Title:

Increasing number of long-lived ancestors associates with up to a decade of healthspan extension and a healthy metabolomic profile in mid-life

Authors:

Niels van den Berg[1,3,*], n.m.a.van_den_berg@lumc.nl

Mar Rodríguez-Girondo[2]

Ingrid K van Dijk[3]

P. Eline Slagboom[1,5,†]

Marian Beekman[1,†]

*: corresponding author, †: jointly supervising authors

Affiliations:

[1]. Department of Biomedical Data Sciences, section of Molecular Epidemiology, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden the Netherlands

[2]. Department of Biomedical Data Sciences, section of Medical Statistics, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden the Netherlands

[3]. Centre for Economic Demography, Department of Economic History, Lund University, Scheelevägen 15B, 223 63 Lund, Sweden

[4]. 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

25 Address: Albinusdreef 2, 2333 ZA Leiden the Netherlands

26 Phone: +31 (0) 71 5269738

27

28 **Keywords:**

29 ageing, longevity, disease incidence, medication, metabolome, LRC, Swedish register data,

30 healthy lifespan, aging

Abstract

Globally, the lifespan of populations increases but the healthspan is lagging behind. Previous research showed that survival into extreme ages (longevity) clusters in families as illustrated by the increasing lifespan of study participants with each additional long-lived family member. Here we investigate whether the healthspan in such families follows a similar quantitative pattern using three-generational data from two databases, LLS (Netherlands), and SEDD (Sweden). We study healthspan in 2,143 families containing index persons and two ancestral generations, comprising 17,539 persons with 25 follow-up years. Our results provide strong evidence that an increasing number of long-lived ancestors associates with up to a decade of healthspan extension. Further evidence indicates that members of long-lived families have a delayed onset of medication use, multimorbidity and, in mid-life, healthier metabolomic profiles than their partners. We conclude that in longevity families, both lifespan and healthspan are quantitatively linked to ancestral longevity, making such families highly suitable to identify protective mechanisms of multimorbidity.

Main

The human life expectancy has doubled over the past two centuries[1], reaching 82.1 years in Western European countries[2]. Although people started to live longer, the time spent in good physical and cognitive health did not rise similarly[2]. In fact, over 70% of the 65 year olds have at least one disease and over 50% have multimorbidity (2 disease or more)[3]. In contrast to the general population, some persons seem to become exceptionally old with a significantly lower chronic age-related disease burden (e.g. high blood pressure, malignancies, and type 2 diabetes) than the general population[4–17]. Moreover, the children of these exceptionally old persons have a delayed first disease onset[11,14,18,19]. These observations are mostly based on cross-sectional designs[4,10–14,17–20], so prospective studies into the development of first diseases and (multi)morbidity are needed. The study of long-lived families is important as they likely harbor gene-environment interactions which beneficially regulate molecular pathways involved in longevity, resistance to disease, resilience to negative side-effects of treatment and therefore healthy aging[8,21].

In our previous work we used data from millions of subjects in contemporary medical and historical family-tree databases to investigate the intergenerational transmission of human longevity[22–25]. We concluded that longevity, as a heritable trait, is primarily transmitted if persons belong to the oldest 10% survivors of their birth cohort and if at least 30% of their ancestors also belonged to the oldest 10% survivors[22,23]. Subsequently, we developed the Longevity Relatives Count (LRC) score as an instrument to quantify the number of long-lived family members and observed that the survival advantage of study participants increased with each additional long-lived family member, indicating additive effects[22]. As such, the LRC

score is an indicator of increased survival and longevity, and can therefore be used to enlarge the survival contrast in epidemiological data, thereby leading to more powerful genetic longevity studies. If the LRC score also represents healthspan as a quantitative trait (additive effects), this instrument can potentially be used in (genetic) studies to elucidate multi-morbidity limiting mechanisms.

We identified two issues that have not yet thoroughly been investigated: (1) whether from mid-life onward, health, medication use, disease incidence as well as the development of multi-morbidity are delayed over time, and (2) whether an increasing number of long-lived ancestors, as measured with the LRC score, represents not only lifespan as a quantitative trait but also healthspan. To address these issues, longitudinal life course and health data should ideally be investigated, preferably in large numbers of individuals. In addition, multiple generational family-tree information is required to investigate how the number of ancestral long-lived relatives relates to morbidity. Therefore we investigate chronic disease incidence and multimorbidity in long-lived families using up to 25 years of follow-up data. We further study whether an increasing number of long-lived ancestors, as measured by the LRC score, associates with a decreased incidence of chronic diseases. In addition, we investigate whether the families with the most extreme LRC scores have a healthy metabolomic profile in mid-life, representing overall health to complement the information on morbidity.

We use the data available in the Leiden Longevity Study (LLS, Netherlands) and the Swedish register data available in the Scanian Economic-Demographic Database (SEDD, Sweden). The LLS, initiated in 2002, was based on the inclusion of nonagenarian siblings. Also the middle

aged children (called index persons (IPs) in the current study) and their partners, as adult environment-matched controls, were included. The SEDD contains the entire population of 5 parishes and a town in Scania (southern Sweden), and as such does not contain any initial inclusion criteria. For the current study we identified in both datasets combined 2,143 three-generational families (F1-F3) containing IPs (F3) and their family members, comprising 17,539 persons in total. First, we examine whether LLS IPs and their partners differ in terms of disease and medication prevalence at the moment of study inclusion (2002-2004). Second, we investigate differences in disease incidence towards multimorbidity. Third, we study whether an increasing number of long-lived ancestors is associated with a decreased disease incidence in IPs using the Longevity Relatives Count (LRC) score[23] in both LLS and SEDD datasets. Finally, we compare mid-life health of LLS IPs with the highest LRC scores and their partners, using a previously developed metabolomics based score predicting mortality[26].

Results

Study populations

LLS IPs and their partners, serving as environment-matched controls, were included between 2002 and 2006 at the average age of 59 years. The study inclusion was based on nonagenarian siblings in the F2 generation. Hence, IPs (F3) were included if they had at least one long-lived F2 parent and F2 aunt or uncle (females ≥ 91 years and males ≥ 89 years). From inclusion onward, the IPs and their partners were followed over time, with a maximum mortality follow-up of 19 years (2002-2021) and maximum morbidity follow-up of 16 years (2002-2018). In 2021, 227 (14%) IPs and 113 (15%) partners were deceased and 1409 (84%) IPs and 619 (83%) partners were still alive. In 2018, 671 (40%) IPs and 324 (43%) partners had a disease diagnosis whereas 535 (32%) IPs and 206 (28%) partners did not have a disease diagnosis (Figure 1A and Table 1A).

SEDD IPs were followed over time from 1990, at an average age of 52 years, with a maximum mortality and morbidity follow-up of 25 years (1990-2015). In 2015, 694 (28%) IPs were deceased whereas 1,803 (72%) were still alive. Moreover, 1,190 (48%) IPs had a disease diagnosis whereas 1,307 (52%) IPs did not have a disease diagnosis (Figure 1B and Table 1B). From here we will refer to disease diagnoses as diseases, disease prevalence in cross-sectional analyses, and disease incidence in longitudinal analyses.

