=0), N= group sample size. Exact CIs were derived
509 (63) and compared to bootstrap CIs for family data (12). Both methods provided identical CIs
510 and thus, to reduce the amount of computational time necessary to estimate bootstrap CIs,
511 we estimated exact CIs.
512

513 **Longevity Relatives Count score**

514 Based on the results of a recent study which shows that longevity is heritable beyond the
515 10% survivors of their birth cohort and that multiple family members, such as parents and/or
516 aunts and uncles, should belong to the top 10% survivors (13) we constructed a novel score
517 that summarizes the familial history of longevity, the Longevity Relatives Count score (LRC).
518

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

519

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

537

538 **Survival analysis (Cox-type random effects regression model)**

539 To investigate the extent of a survival difference between the family F3 case and control
540 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 t_{ij} is the age at death for person 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 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 KR, Janssens A, Slagboom PE (2017) Historical demography and longevity genetics: Back to the future. *Ageing Res Rev* 38:28–39.
2. Herskind AM, et al. (1996) The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900. *Hum Genet* 97(3):319–323.
3. Kaplanis J, et al. (2018) Quantitative analysis of population-scale family trees with millions of relatives. *Science* 360(6385):171–175.
4. Ruby JG, et al. (2018) Estimates of the Heritability of Human Longevity Are Substantially Inflated due to Assortative Mating. *Genetics* 210(3):1109–1124.
5. Houde L, Tremblay M, Vézina H (2008) Intergenerational and Genealogical Approaches for the Study of Longevity in the Saguenay-Lac-St-Jean Population. *Hum Nat* 19(1):70–86.
6. Perls TT, et al. (2002) Life-long sustained mortality advantage of siblings of centenarians. *Proc Natl Acad Sci* 99(12):8442–8447.
7. Willcox BJ, Willcox DC, He Q, Curb JD, Suzuki M (2006) Siblings of Okinawan centenarians share lifelong mortality advantages. *J Gerontol A Biol Sci Med Sci* 61(4):345–54.
8. Jarry V, Gagnon A, Bourbeau R (2013) Survival advantage of siblings and spouses of centenarians in 20th-century Quebec. *Can Stud Popul* 39(3–4):67.
9. Dutta A, et al. (2013) Longer lived parents: Protective associations with cancer incidence and overall mortality. *Journals Gerontol - Ser A Biol Sci Med Sci* 68(11):1409–1418.
10. Gudmundsson H, Gudbjartsson DF, Kong A (2000) Inheritance of human longevity in

Iceland. *Eur J Hum Genet EJHG* 8(10):743–749.

11. Sebastiani P, Nussbaum L, Andersen SL, Black MJ, Perls TT (2016) 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(3):340–346.

12. Berg N van den, et al. (2018) Longevity Around the Turn of the 20th Century: Life-Long Sustained Survival Advantage for Parents of Today's Nonagenarians. *Journals Gerontol Ser A* 73(10):1295–1302.

13. van den Berg N, et al. (2019) Longevity defined as top 10% survivors and beyond is transmitted as a quantitative genetic trait. *Nat Commun* 10(1):35.

14. Pedersen JK, et al. (2017) The Survival of Spouses Marrying Into Longevity-Enriched Families. *Journals Gerontol Ser A Biol Sci Med Sci* 72(1):109–114.

15. Schoenmaker M, et al. (2006) Evidence of genetic enrichment for exceptional survival using a family approach: the Leiden Longevity Study. *Eur J Hum Genet* 14(1):79–84.

16. Ljungquist B, Berg S, Lanke J, McClearn GE, Pedersen NL (1998) 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(6):441–446.

17. Terry DF, et al. (2004) Lower All-Cause, Cardiovascular, and Cancer Mortality in Centenarians' Offspring. *J Am Geriatr Soc* 52(12):2074–2076.

18. Westendorp RGJ, et al. (2009) Nonagenarian Siblings and Their Offspring Display Lower Risk of Mortality and Morbidity than Sporadic Nonagenarians: The Leiden Longevity Study. *J Am Geriatr Soc* 57(9):1634–1637.

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

382(9903):1507–1513.

20. Christensen K, McGue M, Petersen I, Jeune B, Vaupel JW (2008) Exceptional longevity does not result in excessive levels of disability. *Proc Natl Acad Sci* 105(36):13274–13279.

