Title:

Longevity Relatives Count score identifies heritable longevity carriers and suggests 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:**

Name: Niels van den Berg

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

Address: Albinusdreef 2, 2333 ZA Leiden the Netherlands

Phone: +31 (0) 71 5269738

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[5–17]. Identifying longevity loci is important because these loci likely represent key mechanisms of a life-long decreased mortality[15,16], decreased morbidity[9,12,18] 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[1,16]. 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,

66 on selecting multiple family members with the same trait[15,38,39,43,44] and usually focus on a
67 single group of relatives, such as parents[15,43] or siblings[39] of cases.
68
69 As the definition of heritable longevity was not yet established, the construction and
70 application of the family selection scores have not yet been adopted as state of the art in
71 longevity research. As such, the majority of genealogical[5,6,10–14,45] and genetic studies[24–26,31–
72 34] focus only on single, and thus including sporadic, long-lived individuals (singletons), with
73 some exceptions focusing for example on parental age[28,29] or multiple siblings[7,25]. In previous
74 work, we showed that longevity defined as top 10% survivors or more extreme is
75 transmitted to subsequent generations[16]. With this, a consistent definition of longevity was
76 provided that is also adopted in the largest longevity GWAS up to now[32]. In addition, we
77 showed that every additional long-lived relative independently contributes to the survival
78 advantage of study participants, according to their genetic distance[16]. As such, there is room
79 to incorporate these novel insights into family selection scores to gain knowledge about the
80 extent that longevity needs to cluster in families in order to include individuals with the
81 heritable longevity trait and increase the power of genetic studies.
82
83 Here, we aim to establish the proportion of ancestral blood relatives that should be long-
84 lived (top 10% survivors of their birth cohort or more extreme) in order to observe a survival
85 advantage in their descendants and incorporate these insights into a novel family score to
86 define cases with the heritable longevity trait for inclusion in genetic studies. For our
87 analyses we use the data available in the Historical Sample of the Netherlands (HSN) for the
88 period between 1860 and 1875 which is based on Dutch citizens[46–48]. We primarily identify
89 cases who died beyond 80 years (N=884, on average top 10% survivors of their birth cohort),

90 allowing us to select on more extreme ages at death, and controls who died between 40 and
91 59 years (N=442). We extend this filial (F) 1 generation data with a parental and 3
92 descendant generations of individual life course and mortality data and refer to the data as
93 the HSN case/control dataset. We subsequently exclude groups with high rates of missing
94 mortality information and where the majority was still alive (Supplementary Figure 4). This
95 study covers 37,825 persons from 1,326 three-generational families (F1-F3) and contains F1
96 index persons (IPs), 2 consecutive generations of descendants (F2-F3) and 2 generations of
97 spouses (F2-F3) (Table 1). The dataset is unique in that it covers multiple generations and
98 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

122 selected either F3 descendants with at least one top 10% grandparent, at least one top 10%
123 parent, or with grandparents who died between 40 and 59 years (their children (parents)
124 resembled the general population). In a final step, we focused on the F3 descendants with at
125 least one long-lived parent and calculated LRC scores within this F3 group to determine if
126 parents transmitted their longevity more frequently if they were part of a long-lived
127 (LRC≥0.30) family (Figure 1B). The analysis steps are summarized in Supplementary Figure 5
128 and an overview of the available data per group and generation is shown in Table 1.

129

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

131 Focusing on the original approach (Figure 1A), we determined to what extent longevity is
132 transmitted in the original case and the control group by estimating SMRs per generation for
133 all cases and controls separately. Table 2 shows that F1 cases had a similar survival pattern
134 to birth cohort members of the same sex, indicating that they resemble a representative
135 group of random Dutch persons aged ≥ 80 years and born between 1860 and 1875. The SMR
136 for the descendants of the cases (F2 case descendants) was 0.87 (95%CI=0.84-0.89),
137 indicating 13% less deaths than expected based on individuals from a similar birth cohort
138 and sex. From here we refer to this as 13% excess survival (or, if appropriate, excess
139 mortality) compared to the general population. The descendants of controls (F2 control
140 descendants) had a similar survival pattern to the general population (SMR=1.01
141 (95%CI=0.96-1.05)). The spouses of the F2 case and control descendants surprisingly also
142 showed a pattern of excess survival ($SMR_{case_F2spouses}$=0.89 (95%CI=0.85-0.94) and
143 $SMR_{control_F2spouses}$ =0.9 (95%CI=0.83-0.97)). Next we observed 14% (95%CI=11%-16%) excess
144 survival compared to the general population for F3 descendants of the F1 cases, whereas F3

145 control descendants resembled the general population (SMR=0.96 (95%CI=0.93-1.00)) just as
146 observed in the F2 generation. The spouses of both F3 groups resembled the general
147 population ($SMR_{case_F3spouses}$=1.00 (95%CI=0.95-1.05) & $SMR_{control_F3spouses}$=1.07 (95%CI=0.99-
148 1.15)). We conclude that two descendant generations of cases, who belong on average to
149 the top 10% survivors, have 13-14% excess survival compared to the general populations
150 and that the descendants of controls resemble the general population.

151

152 To explore to what extent the survival of F2 and F3 descendants depends on the extremity of
153 the longevity of their parents, we calculated SMRs for F2 and F3 case and control
154 descendants with increasing parental longevity (for example, a parent belonged to the top
155 10%, 5%, or 1% survivors). We observed that the SMR decreased in descendants when
156 defining parental longevity in terms of more extreme survival percentiles. This was the case
157 for descendants of both the IP cases and controls although the effects were stronger in the
158 descendants of the cases, especially in F3, since this group is now selected to have long-lived
159 parents and grandparents (Supplementary table 1). This illustrates that selection on single
160 long-lived persons belonging on average to the top 10% survivors, as we did for the IP
161 selection, leads only to a modest transmission of longevity in two generations (max 14%).
162 Likely, the control group includes misclassified persons of which the descendants do live
163 longer, whereas the case group includes long-lived persons that do not transmit longevity to
164 their descendants (potentially these are phenocopies). Such misclassification can jeopardize
165 genetic studies immensely. To be able to evaluate living persons as potential carriers of the
166 heritable longevity trait in genetic studies, we constructed and validated a familial longevity
167 score.