Disease prevalence at LLS study inclusion

To investigate whether LLS IPs and their partners differ in terms of disease and medication prevalence at the moment of study inclusion we conducted mixed-model logistic regression analysis. Figure 2A shows a 13% lower odds for age-related disease prevalence in IPs (OR=0.87 (95% CI=0.64−1.17)) compared to their partners. We further observed that IPs had a 20% lower odds for

metabolic diseases (OR=0.80 (95% CI=0.58−1.15)) and 16% higher odds for malignant diseases (OR=1.16 (95% CI=0.64−2.11) compared to their partners, albeit not statistically significant.

LLS IPs have a lower risk of using medication early in the study

Data on medication use was collected in the LLS between 2006 and 2008 (Supplementary Figure 1) and was available for 1254 LLS IPs (75%) and 588 partners (79%). We focused on ATC A-C type medication because they match the disease groups we investigate (see methods). To study whether LLS IPs had a lower medication use compared to their partners, we fitted mixed-model logistic regression analyses. Figure 2B shows that the odds of using ATC-A (alimentary tract and metabolism) type medication is 45% (OR=0.55 (95% CI=0.42−0.71)) lower for the offspring than for their partners. Similarly, the odds of using ATC-B (blood and blood forming organs) and ATC-C (cardiovascular system) type medication is 42% (OR=0.58 (95% CI=0.37−0.92)) and 48% (OR=0.52 (95% CI=0.38−0.71)) lower for the IPs. Our analyses thus indicate that, early on in the study, the IPs already had a significantly lower intake of metabolic and cardiovascular disease medication.

LLS IPs have a delayed first disease onset during follow-up

To investigate whether and to what extent the onset of first disease was delayed for the LLS IPs compared to their partners during 16 years of follow-up, we excluded persons who had ≥1 disease at inclusion. We therefore include 917 LLS IPs of whom 39 (4.3%) were deceased at the end of disease follow-up (2018) and 395 partners of whom 17 (4.3%) were deceased. We fitted random effect (frailty) Cox regressions and observed a Hazard Ratio (HR) of 0.79 (95% CI=0.65-0.97) for the age-related disease incidence between LLS IPS and their partners. This HR indicates that the yearly risk of age-related disease was 21% lower for the LLS IPs as compared to their partners. The LLS IPs had a 29% (HR=0.71 (95% CI=0.55-0.90)) lower risk of metabolic diseases and a 5% (HR=0.95 (95% CI=(0.70-1.31)) lower risk of malignant diseases (Table 2A and Supplementary Table 1A). In addition, Supplementary Figure 2 shows that 50% of the LLS IPs had an age-related disease at the age of 68

years whereas this was the case at the earlier age of 65.8 years for their partners. 50% of the LLS IPs had a metabolic disease at an age of 74.8 years, while this was the case at 68.6 years for their partners, indicating a median delay of metabolic disease diagnosis for LLS IPs of 6.2 years.

LLS IPs have a delayed onset of multimorbidity during follow-up

To study whether the delayed onset of first disease for LLS IPs extended to developing more than one disease (multimorbidity) during the 16 years of follow-up, we investigated the difference in time from inclusion to having two diseases within the same category (2 age-related, metabolic, or malignant diseases; Table 2B and Supplementary Table 1B). We observed that the yearly risk to develop 2 age-related diseases was 45% (HR=0.55 (95% CI=0.36-0.85) lower for the LLS IPs than for their partners, maximizing to a 49% (HR=0.51 (95% CI=0.26-0.97) difference for metabolic diseases. However, the yearly risk to develop 2 malignant diseases (HR=1.39 (95% CI=0.29-6.70)) did not significantly differ between LLS IPs and their partners. Supplementary Figure 3 shows the survival curves corresponding to Table 2B.

To study whether LLS IPs, who already had a disease, had a lower risk of getting a second disease, we investigated whether the time between first and specific second disease was longer for the LLS IPs than for their partners. Table 2C and Supplementary Table 2 show that the yearly risk to develop an age-related or a metabolic disease after already being diagnosed with an age-related disease, was 54% (HR=0.46 (95% CI=0.26-0.83), and 66% (HR=0.33 (95% CI=0.14-0.81)) lower for the IPs, respectively. For a malignant disease following the initial diagnosis of an age-related disease, no significant difference between LLS IPs and their partners was observed (HR=0.58 (95% CI=0.27-1.25)). Supplementary Figure 4 shows the survival curves corresponding to Table 2C. Moreover, sensitivity analyses showed that the HRs representing time from first to second disease were not affected by the group (metabolic or malignant) of first disease.

Increasing numbers of long-lived ancestors indicate a later disease onset in LLS and SEDD

In our previous work we developed the Longevity Relatives Count (LRC) score to quantify a person's number of long-lived ancestors[22,23]. The LRC score can be interpreted as a weighted proportion (ranging between 0 and 1)[23]. For example, an LRC score of 0.5 for an IP indicates 50% long-lived ancestors weighted by the genetic distance between IPs (and partners in LLS) and their ancestors. Here we investigate whether healthspan in LLS and SEDD is associated with the number of long-lived ancestors by testing whether an increasing LRC score of IPs is associated with a delay in disease onset and lower medication use in a longitudinal study design of the two independent databases; LLS and SEDD.

We conducted our analyses using two approaches. In the first approach we used the LRC score to enlarge the contrasts between the LLS IPs and their partners by defining four mutually exclusive groups: LRC_g1: IPs with an LRC ≥0.60, LRC_g2: IPs with an LRC [≥0.1 & <0.60], LRC_g3: partners with an LRC >0, and LRC_g4: partners with an LRC =0. We subsequently compared the disease incidence and medication use of LRC_g1-3 with LRC_g4, using Cox-type random effect (frailty) and linear mixed model regression analysis respectively. In the second approach, we calculated the LRC score in the LLS IPs and partners combined, allowing a quantitative definition of the LRC-score instead of defining groups. Using the quantitatively defined LRC-score we investigate whether an increasing LRC score associates with a decreasing first disease incidence, using Cox-type random effect (frailty) regression analysis. Finally, we validate the results obtained in the LLS by replicating our analysis in Swedish register data (SEDD).

First approach. When comparing the LLS IPs with an LRC score ≥0.60 (LRC_g1) with the partners who had an LRC score of 0 (LRC_g4) we observed a HR of 0.56 (95% CI=0.34-0.92) and 0.69 (95% CI=0.31-1.53) for time to first age-related and malignant disease, respectively (Table 3). Table 3 further shows that the healthspan benefit of the LRC_g1 group was most striking for the incidence of first metabolic

disease, for which the yearly risk was 53% lower (HR=0.47 (95% CI=0.25-0.87)). For comparison: HR's in Table 2 (not applying LRC score) are 0.79, 0.95 and 0.71 for age-related, malignant and metabolic diseases respectively, providing a strong indication that increasing numbers of long-lived ancestors are associated to a later disease incidence. To illustrate this comparison, Figure 3 shows the survival curves for the LLS IPs and partners (Panel A corresponding to Table 2) and the LRC groups (Panel B corresponding to Table 3). The figure shows how the LRC score maximizes the contrast: 50% of the LRC_g4 persons developed a first metabolic disease at the age of 68 years, whereas 50% of the LRC_g1 persons developed a first metabolic disease at the age of 81 years. Hence, the LRC_g1 persons delayed the age of metabolic disease onset with a pronounced 13 years difference. The survival curves of the other disease categories are presented in Supplementary Figure 5. Further benefit for LRC_g1 over LRC_g4 concerns development of multimorbidity and medication use, since an HR of 0.14 (95% CI=0.03-0.70) was observed for the time to develop 2 age-related diseases and an OR of 0.26 (95% CI=0.12-0.57) for medication use.