21. Andersen SL, Sebastiani P, Dworkis D a., Feldman L, Perls TT (2012) 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(4):395–405.

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

23. Eline Slagboom P, van den Berg N, Deelen J (2018) Phenome and genome based studies into human ageing and longevity: An overview. *Biochim Biophys Acta - Mol Basis Dis* 1864(9):2742–2751.

24. Deelen J, et al. (2014) Genome-wide association meta-analysis of human longevity identifies a novel locus conferring survival beyond 90 years of age. *Hum Mol Genet* 23(16):4420–4432.

25. Broer L, et al. (2015) GWAS of longevity in CHARGE consortium confirms APOE and FOXO3 candidacy. *Journals Gerontol - Ser A Biol Sci Med Sci* 70(1):110–118.

26. Willcox BJ, et al. (2008) FOXO3A genotype is strongly associated with human longevity. *Proc Natl Acad Sci* 105(37):13987–13992.

27. Flachsbart F, et al. (2009) Association of FOXO3A variation with human longevity confirmed in German centenarians. *Proc Natl Acad Sci* 106(8):2700–2705.

28. Pilling LC, et al. (2017) Human longevity: 25 genetic loci associated in 389,166 UK biobank participants. *Aging (Albany NY)* 9(12):2504–2520.

630 29. Joshi PK, et al. (2017) Genome-wide meta-analysis associates HLA-DQA1/DRB1 and
631 LPA and lifestyle factors with human longevity. *Nat Commun* 8(1):1–13.
632 30. Partridge L, Deelen J, Slagboom PE (2018) Facing up to the global challenges of ageing.
633 *Nature* 561(7721):45–56.
634 31. Sebastiani P, et al. (2017) Four Genome-Wide Association Studies Identify New
635 Extreme Longevity Variants. *J Gerontol A Biol Sci Med Sci* 72(11):1453–1464.
636 32. Flachsbart F, et al. (2016) Immunochip analysis identifies association of the
637 RAD50/IL13 region with human longevity. *Aging Cell* 15(3):585–588.
638 33. Zeng Y, et al. (2016) Novel loci and pathways significantly associated with longevity.
639 *Sci Rep* 6(1):21243.
640 34. Oeppen J, Vaupel JW (2002) Broken limits to life expectancy. *Science* 296(5570):1029–
641 1031.
642 35. Vaupel JW (1998) Biodemographic Trajectories of Longevity. *Science (80-)*
643 280(5365):855–860.
644 36. Deluty JA, Atzmon G, Crandall J, Barzilai N, Milman S (2015) The influence of gender
645 on inheritance of exceptional longevity. *Aging (Albany NY)* 7(6):412–418.
646 37. Gordon D (2005) Factors affecting statistical power in the detection of genetic
647 association. *J Clin Invest* 115(6):1408–1418.
648 38. Mandemakers K (2000) Historical sample of the Netherlands. In P. K. Hall, R. McCaa, &
649 G. Thorvaldsen (Eds.). *Handbook of International Historical Microdata for Population*
650 *Research*, pp 149–177.
651 39. Mandemakers K (2010) https://socialhistory.org/en/hsn/hsn-releases. *HSN 201001*
652 *release*.
653 40. van den Berg N, et al. (2018) Families in Comparison: An individual-level comparison

of life course and family reconstructions between population and vital event registers. *SocArXiv*. Available at: https://osf.io/preprints/socarxiv/h2w8t/.

41. Kemkes-Grottenthaler A (2004) Parental effects on offspring longevity—evidence from 17th to 19th century reproductive histories. *Ann Hum Biol* 31(2):139–158.

42. Rozing MP, et al. (2010) Familial longevity is associated with decreased thyroid function. *J Clin Endocrinol Metab* 95(11):4979–4984.

43. Sebastiani P, et al. (2009) A Family Longevity Selection Score: Ranking Sibships by Their Longevity, Size, and Availability for Study. *Am J Epidemiol* 170(12):1555–1562.

44. Arbeeva LS, et al. (2018) How Well Does the Family Longevity Selection Score Work: A Validation Test Using the Utah Population Database. *Front Public Heal* 6(October). doi:10.3389/fpubh.2018.00277.

45. Kerber RA, Brien EO, Smith KR, Cawthon RM (2001) Familial Excess Longevity in Utah Genealogies. *Journals Gerontol Ser A Biol Sci Med Sci* 56(3):130–139.