168

169 **Constructing the Longevity Relatives Count score**

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

174

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

175

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

184

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

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

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

207

208 Using the LRC score family method we defined a new case and control group in the F3
209 generation, which is based on the presence or absence of longevity among the ancestors of
210 the F3 generation and potential excess survival or mortality in the F3 generation itself
211 (Figure 1B). The F3 family controls include all F3 descendants without any long-lived

212 ancestors (LRC score of 0, N=4,166). To define the F3 family cases we chose an LRC cutoff
213 based on a trade-off between the size and the uncertainty, given by the sample size, of the
214 SMR. The F3 family cases include all F3 descendants with at least 30% long-lived ancestors
215 (LRC score ≥ 0.30 (N=2,526)). Even if F3 family cases are not long-lived themselves, their
216 survival reflects the presence of longevity of their ancestors, which is transmitted by their
217 parents. Similarly, F3 controls reflect the absence of longevity of their ancestors.
218 Supplementary Figure 1 shows the variation in lifespan of the F3 family case and control
219 descendants. F3 descendants with more than 0% and up to 20% long-lived ancestors (LRC
220 score >0 and < 0.2) did not express excess survival (N=5,340). The F3 descendants with an
221 LRC score ≥0.2 and < 0.30 showed some excess survival compared to the general population,
222 but the size of the SMR was considered too low to enter our family case definition. Hence,
223 we denoted them as non-classified (N=2,639).

224

225 **Strong survival advantage and genetic enrichment for F3 family cases**

226 To validate the LRC score, we investigate survival differences, measured as age at death or
227 last observation, between the F3 family cases and controls and used a Cox-type random
228 effects (frailty) regression model to adjust for within-family relations of the F3 descendants.
229 Figure 4 and table 3A show that F3 cases have a 25% (95%CI=18-31%) lower hazard of dying
230 than F3 controls, even after adjustment for sibship size, birth year, and sex. The difference
231 between the cases and controls became increasingly more pronounced when confining the
232 cases to a higher proportion of long-lived ancestors, for example an LRC score of 0.40, 0.50,
233 or 0.60, reflecting 40%, 50%, or 60% long-lived ancestors (Supplementary figure 2). The
234 strongest effect was observed for those with an LRC score ≥ 0.60 (hazard ratio (HR) of 0.62

235 (95%CI=0.50-0.77)). The mortality pattern for the spouses of these F3 cases resembled that
236 of the F3 controls (HR=0.94 (95%CI=0.82-1.07),Table 3B) and the general population
237 (SMR=0.92 (95%CI=0.83-1.02)). The survival of the spouses, equal to the F3 controls and the
238 general population, in addition to the absence of effects of environmental covariate
239 adjustment, indicates that environmental factors were likely of limited influence to the
240 observed survival benefit of the F3 cases as defined by our novel family based definition.
241 Hence, the observed survival benefit of F3 cases likely represents a genetic longevity
242 component.

243

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

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

258 cases, we observed 26% (95%CI=22%-30%) excess survival compared to the general
259 population and this increased to 38% (95%CI=31%-45%) when applying a 50% threshold to
260 the family method. For the identification of controls both methods seem to preform equally
261 well, with almost identical SMRs of around 1. This indicates that the F3 controls, whether
262 defined by having no long-lived ancestors or by grandparents dying between 40 and 50
263 years, have a similar survival pattern to the general population. We conclude that, at least
264 for cases, the family method provides a better contrast in excess survival compared to the
265 general population and seems to better represent the heritable longevity trait.

266

267 Since the F3 descendants with ≥ 30% long-lived ancestors have a stronger survival advantage
268 than those with at least one long-lived parent, it is possible to get an indication of how many
269 F3 descendants did not appear to have a survival advantage compared to the general
270 population, even though at least one parent was long-lived. This is relevant in view of case
271 definitions used in large genetic studies into longevity. Figure 3 and Supplementary Figure 3
272 show that 919 F3 descendants had a long-lived parent. Out of those 919 F3 descendants,
273 247 (27%) had more than 0% but less than 20% long-lived ancestors (LRC > 0 and < 0.20) and
274 thus as a group had an SMR that resembled the general population (Supplementary Figure
275 3D). The other 672 (73%) had exactly, or more than 20% long-lived ancestors (LRC ≥ 0.20)
276 and thus, as a group, showed excess survival compared to the general population
277 (Supplementary Figure 3B and C). These results suggest that if living persons are selected as
278 case in genetic studies on the basis of one long-lived parent, 27% of these persons is unlikely
279 to be a carrier of the longevity trait. Persons defined as 30% long-lived ancestors, on the
280 other hand would be potential carriers.

281 # Discussion

282 Human longevity is heritable and clusters in specific families. Studying the familial clustering
283 of longevity in these families is important to improve our understanding of genetic factors
284 promoting longevity and healthy aging. The main observations supporting this are (1) In the
285 original approach, we observed 14% excess survival of the cases compared to their birth
286 cohort for two subsequent generations (F2-F3) while in the controls no such benefit was
287 observed, (2) in the combined approach, the excess survival of the F3 cases compared to the
288 general population was 26-38% depending on the proportion of long-lived family members
289 being 30-50% and these estimates strongly overlap to the survival difference between the F3
290 family cases and controls based on the Cox models, (3) no excess survival as compared to
291 the birth cohort and general population was observed for F3 controls, spouses of cases or
292 controls and neither for F3 cases with up to 20% long-lived ancestors. The analyses in the
293 HSN case/control dataset provides strong evidence that longevity is transmitted for at least
294 2 subsequent generations and only when at least 20% of all ancestors are long-lived.
295 Moreover, the family cases seem to be genetically enriched for longevity while the controls
296 resemble the general population. Finally, 27% of the F3 descendants showed a survival
297 pattern similar to the general population even though they had at least one long-lived
298 parent.

299

300 Previous family studies, usually focusing on 2 generations and single individuals, showed that
301 siblings and children of long-lived persons lived longer than first degree ancestors of non-
302 long-lived persons or population controls[5–7,9–15,45,49]. This knowledge about the familial
303 clustering of longevity was utilized to construct longevity ranking scores such as the Family

304 Mortality History Score (FMHS)[43], the est(SE) which subsequently was developed into the
305 FLOSS[39,44], the Longevity Family Score (LFS) which is an adaptation to the est(SE) and the
306 FMHS[15], and finally a method was developed to rank individuals by the survival of their
307 ancestors, the Familial Excess Longevity (FEL) score[38]. The FMHS, FLOSS, and LFS all resemble
308 excess survival of a family (FMHS focus on parents and FLOSS and LFS focus on siblings)
309 compared to the general population. The FEL score focuses on excess survival, defined as the
310 difference between a person's attained and expected age, derived from an accelerated
311 failure time model. This excess survival was estimated for ancestors and from this a score
312 was created for individuals. Although these scores all resemble a continuous familial
313 estimate of a lifespan advantage and not necessarily longevity, they might be used as an
314 inclusion tool for cases in genetic (association) studies[39]. However, these scores are not
315 based on a clear longevity definition that represents the heritable longevity trait and they
316 always require an arbitrary and difficult to interpret decision to make a cutoff in the scores
317 so that they resemble longevity. In addition, the majority of the scores are not based on
318 ancestors and thus do not capture the full family history of longevity. As such, the scores are
319 not suitable to establish the proportion of family members that should be long-lived in order
320 to properly define long-lived cases with a heritable longevity trait and thus, increase the
321 power of genetic longevity studies.
322
323 To overcome these issues, we developed a novel tool based on mapping the longevity of a
324 person's ancestors, the LRC score. The LRC score can be used to select carriers of the
325 heritable longevity trait (cases) and controls who resemble the general population. Another
326 interesting group, which we did not address in this article, is composed of persons without
327 any long-lived ancestors who themselves are long-lived. It may be interesting to study