Second approach. We calculated the LRC score for the LLS IPs and their partners combined to avoid any grouping. Table 4A shows that with every 0.1 (10%) increase in LRC score, LLS F3 participants had a 5% (HR=0.95 (95% CI=0.91-0.99)) lower yearly risk to develop a first age-related disease. To illustrate the magnitude, this effect increases to 50% when all ancestors were long-lived (LRC score = 1). We further observed a 7% (HR=0.93 (95% CI=0.88-0.98)) lower yearly first metabolic and 3% (HR=0.97 (0.91-1.04)) lower malignant disease risk, though the latter effect was not statistically significant.

We validate the results obtained in the LLS by replicating our analysis in Swedish register data (SEDD). Table 4B shows that with every 10% increase in LRC score, the SEDD IPs have a 6% (HR=0.94 (95% CI=0.89-0.98)) lower yearly risk to develop a first age-related disease, 9% (HR=0.91 (0.87-0.96))

lower risk for metabolic and 5% (HR=0.95 (0.90-0.99)) for malignant disease. Moreover, the yearly risk of dying decreases 8% (HR=0.92 (0.87-0.97)) with every 10% increase in LRC score.

LLS IPs with an LRC score ≥0.60 already had a healthy metabolomic profile at inclusion

Our results point strongly towards protection from metabolic diseases for persons with an increasing number of long-lived ancestors as established with a high LRC score. We therefore investigated whether those with a high LRC score at the moment of inclusion in the LLS, indeed have a healthy circulating metabolomic profile that marks protection from disease at midlife. To estimate health in a quantitative parameter, we use a recently developed NMR-metabolomics based predictor of 5-10 years all-cause mortality across all ages from midlife onwards (from here MetaboHealth score)[26]. Hence, we explored whether the MetaboHealth score associates with differences between LRC groups as defined in the first approach of the analysis above (LRC_g1-3 compared to LRC_g4) and conducted a mixed model linear regression analysis.

We observed that the IPs with an LRC score ≥0.60 (LRC_g1 IPs) had a 0.098 (95%CI: [-0.184] – [-0.012]) lower MetaboHealth score than the partners who had an LRC score of 0 (LRC_g4 IPs; Figure 4 and Supplementary Table 3). The LRC_g2 and LRC_g3 IPs had a 0.032 (95%CI: [-0.077 - [0.012]) and a 0.016 (95%CI: [-0.091] - [0.058]) lower score than the LRC_g4 IPs respectively. Though the effects are relatively small (Figure 4), Indeed we observed that the LLS IPs with ≥60% long-lived ancestors who show delayed onset of disease, also have a healthier circulating metabolic profile in mid-life than the partners with an LRC score of 0.

Discussion

Human longevity is heritable and clusters in specific families. Members of these families live longer and seem to age healthier than the general population. Studying these long-lived families is important to improve our understanding of the molecular and environmental mechanisms that protect from (multi)morbidity and promote a healthy survival up to high ages. In this study we investigated the development of diseases from mid-life onwards in big multigenerational and prospective data, covering up to 25 years of follow-up, in family based (LLS, Netherlands) and population based (SEDD, Sweden) data. We showed that members of long-lived families have a delayed onset of disease, multimorbidity and medication use as compared to their partners, thereby extending their healthspan with up to a decade. These members also postponed multimorbidity since those who were already diagnosed with an age-related disease had a 54% lower risk of having a second age-related disease compared to their partners. When defining familial longevity quantitatively using the LRC score, we demonstrated that an increasing number of long-lived ancestors associates with an increasing delay in disease onset and lower medication use. Finally we demonstrated that at the moment of LLS study inclusion, those with an LRC score ≥0.60 (LRC_g1)) had a better MetaboHealth score than their partners with an LRC score of 0 (LRC_g4), indicating better immune and metabolic health, and lower 5-years mortality risk. We conclude that an increasing number of long-lived ancestors, as measured with the LRC score, is a quantitative indicator of familial longevity, capturing delayed mortality, protection against multimorbidity, and improved health in selected families as well as the population at large. The LRC score can thus potentially be used in genetic studies to elucidate multi-morbidity limiting mechanisms that promote healthspan already in mid-life.

277

278 Our analyses, using ancestral mortality data, in the selected Dutch longevity families and the
279 Swedish register data led to remarkably similar conclusions. An increasing number of long-
280 lived ancestors, as measured with the LRC score, not only associates with a lower mortality
281 at any moment in life[22,23], it also associates, in a similar way, with a lower disease incidence
282 during mid and later life (60 to 75 years): With every 10% increase in LRC score the yearly
283 risk to develop an age-related disease decreased with 39% in the LLS, maximizing to 46% in
284 the SEDD.

285

286 We did observe stronger effects in the SEDD data than in the LLS data, with consistently
287 lower hazard ratio's (HRs) for age-related and metabolic disease incidence. This may be
288 explained firstly because LLS IPs are compared with their partners as controls, either as
289 separate or combined groups in the LRC analyses. IPs and partners share the same adult
290 household and thus, the LLS design controls for shared resources and behavior (such as
291 socio-economic status, social network, and lifestyle). In the SEDD data we did not compare
292 IPs with their partners. The effect size difference between LLS and SEDD may therefore
293 represent the influence of shared resources and behavior. Secondly, in the LLS disease
294 diagnoses were obtained from the general practitioners (GPs) whereas in the SEDD, disease
295 diagnoses were obtained from hospital records available in the Swedish national register
296 data. It may be that stronger effects were observed in the SEDD because hospitalization is on
297 average an indication of more extreme health events than receiving a GP diagnosis.
298 Nevertheless, for many of the GP diagnosed diseases, such as a myocardial infarction,
299 hospitalization is also required.

300

We did not observe statistically significant results for malignancies in the LLS data whereas we did observe significant effects in the SEDD data. However, within the SEDD data the effects for malignant disease incidence were considerably smaller (HR closer to reference group) than for metabolic and age-related diseases. A first explanation for this observation relates to differences in study population and follow-up time. LLS IPs and partners were followed-up for a maximum of 16 years from the average age of 59 years whereas for the SEDD IPs this was a maximum of 25 years from the average age of 52 years. As a result, we may have missed early onset (around 50 years) malignancies in the LLS whereas those are included in the SEDD. This may also explain why a considerably lower proportion of malignant diseases was available in LLS than in the SEDD whereas this was not the case for metabolic diseases. Secondly, inherited genetic factors have a limited effect on many types of malignancies, with heritability estimates ranging between 20% and 30%[27]. However, the chronic diseases, as measured in our study, are much more heritable[28–37], with over 70% heritability for type 2 diabetes[38,39]. As the LRC score captures additive genetic effects[22,23], the lower heritability of malignancies could explain why the small number of malignant disease observations in the LLS did not provide enough power to detect effects and why the effect sizes are lower in the SEDD compared to metabolic diseases. Previous research focusing on malignant diseases in long-lived individuals and their offspring obtained very heterogenous results[4,5,7,8,10,11,13–15,17,40–44] which may also be due to study population and selection differences.