46. Buyske S, Yang G, Matise TC, Gordon D (2009) 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(4):287–292.

47. Gordon D, Haynes C, Yang Y, Kramer PL, Finch SJ (2007) Linear trend tests for case–control genetic association that incorporate random phenotype and genotype misclassification error. *Genet Epidemiol* 31(8):853–870.

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

49. Rekaya R, Smith S, Hay EH, Farhat N, Aggrey S (2016) Analysis of binary responses with outcome-specific misclassification probability in genome-wide association studies.

Appl Clin Genet Volume 9:169–177.

50. Ji F, Yang Y, Haynes C, Finch SJ, Gordon D (2006) 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(1). doi:10.2202/1544-6115.1184.

51. Edwards BJ, Haynes C, Levenstien MA, Finch SJ, Gordon D (2005) Power and sample size calculations in the presence of phenotype errors for case/control genetic association studies. *BMC Genet* 6:18.

52. Royall DR, Chiodo LK, Polk MJ (2004) Misclassification Is Likely in the Assessment of Mild Cognitive Impairment. *Neuroepidemiology* 23(4):185–191.

53. Bross I (1954) Misclassification in 2 X 2 Tables. *Biometrics* 10(4):478.

54. Platz EA, De Marzo AM, Giovannucci E (2004) Prostate cancer association studies: Pitfalls and solutions to cancer misclassification in the PSA era. *J Cell Biochem* 91(3):553–571.

55. Smith S, Hay E, Farhat N, Rekaya R (2013) Genome wide association studies in presence of misclassified binary responses. *BMC Genet* 14(1):124.

56. Danzhen You, Danan Gu, Zeng Yi (2010) Familial Transmission of Human Longevity Among the Oldest-Old in China. *J Appl Gerontol* 29(3):308–332.

57. Gavrilov LA, Gavrilova NS (2015) Predictors of Exceptional Longevity: Effects of Early-Life and Midlife Conditions, and Familial Longevity. *North Am Actuar J* 19(3):174–186.

58. Montesanto A, et al. (2011) The genetic component of human longevity: analysis of the survival advantage of parents and siblings of Italian nonagenarians. *Eur J Hum Genet* 19(8):882–886.

59. Mandemakers K (2002) Building life courses datasets from population registers by the

historical sample of the Netherlands. *History and Computing*, pp 87–107.

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

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

62. Carolina T, Uijvenhoven L, van der Laan J (2009) Overlevingstafels en longitudinale analyse. *CBS* (10010):1–25.

63. Ulm K (1990) Simple method to calculate the confidence interal of a standardized mortality ratio (SMR). *Am J Epidemiol* 131(2):373–375.

Figure legends

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.

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.

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%).

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.

Tables

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

Role	Number	Deceased (%)	Alive (%)	Female (%)	Range Birth cohort	Mean age (sd)	Median age (sd)	missing_age (%)
Cases (Original design)								
F0 parents*	1768	899 (51)	0 (0)	884 (50)	1788-1858	64.23 (16.17)	66.52 (18.01)	869 (49)
F1 IPs	884	884 (100)	0 (0)	422 (50)	1860-1875	85.79 (4.59)	84.99 (4.95)	0 (0)
F1 siblings*	3439	1889 (55)	0 (0)	1699 (50)	1843-1908	39.12 (33.06)	42.45 (51.79)	1550 (45)
F1 spouses*	944	581 (62)	0 (0)	502 (53)	1833-1904	66.89 (15.1)	69.8 (15.48)	363 (38)
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)								
F0 parents*	884	476 (54)	0 (0)	442 (50)	1791-1858	61.77 (15.49)	63.34 (17.7)	408 (46)
F1 IPs	442	442 (100)	0 (0)	214 (48)	1860-1875	51.71 (5.71)	52.88 (6.21)	0 (0)
F1 siblings*	1740	1039 (60)	0 (0)	832 (48)	1851-1897	34.9 (32.39)	28.33 (41.44)	701 (40)
F1 spouses*	465	233 (50)	0 (0)	246 (53)	1837-1890	64.66 (16.64)	67.92 (16.74)	232 (50)
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)

761 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
762 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

763 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
764 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
765 standard deviation.