328 environmental factors contributing to a long and healthy life in this group. Here we used the
329 LRC score to construct a novel family case and control group and observed a survival
330 advantage for F3 case descendants, even when their parents were not necessarily long-lived,
331 supporting the idea that a beneficial genetic component was transmitted. Likewise, the
332 increase in the LRC score ≥ 20% associated with an increase in survival advantage for F3
333 descendants. This indicates that every additional ancestor contributes to the survival
334 advantage of F3 descendants and confirms our previous findings in the LINKing System for
335 historical demography (LINKS) data and the Utah Population Database (UPDB)[16]. This
336 additive pattern is not readily expected if the observations are due to non-genetic factors,
337 such as wealth, that cluster in families. The fact that none of the environmental confounders
338 (sex, birth year, and sibship size) affected the survival differences between the family cases
339 and controls provided additional evidence for the transmission of a genetic component. A
340 final indication for the genetic enrichment of the family cases is based on the observed
341 mortality pattern for the spouses of the family cases and controls which resembled the
342 family controls themselves and the general population.
343
344 We observed that F3 descendants with at least one long-lived parent had less excess survival
345 than a subset of these F3 descendants who had at least 30% long-lived ancestors and this
346 difference increased when at least 50% of their ancestors were long-lived. These results
347 indicate that some parents were long-lived but might not have transmitted their longevity to
348 the subsequent F3 generation. In fact, 27% of the F3 descendants with at least one long-
349 lived parent did not have an LRC ≥0.20 and, as a group, did not express excess survival.
350 Hence the parents of theses 27% F3 descendants were sporadically long-lived as they did not
351 transmit their longevity. Thus, genetic studies may benefit from a case definition, where

352 cases are long-lived and have at least 30% long-lived ancestors, as current genetic studies,
353 based on long-lived cases, often not include ancestral longevity in their case selection. Even
354 though our data did not allow for an exact misclassification analysis, studies showed that the
355 level of phenotypic misclassification in case and control annotation has a strong inhibiting
356 effect on the power to identify variants in genetic association studies, including GWAS[42,50–58].
357 Moreover, it was shown that the power to identify genetic variants decreases at an equal
358 rate to the level of misclassification[42]. For example, a study with 95% power to detect an
359 association based on a sample of 100 cases and controls when there are no phenotypic
360 errors may actually have only 75% power when 20% of the cases are misclassified as controls
361 and vice versa[42]. Interestingly, when known, methods exist to adjust for the level of
362 phenotypic misclassification[51–53,55,59], providing opportunities for specific application in
363 genetic longevity research.
364
365 Due to the nature of the HSN data we could not use the mortality data for the parents (F0),
366 siblings (F1), and spouses (F1) of the F1 IPs. Mortality data was less incomplete for the F2
367 and F3 spouses (table 1A) but there was still a relatively large number of missing mortality
368 data. Thus, for future studies with this dataset it might be interesting to extend the mortality
369 information for these groups. Furthermore, life course data was only present for persons
370 with an identified personal card or personal list (details in the methods section).
371 Consequently, socio-economic status and religion was only available for a small part (around
372 15%) of the F3 descendants with an unequal share of availability between men and women.
373 This led to the exclusion of these environmental factors from our analyses. Even though we
374 could not adjust our models for socio-economic status and religion, it is known from other
375 studies that those factors are not influencing the association between parental longevity and

376 offspring survival[16]. Similarly, previous studies showed only a minor[60] or no[16,61] influence of
377 early and mid-life environmental covariates, such as farm ownership, parental literacy,
378 parental and own occupation, and birth intervals, on the association between parental
379 longevity and offspring survival. We, however, cannot completely rule out that other,
380 unobserved non-genetic familial effects may affect our results. The observed excess survival
381 of F2 case and control group spouses in the original approach seem to be an exception, as
382 we observed a survival advantage for both groups. This is likely a form of ascertainment bias
383 because mortality data for this group was difficult to obtain in the Dutch Personal Records
384 Database, leading to an overrepresentation of high ages at death. These observations add to
385 the mixed results about whether spouses married to a long-lived person have a survival
386 advantage themselves[7,11,15–17,62].
387
388 Our results have two important implications. First, existing studies based on living study
389 participants who have not yet reached the ages to express longevity, but have ancestral
390 survival data, such as UK Biobank, can now better distinguish cases by incorporating a
391 liability based on the LRC score. Second, new studies would obtain a maximum power to
392 identify loci that promote survival to the highest ages in the population when cases are
393 included with at least 30% (LRC≥0.30) ancestors who belong at least to the top 10% survivors
394 of their birth cohort and are themselves among the 10% longest lived. More extreme
395 selections can be made on the survival percentile by for example focusing on the top 5% or
396 1% survivors, and/or on the proportion of long-lived family members, for example 50%.
397 However, this is not strictly necessary and might unnecessarily lead to limited sample sizes[16].
398 In addition, controls without any ancestors living to the top 10% survivors of their birth
399 cohort should be included, as their mortality pattern resembles that of the general

400 population. Finally, for future research it may be interesting to study the environmental
401 factors causing the longevity in those individuals who were long-lived but had no long-lived
402 ancestors. If our proposed method is consistently applied across studies, the comparative
403 nature of longevity studies may improve and facilitate the discovery of novel genetic
404 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