Past research primarily focused on studying disease prevalence of long-lived individuals, such as centenarians, and their children[4,10–14,17–20] in cross sectional designs. Our data covers up to 25 years of follow-up and provides a unique combination between ancestral mortality

information and deep phenotyping of chronic age-related diseases, medication use, as well as metabolomics. This allowed us to closely link familial longevity to medication use and the incidence of multiple diseases. Detailed information about disease incidence was provided by the treating physicians (General Practitioners, GPs) of the LLS IPs and their partners. In the SEDD, we used hospitalization records from the Swedish national registers (see methods for more details). The combination between these two types of data ensured robustness against healthy participant bias. Apart from this, the LLS was initiated with the inclusion of (1) LLS IPs who had at least one long-lived parent and aunt or uncle and (2) the partners of the LLS IPs. LLS IPs are likely to share social (e.g. social network, socio-economic status) and behavioral (e.g. lifestyle, drinking, sporting) traits important for healthy aging and longevity, for example because they share the same household or due to assortative mating[45]. The LLS study design thus corrects for such similarities between LLS IPs and partners, potentially resulting in an underestimation of differences between LLS IPs and partners as compared to general population controls. However, replication in the SEDD, which does not contain any initial inclusion criteria, guarantees results which are not influenced by partner similarities.

Genetic longevity studies so far mainly focused on survival to exceptional ages. Using the LRC score, disease-free survival, possibly in combination with the MetaboHeath score, may be explored as a broader phenotype to increase the power of longevity genetic studies. In addition, the association between LRC score and delayed disease incidence may be explained by the presence of variants protecting from development of disease and/or the absence of disease loci in the long-lived families. Though, previous research showed no evidence that long-lived persons were characterized by the absence of disease loci[46], GWAS studies identifying disease susceptibility variants for example, for hypertension[47],

Alzheimer's disease[48] has progressed significantly. Hence, it is interesting re-investigate if the absence of disease susceptibility loci associates to the LRC score. As mentioned in the previous paragraphs, the larger effect sizes in the SEDD likely illustrate the importance of shared resources and behavior in long-lived families. Further evidence for the clustering of socio-behavioral traits was provided in a recent study which showed that members of long-lived families were less frequently hospitalized with smoking-related cancers as a first disease[8]. As socio-behavioral traits are influenced by complex combinations between genes and environment, further investigation may aid genetic research while providing an interesting basis to investigate the social complexity underlying familial longevity.

Our results provide strong evidence that an increasing number of long-lived ancestors associates with up to a decade of healthspan extension and a healthy metabolomic profile in mid-life. Our results have two important implications. First, future genetic research aimed at identifying protective longevity mechanisms beneficially influencing the risk of multimorbidity could focus on a broader definition of longevity entailing survival to exceptional ages as well as disease free survival and possibly the MetaboHealth score metabolites. Second, our results highlight the importance of integrating multiple generations of ancestral mortality data to existing and novel studies. In the past it was difficult to obtain such ancestral information but currently it is much more feasible to do so, as population scale family tree data is becoming increasingly available[22,49,50]. Moreover, in an increasing number of countries ancestral data can be retrieved from the national statistics bureaus, such as Statistics Sweden or Statistics Netherlands. Finally, next to genetic drivers of longevity and disease incidence, it is important to investigate if and how potential socio-behavioral resources[51], such as socio-economic status and stress, associate to both longevity

373 and disease incidence. If these novel insights are consistently applied across studies, the
374 comparative nature of longevity studies may improve and facilitate the discovery of novel
375 genetic variants and mechanisms promoting healthy ageing.

Methods

Leiden Longevity Study

The Leiden Longevity Study (LLS) was initiated in 2002 to study the mechanisms that lead to exceptional survival. The LLS currently consist of 650 three-generational families, defined by siblings who have the same parents (Figure 1). Inclusion took place between 2002 and 2006 and initially started with the recruitment of living sibling pairs. Within a sibling pair, males were invited to participate if they were 89 years or older and females if they were 91 years or older. Inclusion was subsequently extended to the children of the sibling pairs and their partners. This study focuses on the children of the sibling pairs and their partners, referring to them as LLS IPs and partners. From their perspective, IPs were included if they had at least one long-lived parent and aunt or uncle (females ≥ 91 years and males ≥ 89 years). In total, 1,674 Index Persons (IPs, F3), 745 partners (F3), 1,295 parents (F2), 2,370 aunts and uncles (F2), 760 grandparents (F1), and 1,237 parents of the partners (F2) were included in this study.

Mortality information was verified by birth or marriage certificates and passports whenever possible. Additionally, verification took place via personal cards which were obtained from the Dutch Central Bureau of Genealogy. In January 2021 all mortality information was updated through the Personal Records Database (PRD) which is managed by Dutch governmental service for identity information. https://www.government.nl/topics/personal-data/personal-records-database-brp. The combination of officially documented information provides very reliable and complete ancestral as well as current mortality information.

399

400 Disease data has been retrieved from the General Practitioners (GPs) of the LLS IPs and
401 partners and covers the period from birth until 2018. GPs extracted the presence of chronic
402 age-related diseases as specified in Supplementary Table 4 and the year the disease
403 occurred from their electronic health records. The GP records are kept up to date when a
404 person switches from one GP practice to another. Diseases were clustered into 3 groups
405 based on the International Statistical Classification of Diseases and Related Health Problems
406 (ICD-10) codes, (1) metabolic diseases, (2) malignant diseases, and (3) age related diseases,
407 which are the combination of metabolic and malignant diseases. Furthermore, cross-
408 sectional information on medication use from pharmacies was obtained for the period 2006-
409 2008, indicating whether a specific medicine was used or not. Medication was grouped
410 according to the international Anatomical Therapeutic Chemical Classification System (ATC)
411 standard. We focused on ATC-A (alimentary tract and metabolism), ATC-B (blood and blood
412 forming organs), and ATC-C (cardiovascular system) type medications because they match
413 the disease groups we investigate.

414

415 Ethylenediamine tetraacetic acid (EDTA) plasma samples were obtained for all LLS IPs and
416 partners at inclusion. From these samples, metabolomics biomarker data was quantified
417 using high-throughput nuclear magnetic resonance (NMR) spectroscopy provided by the
418 Nightingale Health platform. Experimentation and application details of the Nightingale NMR
419 platform has previously been described[52,53]. Moreover, the metabolic biomarkers measured
420 using the nightingale platform were used in a variety of publications (overview can be found
421 here: https://nightingalehealth. com/publications).