766

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

	Case group		Control group		Adjustment for
Role	SMRs	Number (N)	SMRs	Number (N)	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

768 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
769 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
770 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
771 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
772 truncation, but not a combination between the two.
773

774 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			

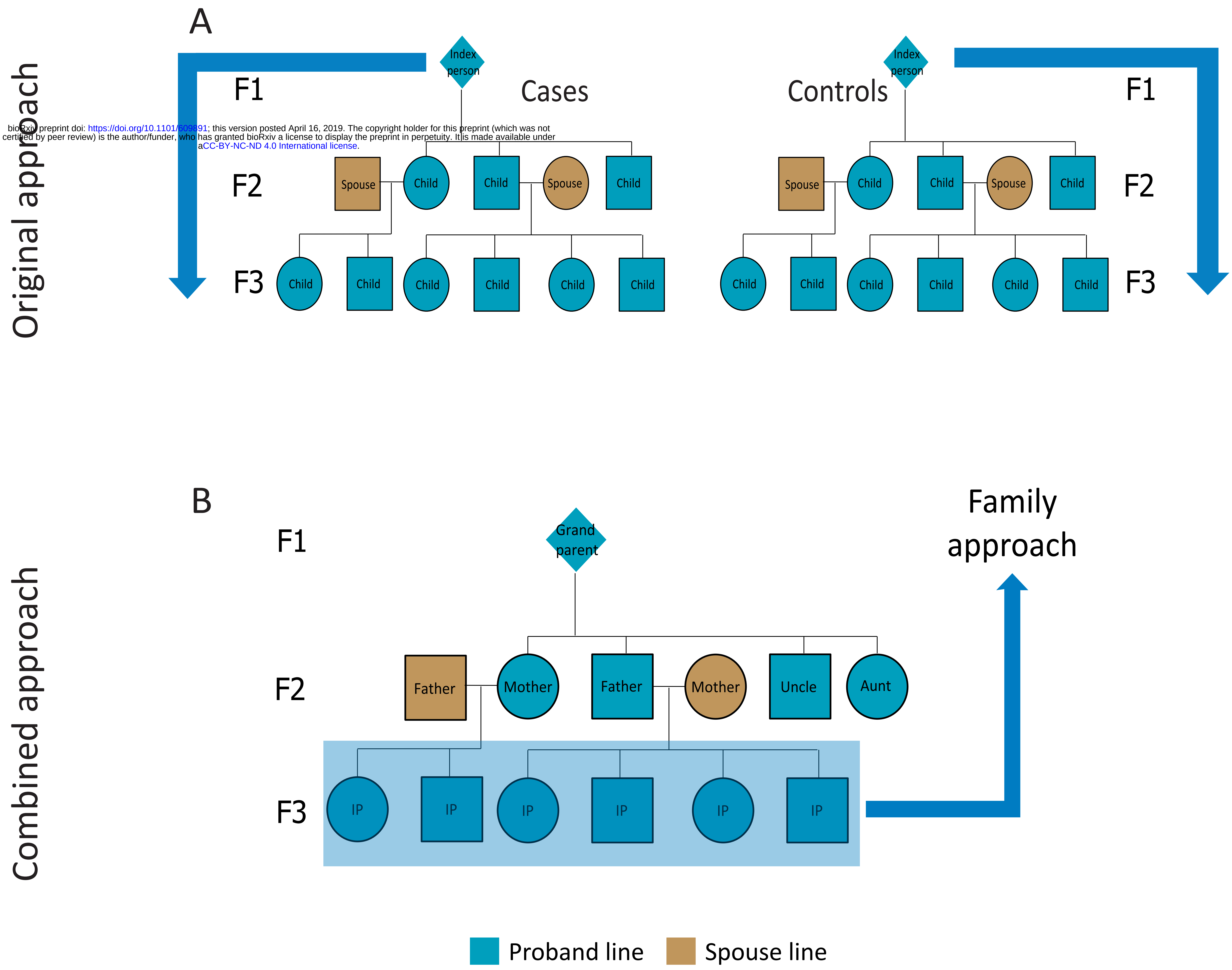
775 Table 3A corresponds to the CH curves of panel a of figure 4. Means represent a mean for a continuous variable and a proportion for a categorical variable. When the p-value was lower than 1.00e-15 we indicated
776 the P-value as <1.00*10-15. SES = socio-economic status, OCC = occupational coding scheme of 1950, CI = confidence interval, CH = cumulative hazard. P-values are estimated with cox regression. F3 children with
777 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
778 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
779 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
780 the cox model.
781