427 Mandemakers and Cor Munnik, Historical Sample of the Netherlands. Project Genes, Germs
428 and Resources. Dataset LongLives. Release 2016.01.
429
430 **Extending the HSN study**
431 For this study we extended the pedigrees until we identified the living descendants for all
432 1326 families. From the population registers we know the names of all F2 descendants and
433 we subsequently identified the F2 descendants on personal cards (PCs) and personal lists
434 (PLs) which were obtained from the Dutch central bureau of genealogy (CBG). These PLs and
435 PCs were respectively introduced in 1939 and 1994 as the individualized and subsequently,
436 digitized form of the population register[48]. The cards contain similar information to the
437 population registers and because of privacy legislation could only be obtained for deceased
438 persons, one year after they passed away (https://cbg.nl/bronnen/cbg-
439 verzamelingen/persoons kaarten-en-lijsten). Hence, from these cards we obtained similar
440 life course and mortality information for the F2 descendants as for the F1 IPs and we
441 obtained the names of their descendants (F3). We repeated this procedure until no cards
442 could be obtained anymore, which was at the F3 generation. Thus the F4 generation was not
443 identified on the PCs of PLs anymore. In conclusion, we identified and obtained information
444 for the F2 descendants, F2 spouses, F3 descendants, F3 spouses, and F4 descendants
445 (Figure1A and Table1). We will refer to this database as the HSN case/control database.
446
447 **Obtaining information for the living descendants**
448 In a final step we obtained as much mortality information as possible for the relatives of the
449 identified persons and we obtained addresses, as contact information for the living
450 descendants. This information was obtained through the Personal Records Database (PRD)

451 which is managed by Dutch governmental service for identity information.
452 https://www.government.nl/topics/personal-data/personal-records-database-brp. The PRD
453 contains PL information on all Dutch citizens (alive and death) and PC information is
454 continuously added. We were granted permission (permission number: 2016-0000364875)
455 to obtain the date of death, date of last observation, current living address, and identifying
456 information such as names of a person's father and mother to double check if the person
457 identified in the PRD was identical to the person in our HSN case/control database. Using the
458 PRD we were able to obtain addresses for F3 and F4 descendants and additional mortality
459 information for F2 descendants, F2 spouses, F3 descendants, F3 spouses, and F4
460 descendants (Figure1A and Table1). The final database covers 57,337 persons from 1,326
461 five-generational families (F0-F4) and contains F1 index persons (IPs), their parents (F0),
462 siblings (F1), spouses (F1), and 3 consecutive generations of descendants (F2-F4) and
463 spouses (F2-F4), connecting deceased persons to their living descendants.
464

465 **Exclusion criteria and study population**

466 Due to the nature of the source data there is a high rate of missing mortality information for
467 F0 parents, F1 spouses and F1 siblings, which we therefore excluded from analyses. We
468 further excluded F4 descendants because 92% is still alive (Table 1 and Figure 1B). The final
469 study population covers 37,825 persons from 1,326 three-generational families (F1-F3) and
470 contains F1 index persons (IPs), 2 consecutive generations of descendants (F2-F3) and 2
471 generations of spouses (F2-F3).
472

473 **Statistical analyses**

474 Statistical analyses were conducted using R version 3.4.1[64]. We reported 95% confidence
475 intervals (CIs) and considered p-values statistically significant at the 5% level ($\alpha = 0.05$).
476
477 **Lifetables**
478 In the Netherlands, population based cohort lifetables are available from 1850 until
479 2019[65,66]. These lifetables contain, for each birth year and sex, an estimate of the hazard of
480 dying between ages x and x + n (hx) based on yearly intervals (n=1) up to 99 years of age.
481 Conditional cumulative hazards (Hx) and survival probabilities (Sx) can be derived using
482 these hazards. In turn, we can determine to which sex and birth year based survival
483 percentile each person of our study belonged to. For example: a person was born in 1876,
484 was a female, and died at age 92. According to the lifetable information this person
485 belonged to the top three percent survivors of her birth cohort, meaning that only three
486 percent of the women born in 1876 reached a higher age. We used the lifetables to calculate
487 the birth cohort and sex specific survival percentiles for all persons in the HSN case/control
488 study. This approach prevents against the effects of secular mortality trends over the last
489 centuries and enables comparisons across study populations[1,14]. Supplementary Figure 6
490 shows the ages at death corresponding to the top 10, 5, and 1 percent survivors of their
491 birth cohorts for the period 1850-1935.
492
493 **Standardized Mortality Ratios**
494 To indicate excess mortality or excess survival of groups, such as F2 case or control group
495 descendants in the HSN case/control study compared to Dutch birth cohort members of the
496 same sex, we used Standardized Mortality Ratios (SMRs). An SMR is estimated by dividing
497 the observed number of deaths by the expected number of deaths. The expected number of

498 deaths are given by the sum of all individual cumulative hazards based on the birth cohort
499 and sex specific lifetables of the Dutch population. An SMR between 1 and 0 indicates excess
500 survival, an SMR of 1 indicates that the study population shows a similar survival to the
501 reference population, and an SMR above 1 indicates excess mortality. The SMR can be
502 estimated conditional on the specific age at which an individual starts to be observed in the
503 study (correction for left truncation). This was necessary to avoid selection bias if individuals
504 in a study population were not at risk of dying before a specific age of entry.
505

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

506

507 Where d_i=dead status (1=dead, 0=alive), H_{t0i}=sex and birth year specific cumulative hazard
508 based on lifetable, t_i=timing, referring to age at death or last observation, t_{0i}=liftable age
509 conditioning, for example from birth (t_{0i}=0), N= group sample size. Exact CIs were derived [67]
510 and compared to bootstrap CIs for family data [15]. Both methods provided identical CIs and
511 thus, to reduce the amount of computational time necessary to estimate bootstrap CIs, we
512 estimated exact CIs.
513

514 **Longevity Relatives Count score**

515 Based on the results of a recent study which shows that longevity is heritable beyond the
516 10% survivors of their birth cohort and that multiple family members, such as parents and/or
517 aunts and uncles, should belong to the top 10% survivors[16] we constructed a novel score
518 that summarizes the familial history of longevity, the Longevity Relatives Count score (LRC).
519

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

Where k=1,…,N_i 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[16]. 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:

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

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 on DANS (https://dans.knaw.nl/en/front-page?set_language=en).