422

423 **Scanian Economic-Demographic Database**

424 The Scanian Economic-Demographic Database (SEDD) is a longitudinal database covering five
425 rural Scanian parishes and the city of Landskrona. It spans the period 1812-1967, with full
426 coverage of the villages from 1812 and for Landskrona from 1904. The SEDD database was
427 constructed using register-type data from catechetical examination registers and was
428 updated with information on births, marriages, and deaths from church books. Unique
429 person numbers were introduced in Sweden by 1947. Through these person numbers
430 individuals can be followed in the national Swedish registration, introduced in 1967. Persons
431 who out-migrated from the research region before the introduction of the person number
432 were linked to the 1950 Census and the Swedish Death Index. The obtained person numbers
433 were subsequently used to track individuals in the Swedish national register for the period
434 1968-2015. The link to the Swedish Death Index yielded ancestral death dates anywhere in
435 Sweden even for individuals who out-migrated from the research region before the person
436 number or nationwide register data were introduced. At present (2022), the SEDD database
437 contains 920,159 unique individuals.
438
439 Index person (IP) identification for this study happened in subsequent steps (Supplementary
440 table 5). First, from the entire SEDD data we identified all persons (from here: IPs) who were
441 part of the national register data in the years 1990-1995 and between ages 45-60, and
442 followed them in the national registers for the period 1990-2015. Second, IPs were selected
443 to have known grandparents on at least one side of the family (maternal or paternal), and
444 whose parents were from an extinct birth cohort (born before 1915) to ensure complete
445 information about their date of death. Third, we included lifespan information of their
446 parents, aunts and uncles, and their grandparents. Fourth, IPs who were found in the

hospital records in the year preceding their eligibility for the study (1989-1994) were excluded to minimize the number of IPs with existing conditions receiving hospital treatments. Lastly, partners of IPs were excluded to ensure mutually exclusive ancestral information. In total, 1,493 Index Persons (IPs, F3), 2,969 parents (F2), 5,830 aunts and uncles (F2), and 3,028 grandparents (F1) were included in this study.

The Swedish hospital registers reached nationwide coverage in 1987 and records are considered complete from 1989. The main diagnosis for each hospitalization has been recorded in ICD-9 coding from 1987-1997 and ICD-10 coding 1997-2015. We recoded ICD-9 diagnoses to ICD-10 using the official crosswalk provided by Statistics Sweden. Diseases are specified identical to the LLS (Supplementary Table 4) to ensure comparability between the databases. It is relevant for our analyses to mention that only 214 IPs (8.6%) die without ever receiving a hospital diagnosis as a higher percentage would have warranted a competing risk analysis (see statistics section for more details).

Lifetables

In the Netherlands and Sweden, population based cohort lifetables are available from 1850 and 1800 respectively, until 2021[54,55]. 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 LLS and SEDD. This approach prevents against the effects of secular mortality trends over the last centuries and enables comparisons across study populations[56,57]. In SEDD, we focused only on extinct birth cohorts and death ancestors. However, In the LLS some ancestors (only aunts/uncles) were still alive (right censoring). To deal with non-extinct birth cohorts, we used the prognostic lifetables provided by Statistics Netherlands[54,55] and to deal with right censoring we used single imputation where we estimated an age of death based on the remaining life expectancy at the age of censoring.

Scores

The Longevity Relatives Count (LRC) score was used in LLS and SEDD to map the offspring's family history of longevity. The LRC score indicates the proportion of ancestors that became long-lived, weighted by the genetic distance between IPs (and partners in LLS) and their ancestors. For example, an LRC of 0.5 indicates 50% long-lived ancestors. For this study, two generations of ancestors were available to calculate the LRC score for IPs and one generation for the partners of the LLS IPs (Figure 1 and Supplementary Figure 1). In the LLS, the LRC score was calculated using the mortality information as updated in 2021. In the SEDD, IPs were identified in such a way that all ancestors were deceased at the start of follow-up. The LRC score has been described in detail by van den Berg et al, 2020 in Aging Cell[23]. Supplementary Figure 6A-B depict the LRC score distribution in the LLS and SEDD. Additionally, in de LLS study, the MetaboHealth score was used as an indicator for the (metabolomic) health of the offspring and partners at study inclusion. This previously published score was generated by based on NMR metabolomics data in ~40.000 European

study participants and provides a weighted summary of 14 independent metabolites covering 5-10 years mortality risk and metabolite markers of lipid metabolism, fatty acid metabolism, glycolysis, fluid balance, and inflammation[26]. Supplementary Figure 6C depicts the MetaboHealth score distribution in the LLS.

Statistical analyses

Statistical analyses were conducted using R version 4.0.2[58]. We reported 95% confidence intervals (CIs) and considered p-values statistically significant at the 5% level (α=0.05). A list of used R-packages and version numbers will be made available on gitlab (see code availability statement). In all random effect and frailty models we consider the F3 IPs who share the same parents as a family. Random effect and frailty models were used to adjust for within-family relations of the F3 IPs

Logistic and linear mixed model. To compare disease and medication prevalence between (LRC-based) LLS IPs and partners we fitted a logistic mixed model (1) and to compare the MetaboHealth score between the LRC groups (LRC_g1-3 with LRC_g4) in the LLS we fitted a linear mixed model (2):

$$\text{logit}\,(\boldsymbol{\pi}_{ij}) = \boldsymbol{\beta Z}_{ij} + \boldsymbol{\gamma X}_{ij} + \boldsymbol{u}_i \qquad (1)$$

$$\boldsymbol{Y}_{ij} = \boldsymbol{\beta Z}_{ij} + \boldsymbol{\gamma X}_{ij} + \boldsymbol{u}_i \qquad (2)$$

Where $\boldsymbol{Y}_{ij}$ is a vector of responses for person j in family i. and $\pi_{ij} = P(Y_{ij} = 1|\boldsymbol{Z}_{ij}, \boldsymbol{X}_{ij}, u_i)$ when considering logistic regression. $\boldsymbol{\beta}$ is a vector of regression coefficients for the main effects of interest ($\boldsymbol{Z}$). $\boldsymbol{\gamma}$ is a vector of regression coefficients for the effects of possible

519 confounders ($\boldsymbol{X}$). $\boldsymbol{u}$ is a vector of unobserved random effects shared by each member of the
520 same family i and was assumed to follow a normal distribution. All analyses performed using
521 logistic and linear mixed models have been adjusted for sex and age at study inclusion. In
522 addition, the MetaboHealth score analyses have been adjusted for medication use.
523
524 **Survival analysis (Cox-type random effects regression model).** To compare prospective
525 disease incidences between (LRC-based) offspring and partners we fitted three different
526 Cox-type random effects models:
527

528 $$\lambda(t_{ij}) = u_i \lambda_0(t_{ij}) \exp(\boldsymbol{\beta Z}_{ij} + \boldsymbol{\gamma X}_{ij}) \qquad (3)$$