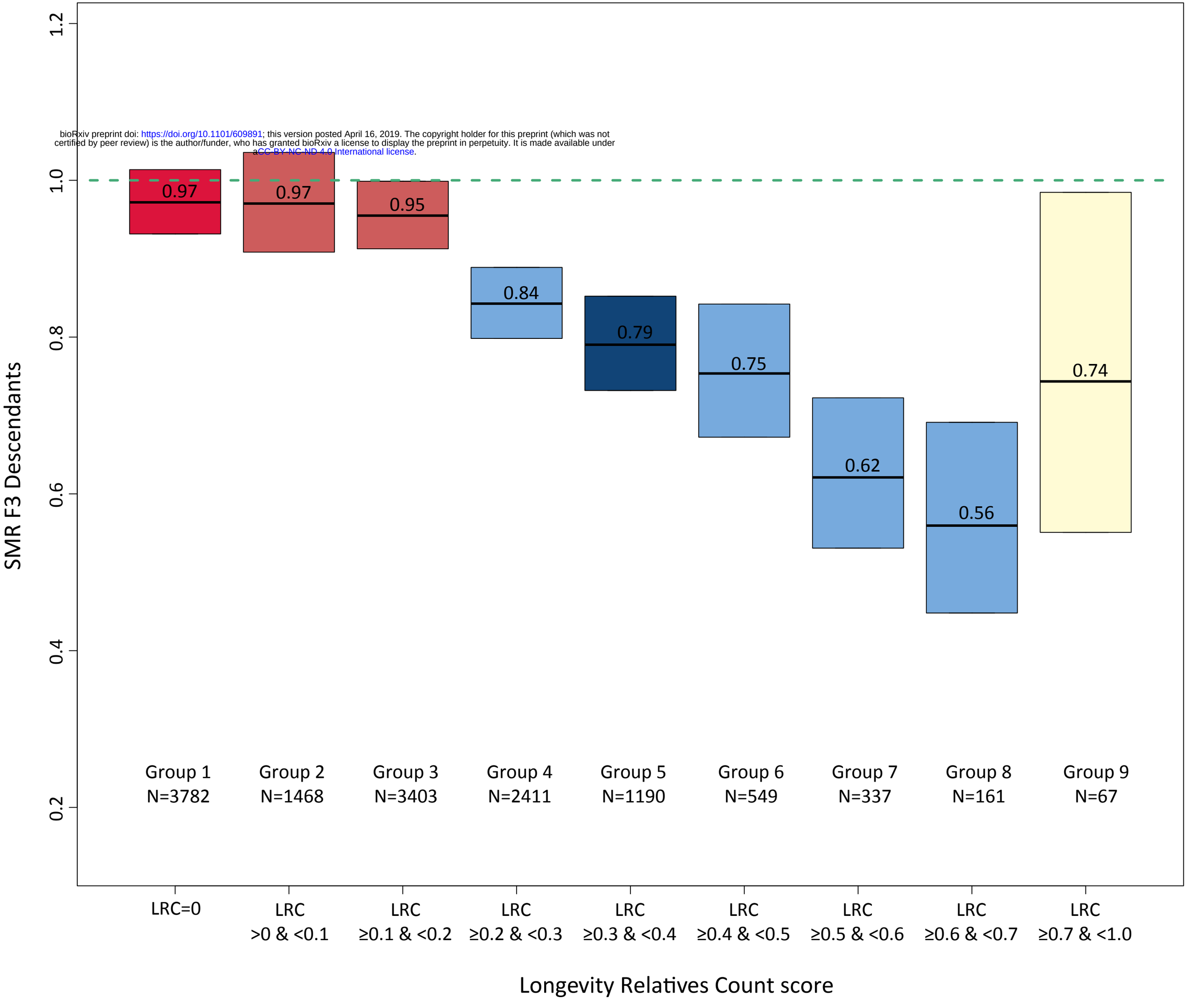
782 Table 4: Standardized Mortality Ratio for different F3 descendant groups