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

583 10. Willcox, B. J., Willcox, D. C., He, Q., Curb, J. D. & Suzuki, M. Siblings of Okinawan
584 centenarians share lifelong mortality advantages. *J. Gerontol. A. Biol. Sci. Med. Sci.* **61**,
585 345–54 (2006).
586 11. Jarry, V., Gagnon, A. & Bourbeau, R. Survival advantage of siblings and spouses of
587 centenarians in 20th-century Quebec. *Can. Stud. Popul.* **39**, 67 (2013).
588 12. Dutta, A. *et al.* Longer lived parents: Protective associations with cancer incidence and
589 overall mortality. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **68**, 1409–1418 (2013).
590 13. Gudmundsson, H., Gudbjartsson, D. F. & Kong, A. Inheritance of human longevity in
591 Iceland. *Eur. J. Hum. Genet. EJHG* **8**, 743–749 (2000).
592 14. Sebastiani, P., Nussbaum, L., Andersen, S. L., Black, M. J. & Perls, T. T. Increasing
593 Sibling Relative Risk of Survival to Older and Older Ages and the Importance of Precise
594 Definitions of "Aging," "Life Span," and "Longevity". *Journals Gerontol. Ser. A Biol. Sci.*
595 *Med. Sci.* **71**, 340–346 (2016).
596 15. Berg, N. van den *et al.* Longevity Around the Turn of the 20th Century: Life-Long
597 Sustained Survival Advantage for Parents of Today's Nonagenarians. *Journals*
598 *Gerontol. Ser. A* **73**, 1295–1302 (2018).
599 16. van den Berg, N. *et al.* Longevity defined as top 10% survivors and beyond is
600 transmitted as a quantitative genetic trait. *Nat. Commun.* **10**, 35 (2019).
601 17. Pedersen, J. K. *et al.* The Survival of Spouses Marrying Into Longevity-Enriched
602 Families. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **72**, 109–114 (2017).
603 18. Westendorp, R. G. J. *et al.* Nonagenarian Siblings and Their Offspring Display Lower
604 Risk of Mortality and Morbidity than Sporadic Nonagenarians: The Leiden Longevity
605 Study. *J. Am. Geriatr. Soc.* **57**, 1634–1637 (2009).
606 19. Christensen, K. *et al.* Physical and cognitive functioning of people older than 90 years:

607 a comparison of two Danish cohorts born 10 years apart. *Lancet* **382**, 1507–1513
608 (2013).
609 20. Christensen, K., McGue, M., Petersen, I., Jeune, B. & Vaupel, J. W. Exceptional
610 longevity does not result in excessive levels of disability. *Proc. Natl. Acad. Sci.* **105**,
611 13274–13279 (2008).
612 21. Andersen, S. L., Sebastiani, P., Dworkis, D. a., Feldman, L. & Perls, T. T. Health Span
613 Approximates Life Span Among Many Supercentenarians: Compression of Morbidity
614 at the Approximate Limit of Life Span. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **67A**,
615 395–405 (2012).
616 22. Shadyab, A. H. & LaCroix, A. Z. Genetic factors associated with longevity: A review of
617 recent findings. *Ageing Res. Rev.* **19**, 1–7 (2015).
618 23. Eline Slagboom, P., van den Berg, N. & Deelen, J. Phenome and genome based studies
619 into human ageing and longevity: An overview. *Biochim. Biophys. Acta - Mol. Basis Dis.*
620 **1864**, 2742–2751 (2018).
621 24. Deelen, J. *et al.* Genome-wide association meta-analysis of human longevity identifies
622 a novel locus conferring survival beyond 90 years of age. *Hum. Mol. Genet.* **23**, 4420–
623 4432 (2014).
624 25. Broer, L. *et al.* GWAS of longevity in CHARGE consortium confirms APOE and FOXO3
625 candidacy. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **70**, 110–118 (2015).
626 26. Willcox, B. J. *et al.* FOXO3A genotype is strongly associated with human longevity.
627 *Proc. Natl. Acad. Sci.* **105**, 13987–13992 (2008).
628 27. Flachsbart, F. *et al.* Association of FOXO3A variation with human longevity confirmed
629 in German centenarians. *Proc. Natl. Acad. Sci.* **106**, 2700–2705 (2009).
630 28. Pilling, L. C. *et al.* Human longevity: 25 genetic loci associated in 389,166 UK biobank

631 participants. *Aging (Albany. NY).* **9**, 2504–2520 (2017).

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

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

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

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

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

642 34. Zeng, Y. *et al.* Novel loci and pathways significantly associated with longevity. *Sci. Rep.*
643 **6**, 21243 (2016).

644 35. Oeppen, J. & Vaupel, J. W. Broken limits to life expectancy. *Science* **296**, 1029–1031
645 (2002).

646 36. Vaupel, J. W. Biodemographic Trajectories of Longevity. *Science (80-.).* **280**, 855–860
647 (1998).

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

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

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

654 40. Hujoel, M. L. A., Gazal, S., Loh, P., Patterson, N. & Alkes, L. Combining case-control

655 status and family history of disease increases association power. 1–20 (2019).
656 41. Liu, J. Z., Erlich, Y. & Pickrell, J. K. Case–control association mapping by proxy using
657 family history of disease. *Nat. Genet.* **49**, 325–331 (2017).
658 42. Gordon, D. Factors affecting statistical power in the detection of genetic association. *J.*
659 *Clin. Invest.* **115**, 1408–1418 (2005).
660 43. Rozing, M. P. *et al.* Familial longevity is associated with decreased thyroid function. *J.*
661 *Clin. Endocrinol. Metab.* **95**, 4979–4984 (2010).
662 44. Arbeeva, L. S. *et al.* How Well Does the Family Longevity Selection Score Work: A
663 Validation Test Using the Utah Population Database. *Front. Public Heal.* **6**, (2018).
664 45. Deluty, J. A., Atzmon, G., Crandall, J., Barzilai, N. & Milman, S. The influence of gender
665 on inheritance of exceptional longevity. *Aging (Albany. NY).* **7**, 412–418 (2015).
666 46. Mandemakers, K. Historical sample of the Netherlands. In P. K. Hall, R. McCaa, & G.
667 Thorvaldsen (Eds.). in *Handbook of International Historical Microdata for Population*
668 *Research* 149–177 (2000).
669 47. Mandemakers, K. https://socialhistory.org/en/hsn/hsn-releases. *HSN 2010.01 release*
670 (2010).
671 48. van den Berg, N. *et al.* Families in Comparison: An individual-level comparison of life
672 course and family reconstructions between population and vital event registers.
673 *SocArXiv* (2018).
674 49. Kemkes-Grottenthaler, A. Parental effects on offspring longevity—evidence from 17th
675 to 19th century reproductive histories. *Ann. Hum. Biol.* **31**, 139–158 (2004).
676 50. Buyske, S., Yang, G., Matise, T. C. & Gordon, D. When a Case Is Not a Case: Effects of
677 Phenotype Misclassification on Power and Sample Size Requirements for the
678 Transmission Disequilibrium Test with Affected Child Trios. *Hum. Hered.* **67**, 287–292

679 (2009).

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

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

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

689 54. Ji, F., Yang, Y., Haynes, C., Finch, S. J. & Gordon, D. Computing Asymptotic Power and
690 Sample Size for Case-Control Genetic Association Studies in the Presence of
691 Phenotype and/or Genotype Misclassification Errors. *Stat. Appl. Genet. Mol. Biol.* **4**,
692 (2006).

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

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

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

699 58. Platz, E. A., De Marzo, A. M. & Giovannucci, E. Prostate cancer association studies:
700 Pitfalls and solutions to cancer misclassification in the PSA era. *J. Cell. Biochem.* **91**,
701 553–571 (2004).