529
530 In the first type of models, we model t_{ij}=age at first disease onset; in the second type of
531 models the outcome of interest is t_{ij}=age at the second disease onset (multi-morbidity
532 onset). In both cases, the models are adjusted for left truncation given by the age at entry in
533 the study. In the third type of models, we consider the time between the first and the
534 second disease onset; i.e. t_{ij}=age at second disease onset where age at the first disease onset
535 is considered as the left-truncation time in this analysis. $\lambda_0(t_{ij})$ refers to the baseline
536 hazard, which is left unspecified in a Cox-type model. $\boldsymbol{\beta}$ is the vector of regression
537 coefficients for the main effects of interest ($\boldsymbol{Z}$). $\boldsymbol{\gamma}$ is a vector of regression coefficients for
538 the effects of possible confounders ($\boldsymbol{X}$). $u_i > 0$ refers to an unobserved random effect
539 (frailty) shared by the members of the same family i and was assumed to follow a gamma
540 distribution. All survival models were adjusted for sex. Additionally, the third type of analysis
541 focusing on the time from first to second disease (Table 3) has been further adjusted for age

542 at study inclusion as we did not limit our sample to persons without any diseases at the start
543 of follow-up.
544
545 **Competing interests**
546 The authors declare no competing interests.
547
548 **Ethical regulations**
549 Leiden Longevity Study: In accordance with the Declaration of Helsinki, we obtained
550 informed consent from all participants prior to their entering the study. Good clinical
551 practice guidelines were maintained. The study protocol was approved by the ethical
552 committee of the Leiden University Medical Center before the start of the study (P01.113).
553
554 SEDD: The SEDD has approval for research from Regionala etikprövningsnämnden, Lund,
555 (dnr 161/2006, dnr 627/2010), and instructions from Datainspektionen, Stockholm (dnr
556 1999-2005).
557
558 **Author contributions**
559 Niels van den Berg is the study investigator and was responsible for initiating the study, data
560 management, data analyses, writing the manuscript and finalizing it. Ingrid van Dijk was
561 responsible for the data organization and analyses of the Swedish data. Mar Rodriguez-
562 Girondo provided overall support on statistical analyses. P. Eline Slagboom and Marian
563 Beekman provided overall coordination and supervision.
564
565 **Code availability**

566 The scripts containing the code for data pre-processing and data analyses can be freely
567 downloaded at: https://git.lumc.nl/publications/longevity-family-diseases
568
569 **Data availability**
570 The individual-level data from the SEDD, the Statistics Sweden, and LLS are protected by
571 Swedish and Dutch personal integrity laws, and other (privacy) regulations. As such,
572 restrictions apply to the availability of the LLS and SEDD data, which were used under license
573 for the current study, and so are not publicly available. For both datasets, summary statistics
574 are available upon request to the corresponding author (Niels van den Berg). The LLS data is
575 available for replication purposes upon reasonable request to P. Eline Slagboom and if
576 replication is conducted within the secure LUMC network environment. Researchers can
577 gain access to the SEDD data as used in this study if relevant permissions have been
578 obtained in accordance with the restrictions stated by the Regional Ethical Review Board,
579 the Swedish Data Inspection Board and Lund University.
580
581 **Acknowledgments**
582 The construction and maintenance of the LLS data has received funding from the European
583 Union's Seventh Framework Programme (FP7/2007-2011) under grant agreement number
584 259679. This study was further supported by the Innovation-Oriented Research Program on
585 Genomics (SenterNovem IGE05007), the Centre for Medical Systems Biology and the
586 Netherlands Consortium for Healthy Ageing (grant 050-060-810), all in the framework of the
587 Netherlands Genomics Initiative, Netherlands Organization for Scientific Research (NWO), by
588 BBMRI-NL, a Research Infrastructure financed by the Dutch government (NWO 184.021.007
589 and 184.033.111). Ingrid van Dijk was funded through the research program "Landskrona

590 Population Study" and Niels van den Berg and Ingrid van Dijk were funded through the
591 research project "An Age Old Advantage?" (P21-0139), the Swedish Foundation for the
592 Humanities and Social Sciences (Riksbankens Jubileumfond, RJ). Niels van den Berg was
593 further funded by the Netherlands Organization for Scientific Research, domain Health
594 Research and Medical Sciences (09120012010052). We thank Bianca Schutte, Rinske van
595 Reijen, and Yotam Raz for digitizing and organizing the LLS medication and disease data.

596

References

1. Oeppen, J. & Vaupel, J. W. Broken Limits to Life Expectancy. *Science (80-.).* **296**, 1029–1031 (2002).
2. Wang, H. *et al.* Global age-sex-specific fertility, mortality, healthy life expectancy (HALE), and population estimates in 204 countries and territories, 1950–2019: a comprehensive demographic analysis for the Global Burden of Disease Study 2019. *Lancet* **396**, 1160–1203 (2020).
3. Barnett, K. *et al.* Epidemiology of multimorbidity and implications for health care, research, and medical education: a cross-sectional study. *Lancet* **380**, 37–43 (2012).
4. Andersen, S. L., Sebastiani, P., Dworkis, D. a., Feldman, L. & Perls, T. T. Health Span Approximates Life Span Among Many Supercentenarians: Compression of Morbidity at the Approximate Limit of Life Span. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **67A**, 395–405 (2012).
5. Ash, A. S. *et al.* Are Members of Long-Lived Families Healthier Than Their Equally Long-Lived Peers? Evidence From the Long Life Family Study. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **70**, 971–976 (2015).
6. Lipton, R. B. *et al.* Exceptional Parental Longevity Associated with Lower Risk of Alzheimer's Disease and Memory Decline. *J. Am. Geriatr. Soc.* **58**, 1043–1049 (2010).
7. Galvin, A., Ukraintseva, S., Arbeev, K., Feitosa, M. & Christensen, K. Physical robustness and resilience among long-lived female siblings: A comparison with sporadic long-livers. *Aging (Albany. NY).* **12**, 15157–15168 (2020).
8. Christensen, K. *et al.* Mechanisms underlying familial aggregation of exceptional

health and survival: A three-generation cohort study. *Aging Cell* **19**, 1–11 (2020).

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

10. Ailshire, J. A., Beltran-Sanchez, H. & Crimmins, E. M. Becoming Centenarians: Disease and Functioning Trajectories of Older U.S. Adults as They Survive to 100. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **70**, 193–201 (2015).

11. Adams, E. R., Nolan, V. G., Andersen, S. L., Perls, T. T. & Terry, D. F. Centenarian Offspring: Start Healthier and Stay Healthier. *J. Am. Geriatr. Soc.* **56**, 2089–2092 (2008).

12. Gellert, P. *et al.* Centenarians Differ in Their Comorbidity Trends During The 6 Years Before Death Compared to Individuals Who Died in Their 80s or 90s. *Journals Gerontol. Ser. A* **73**, 1357–1362 (2018).

13. Ismail, K. *et al.* Compression of Morbidity Is Observed Across Cohorts with Exceptional Longevity. *J. Am. Geriatr. Soc.* **64**, 1583–1591 (2016).

14. Gueresi, P. *et al.* Does the longevity of one or both parents influence the health status of their offspring? *Exp. Gerontol.* **48**, 395–400 (2013).

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

16. Gubbi, S. *et al.* Effect of Exceptional Parental Longevity and Lifestyle Factors on Prevalence of Cardiovascular Disease in Offspring. *Am. J. Cardiol.* **120**, 2170–2175 (2017).