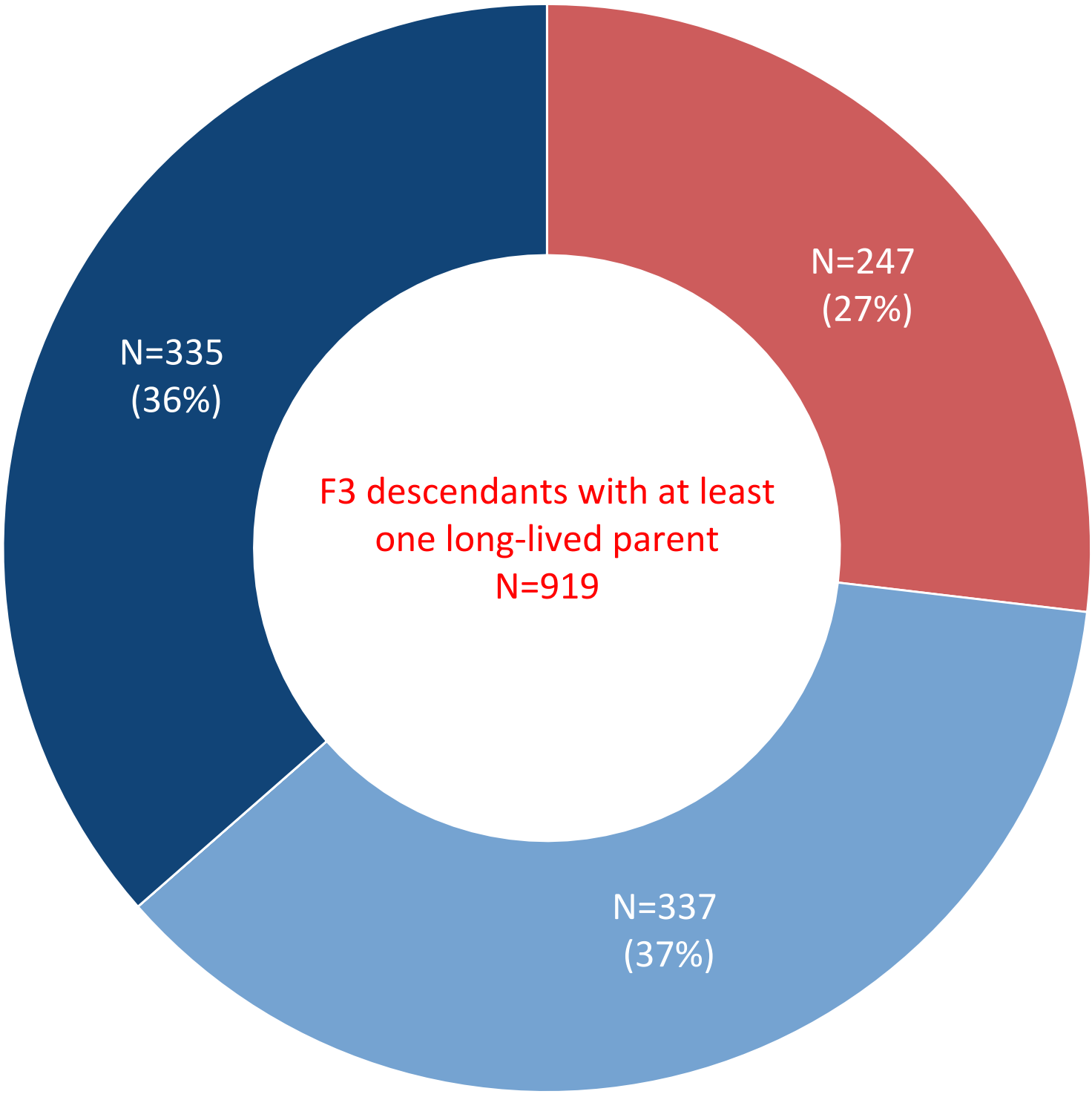
Group	SMR	N
Cases		
F3 descendant with at least one long-lived grandparent	0.86 (95%CI=0.83-0.89)	4986
F3 descendant with at least one long-lived parent	0.84 (95%CI=0.76-0.92)	852
F3 descendant with ≥ 30% long-lived ancestors (LRC ≥ 30%)	0.74 (95%CI=0.70-0.78)	2304
F3 descendant with ≥ 50% long-lived 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