702 59. Smith, S., Hay, E., Farhat, N. & Rekaya, R. Genome wide association studies in

703 presence of misclassified binary responses. *BMC Genet.* **14**, 124 (2013).

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

706 61. Gavrilov, L. A. & Gavrilova, N. S. Predictors of Exceptional Longevity: Effects of Early-
707 Life and Midlife Conditions, and Familial Longevity. *North Am. Actuar. J.* **19**, 174–186
708 (2015).

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

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

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

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

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

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

720

721

722 # Figure legends

723

724 **Figure 1: Pedigree overview of the data structure**

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

735

736 **Figure 2: LRC score in mutually exclusive F3 descendant groups**

737 The figure shows Standardized Mortality Ratios for all F3 descendants without missing
738 mortality information. The F3 descendants are grouped into mutually exclusive groups based
739 on the Longevity Relatives Count (LRC) score. The LRC score represents the family approach
740 as illustrated in figure 1B. The dark red color of group one represents F3 descendants
741 without any long-lived (top 10%) ancestors and are denoted as family controls. The light red
742 represents F3 descendants who had more than 0 and less than 20% long-lived ancestors. The
743 light blue colors represent the F3 descendants with 20% or more long-lived ancestors. The
744 dark blue color represent our cut-off point for the family case definition. Hence all F3

745 descendants with 30% or more long-lived ancestors were considered family cases. The beige
746 color of group 9 shows that this bar represents all F3 ancestors with more than 70% long-
747 lived ancestors as their sample size was very low, we grouped them into one group.
748
749 **Figure 3: LRC score for F3 descendants with at least one long-lived parent**
750 This center of this doughnut figure shows all F3 descendants (N=919) with at least one long-
751 lived (top 10%) parent, ignoring the rest of the ancestors. Thus, at least means that they
752 could have more than 1 long-lived ancestor but we actively selected for the presence of only
753 1 such ancestor. The edges of the doughnut illustrate the number and proportion of these
754 919 F3 descendants with at least one long-lived parent who had 1. 30% or more long-lived
755 ancestors (LRC ≥ 0.30) and excess survival compared to the general population (SMR < 1),
756 N=335 (36%) 2. between 20% and 30% long-lived ancestors (LRC ≥ 0.20 and < 0.30) and
757 excess survival compared to the general population (SMR < 1), N=337 (37%) and 3. between
758 0% and 20% long-lived ancestors (LRC > 0.20 and < 0.20) and a similar survival pattern to the
759 general population (SMR ~ 1), N=247(27%).
760
761 **Figure 4: Survival differences between family based cases and their spouses**
762 This figure shows the survival curve for the difference in survival between the F3 family
763 cases and controls. The figure is connected to Table 3A which shows the Hazard Ratios
764 corresponding to the difference between the two curves. Blue color represent the cases, red
765 color represents the controls.
766

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)								
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.

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.

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

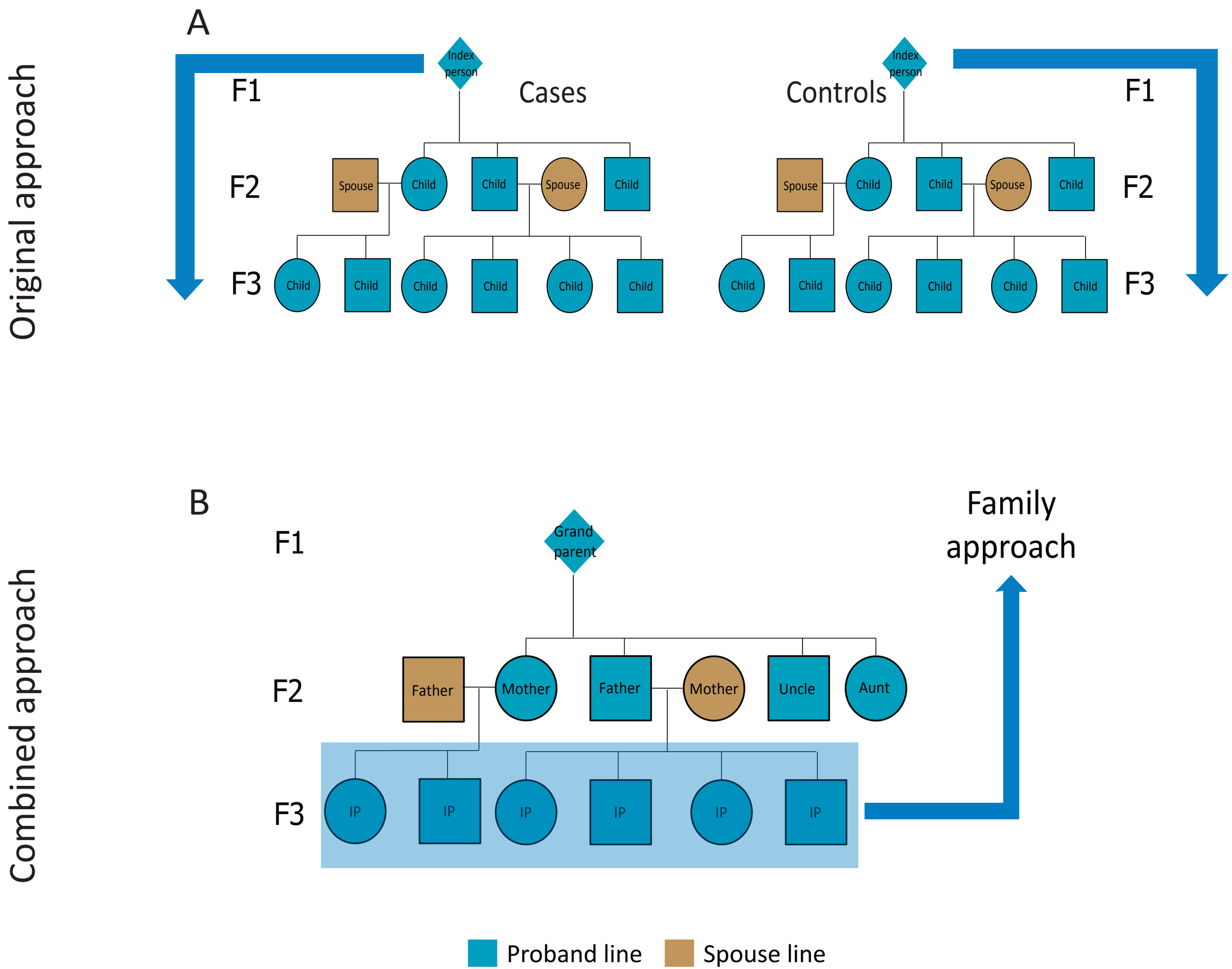
	A			B		
	N (mean)	**HR (95% CI)**	**P-value**	**N (mean)**	**HR (95% CI)**	**P-value**
Family based case/control group						
Control group (ref)	3714 (0.62)			3714 (0.50)		
Case group	2282 (0.38)	0.75 (0.69-0.82)	1.75e-10	2282 (0.30)	0.74 (0.68-0.80)	4.08e-12
Spouses of cases				541 (0.07)	0.94 (0.82-1.07)	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			