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

644 18. Zeng, Y. *et al.* Health consequences of familial longevity influence among the Chinese
645 elderly. *Journals Gerontol. - Ser. A Biol. Sci. Med. Sci.* **68**, 473–482 (2013).
646 19. Richmond, R. L., Law, J. & KayLambkin, F. Morbidity profiles and lifetime health of
647 Australian centenarians. *Australas. J. Ageing* **31**, 227–232 (2012).
648 20. Christensen, K., McGue, M., Petersen, I., Jeune, B. & Vaupel, J. W. Exceptional
649 longevity does not result in excessive levels of disability. *Proc. Natl. Acad. Sci.* **105**,
650 13274–13279 (2008).
651 21. van den Berg, N. Family matters in unraveling human longevity. *Aging (Albany. NY).*
652 **12**, 1–2 (2020).
653 22. van den Berg, N. *et al.* Longevity defined as top 10% survivors and beyond is
654 transmitted as a quantitative genetic trait. *Nat. Commun.* **10**, 35 (2019).
655 23. Berg, N. *et al.* Longevity Relatives Count score identifies heritable longevity carriers
656 and suggests case improvement in genetic studies. *Aging Cell* **19**, 1–15 (2020).
657 24. van den Berg, N. *et al.* Longevity Around the Turn of the 20th Century: Life-Long
658 Sustained Survival Advantage for Parents of Today's Nonagenarians. *Journals*
659 *Gerontol. Ser. A* **73**, 1295–1302 (2018).
660 25. Mourits, R. J. *et al.* Intergenerational transmission of longevity is not affected by other
661 familial factors: evidence from 16,905 Dutch families from Zeeland, 1812-1962. *Hist.*
662 *Fam.* **25**, 484–526 (2020).
663 26. Deelen, J. *et al.* A metabolic profile of all-cause mortality risk identified in an
664 observational study of 44,168 individuals. *Nat. Commun.* **10**, 3346 (2019).
665 27. Lichtenstein, P. *et al.* Environmental and Heritable Factors in the Causation of Cancer
666 — Analyses of Cohorts of Twins from Sweden, Denmark, and Finland. *N. Engl. J. Med.*
667 **343**, 78–85 (2000).

668 28. Prins, B. P., Lagou, V., Asselbergs, F. W., Snieder, H. & Fu, J. Genetics of coronary
669 artery disease: Genome-wide association studies and beyond. *Atherosclerosis* **225**, 1–
670 10 (2012).
671 29. FEINLEIB, M. *et al.* The NHLBI Twin Study of Cardiovascular Disease Risk Factors:
672 Methodology and Summary of Results. *Am. J. Epidemiol.* **106**, 284–295 (1977).
673 30. Wienke, A., Holm, N. V., Skytthe, A. & Yashin, A. I. The Heritability of Mortality Due to
674 Heart Diseases: A Correlated Frailty Model Applied to Danish Twins. *Twin Res.* **4**, 266–
675 274 (2001).
676 31. Doris, P. A. The Genetics of Blood Pressure and Hypertension: The Role of Rare
677 Variation. *Cardiovasc. Ther.* **29**, 37–45 (2011).
678 32. Zdravkovic, S. *et al.* Heritability of death from coronary heart disease: a 36-year
679 follow-up of 20 966 Swedish twins. *J. Intern. Med.* **252**, 247–254 (2002).
680 33. Drobni, Z. D. *et al.* Heritability of Coronary Artery Disease: Insights From a Classical
681 Twin Study. *Circ. Cardiovasc. Imaging* **15**, 133–141 (2022).
682 34. Levy, D. *et al.* Genome-wide association study of blood pressure and hypertension.
683 *Nat. Genet.* **41**, 677–687 (2009).
684 35. Khera, A. V. & Kathiresan, S. Genetics of coronary artery disease: discovery, biology
685 and clinical translation. *Nat. Rev. Genet.* **18**, 331–344 (2017).
686 36. Bevan, S. *et al.* Genetic Heritability of Ischemic Stroke and the Contribution of
687 Previously Reported Candidate Gene and Genomewide Associations. *Stroke* **43**, 3161–
688 3167 (2012).
689 37. Muñoz, M. *et al.* Evaluating the contribution of genetics and familial shared
690 environment to common disease using the UK Biobank. *Nat. Genet.* **48**, 980–983
691 (2016).

692 38. Willemsen, G. *et al.* The Concordance and Heritability of Type 2 Diabetes in 34,166
693 Twin Pairs From International Twin Registers: The Discordant Twin (DISCOTWIN)
694 Consortium. *Twin Res. Hum. Genet.* **18**, 762–771 (2015).
695 39. Fuchsberger, C. *et al.* The genetic architecture of type 2 diabetes. *Nature* **536**, 41–47
696 (2016).
697 40. Tindale, L. C., Salema, D. & Brooks-Wilson, A. R. 10-year follow-up of the Super-
698 Seniors Study: compression of morbidity and genetic factors. *BMC Geriatr.* **19**, 58
699 (2019).
700 41. Terry, D. F., Wilcox, M., McCormick, M. A., Lawler, E. & Perls, T. T. Cardiovascular
701 Advantages Among the Offspring of Centenarians. *Journals Gerontol. Ser. A Biol. Sci.*
702 *Med. Sci.* **58**, M425–M431 (2003).
703 42. Terry, D. F., Wilcox, M. A., McCormick, M. A. & Perls, T. T. Cardiovascular Disease
704 Delay in Centenarian Offspring. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **59**, M385–
705 M389 (2004).
706 43. Terry, D. F. *et al.* Lower All-Cause, Cardiovascular, and Cancer Mortality in
707 Centenarians' Offspring. *J. Am. Geriatr. Soc.* **52**, 2074–2076 (2004).
708 44. Sebastiani, P. *et al.* Families Enriched for Exceptional Longevity also have Increased
709 Health-Span: Findings from the Long Life Family Study. *Front. Public Heal.* **1**, 1–9
710 (2013).
711 45. Pedersen, J. K. *et al.* The Survival of Spouses Marrying Into Longevity-Enriched
712 Families. *Journals Gerontol. Ser. A Biol. Sci. Med. Sci.* **72**, 109–114 (2017).
713 46. Beekman, M. *et al.* Genome-wide association study (GWAS)-identified disease risk
714 alleles do not compromise human longevity. *Proc. Natl. Acad. Sci.* **107**, 18046–18049
715 (2010).