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

786



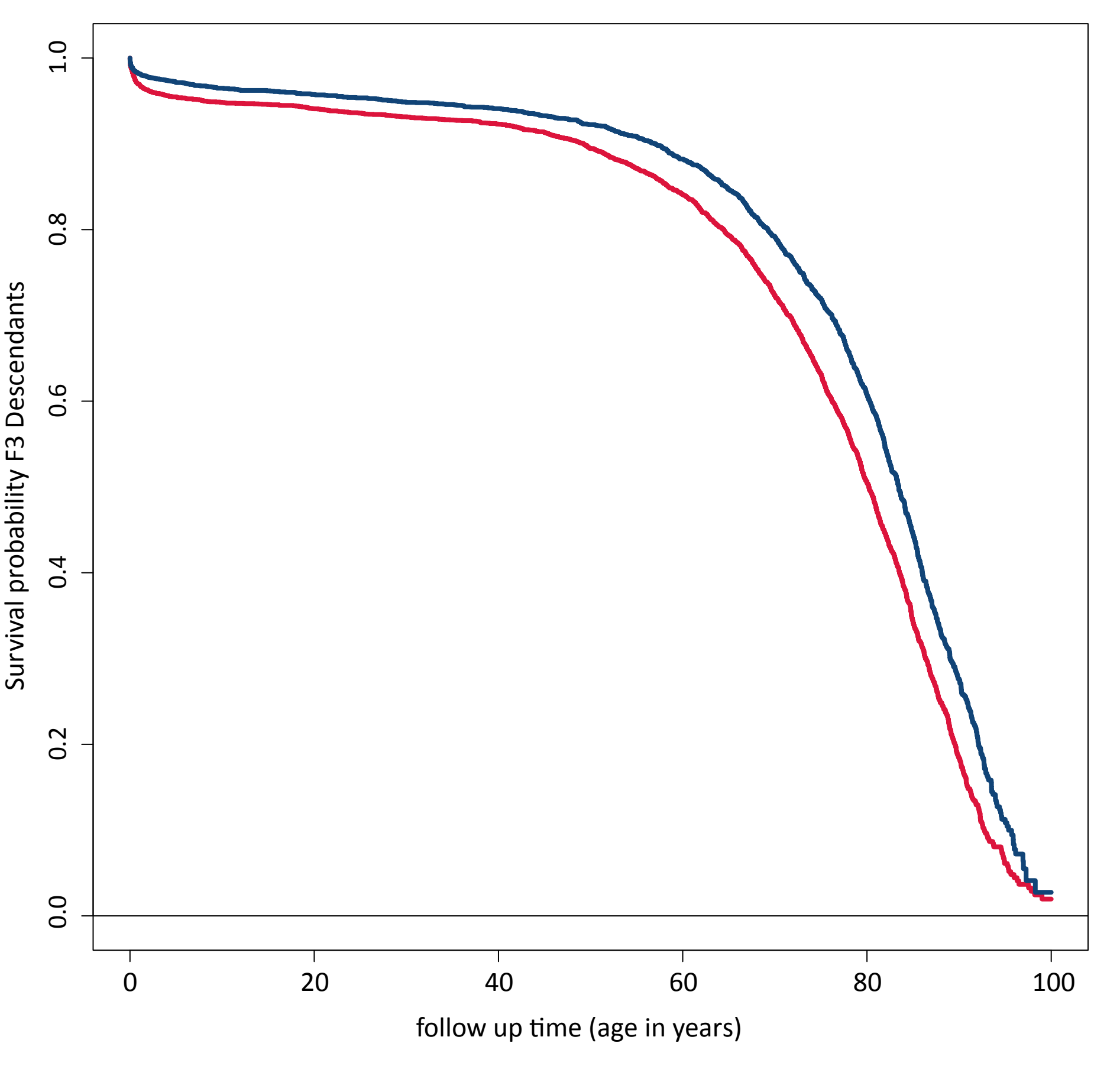




F3 descendants with 30% long-lived ancestors
(LRC ≥ 0.30) and an SMR < 1

F3 descendants with ≥ 20% and less than 30% long-lived ancestors
(LRC ≥ 0.20 and < 0.30) and an SMR <1

F3 descendants with ≥ 0% and less than 20% long-lived ancestors
(LRC ≥ 0 and < 0.20) and an SMR = 1


Survival probability F3 Descendants
1.0
0.8
0.6
0.4
0.2
0.0
0
20
40
60
80
100
follow up time (age in years)