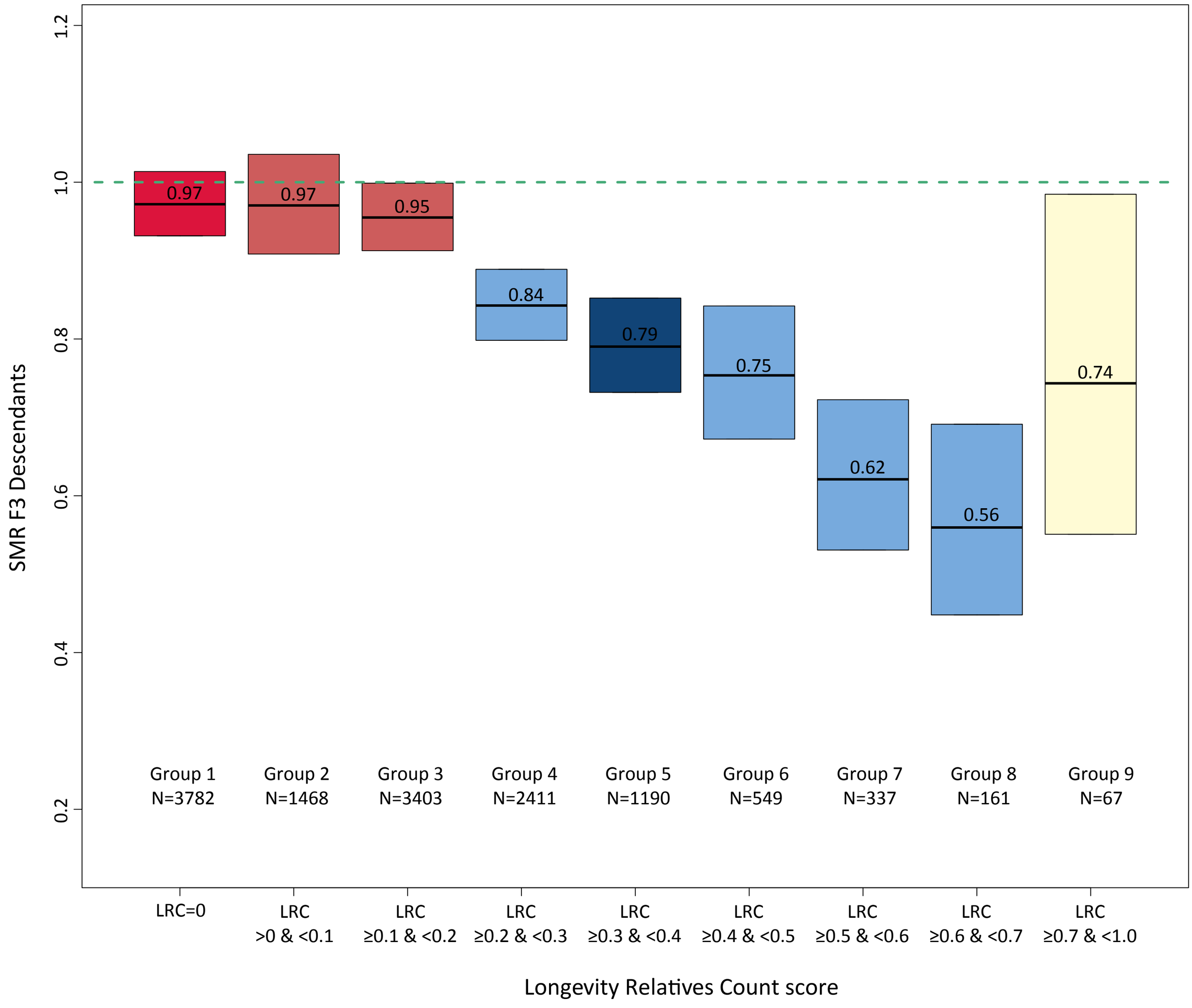
Table 3A corresponds to the CH curves of panel a of figure 4. Means represent a mean for a continuous variable and a proportion for a categorical variable. When the p-value was lower than 1.00e-15 we indicated the P-value as <1.00*10-15. SES = socio-economic status, OCC = occupational coding scheme of 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.

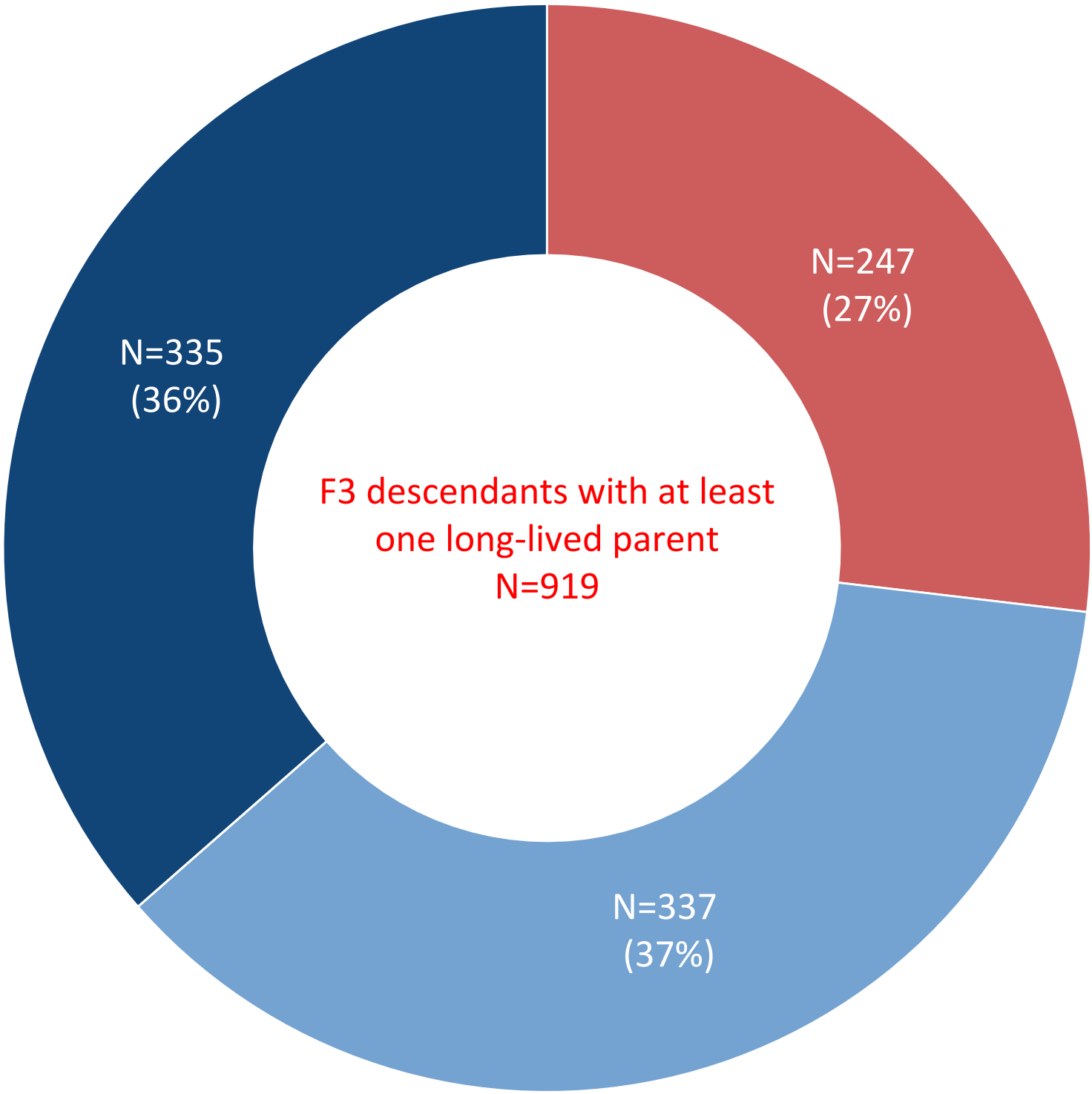
Table 4: Standardized Mortality Ratio for different F3 descendant groups

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.


A
Original approach
F1
F2
F3
Index person
Cases
Spouse
Child
Controls
B
Combined approach
Grand parent
Father
Mother
Uncle
Aunt
IP
Family approach
Proband line
Spouse line

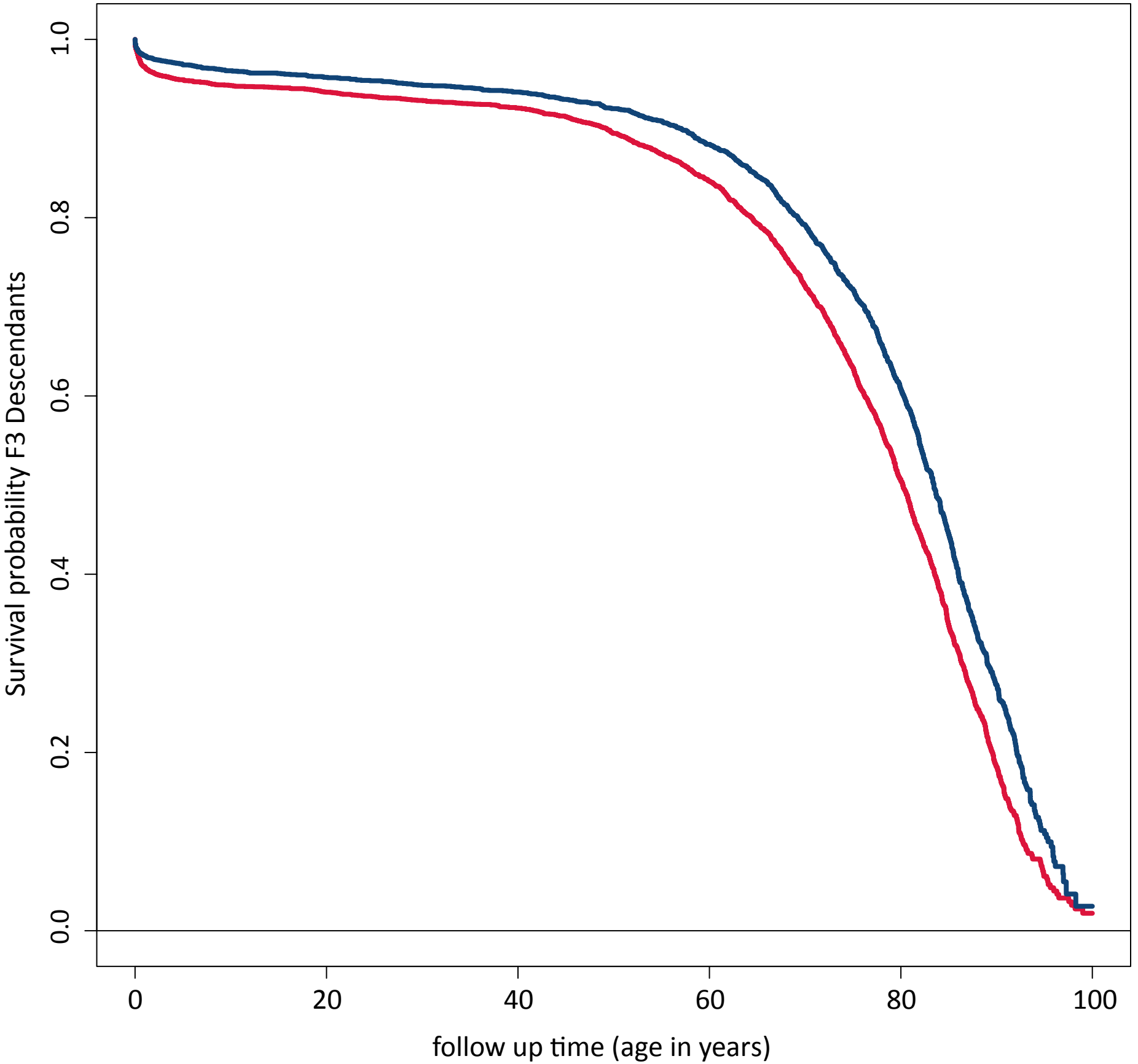

SMR F3 Descendants
1.2
1.0
0.8
0.6
0.4
0.2
0.97
0.97
0.95
0.84
0.79
0.75
0.62
0.56
0.74
Group 1
N=3782
Group 2
N=1468
Group 3
N=3403
Group 4
N=2411
Group 5
N=1190
Group 6
N=549
Group 7
N=337
Group 8
N=161
Group 9
N=67
LRC=0
LRC
>0 & <0.1
LRC
≥0.1 & <0.2
LRC
≥0.2 & <0.3
LRC
≥0.3 & <0.4
LRC
≥0.4 & <0.5
LRC
≥0.5 & <0.6
LRC
≥0.6 & <0.7
LRC
≥0.7 & <1.0
Longevity Relatives Count score



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)