716 47. Surendran, P. *et al.* Trans-ancestry meta-analyses identify rare and common variants
717 associated with blood pressure and hypertension. *Nat. Genet.* **48**, 1151–1161 (2016).
718 48. Bertram, L. & Tanzi, R. E. Alzheimer disease risk genes: 29 and counting. *Nat. Rev.*
719 *Neurol.* **15**, 191–192 (2019).
720 49. Kaplanis, J. *et al.* Quantitative analysis of population-scale family trees with millions of
721 relatives. *Science* **360**, 171–175 (2018).
722 50. Erlich, Y., Shor, T., Pe'er, I. & Carmi, S. Identity inference of genomic data using long-
723 range familial searches. *Science* **362**, 690–694 (2018).
724 51. Crimmins, E. M. Social hallmarks of aging: Suggestions for geroscience research.
725 *Ageing Research Reviews* (2020). doi:10.1016/j.arr.2020.101136
726 52. Soininen, P. *et al.* High-throughput serum NMR metabonomics for cost-effective
727 holistic studies on systemic metabolism. *Analyst* **134**, 1781 (2009).
728 53. Würtz, P. *et al.* Quantitative Serum Nuclear Magnetic Resonance Metabolomics in
729 Large-Scale Epidemiology: A Primer on -Omic Technologies. *Am. J. Epidemiol.* **186**,
730 1084–1096 (2017).
731 54. Van Der Meulen, A. *Life tables and Survival analysis*. (2012).
732 55. Carolina, T., Uijvenhoven, L. & van der Laan, J. Overlevingstafels en longitudinale
733 analyse. *CBS* 1–25 (2009).
734 56. van den Berg, N., Beekman, M., Smith, K. R., Janssens, A. & Slagboom, P. E. Historical
735 demography and longevity genetics: Back to the future. *Ageing Res. Rev.* **38**, 28–39
736 (2017).
737 57. Sebastiani, P., Nussbaum, L., Andersen, S. L., Black, M. J. & Perls, T. T. Increasing
738 Sibling Relative Risk of Survival to Older and Older Ages and the Importance of Precise
739 Definitions of "Aging," "Life Span," and "Longevity". *Journals Gerontol. Ser. A Biol. Sci.*

Med. Sci. **71**, 340–346 (2016).

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

Figure Legends

Figure 1: *Conceptual pedigree of a 3 filial (F) generation LLS family*

This figure corresponds to Table 1 and represents a hypothetical family from the LLS covering 3 filial (F) generations. Circles represent women, Squares represent men. Dark blue: Index persons (IPs, F3), dark green: partners of IPs (F3), light blue: fathers and mothers of IPs (F2), aunts and uncles of IPs (F2), grandmothers and grandfathers of IPs (f1), light green: fathers and mothers of IPs (F2). The dark blue and green colors represent the IPs and their partners who are investigated in this study. The light blue and green colors represent the ancestors of the IPs and partners and were used in this study to calculate the Longevity Relatives Count (LRC) score.

Figure 2: *Disease prevalence and medication use in the LLS*

This figure depicts the odds ratio's (ORs) for disease prevalence (panel A) and medication use (Panel B). Blue bars represent LLS IPs and green bars represent their partners, similar to the colors used in Figure 1. The y-axis of panel A represent the percentage of LLS IPs and partners who had an age-related, metabolic, or malignant disease (x-axis). The y-axis of panel B represent the percentage of LLS IPs and partners who used ATC-A (alimentary tract and metabolism), ATC-B (blood and blood forming organs), or ATC-C (cardiovascular system) type medications (x-axis). CI is the abbreviation for confidence interval and N represents the numbers of the LLS IPs and partners in the specific disease groups. All estimates are adjusted for age at inclusion and sex.

766 **Figure 3:** LLS m*etabolic disease incidence with and without LRC-defined groups*

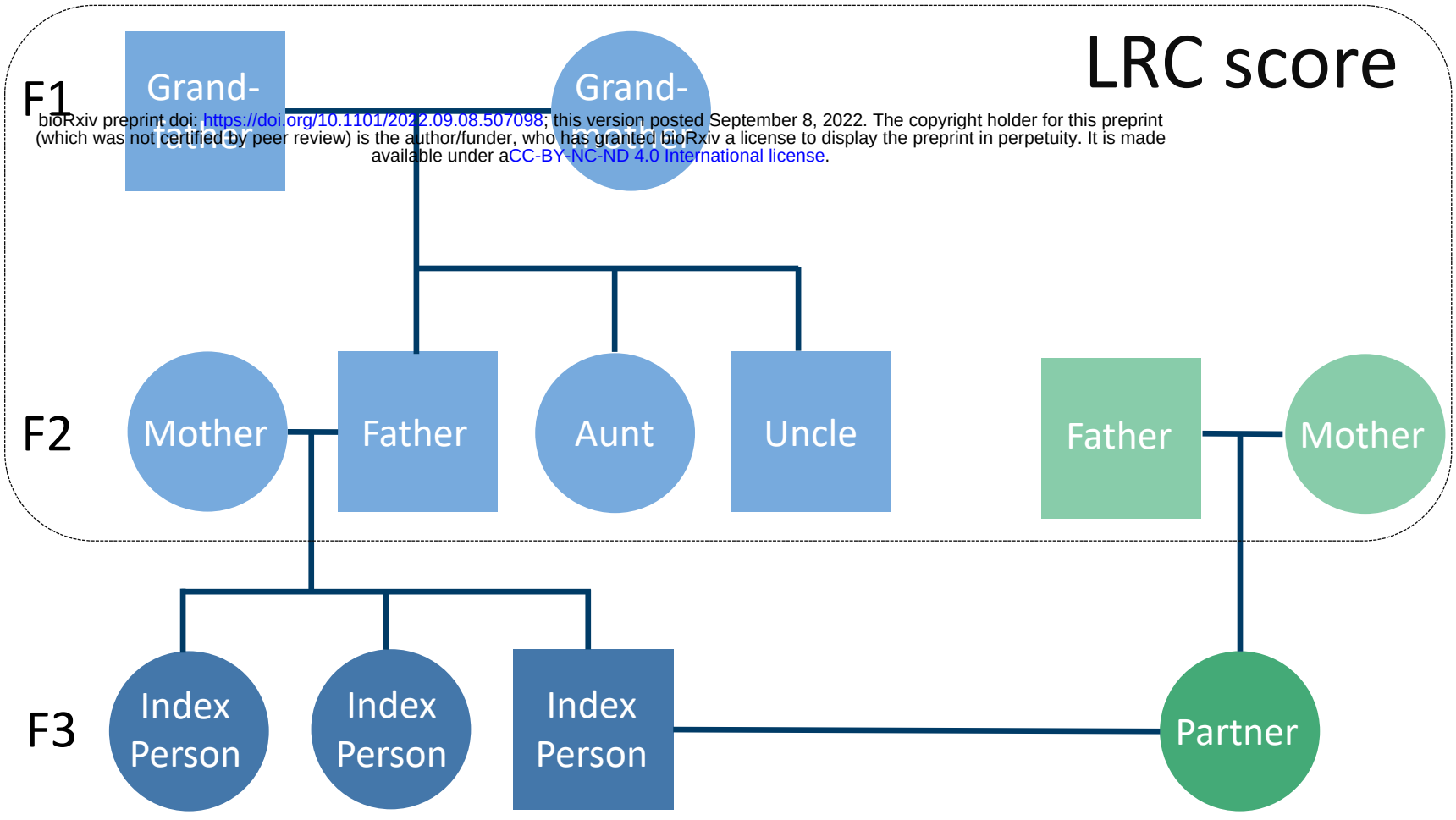
767 This figure depicts survival curves reflecting metabolic disease incidence within the Leiden
768 Longevity Study (LLS). The x-axis show age in years and the y-axis show metabolic disease
769 incidence. Dotted lines represent the age at which 50% of the members of a specific group
770 had their first metabolic disease. Panel A depicts two groups; the blue line represents LLS
771 Index Persons (IPs) and the green line represents the partners. The mean difference
772 between the lines represents the Hazard Ratio (HR) shown in Table 2. Panel B depicts four
773 groups; LRC_g1: IPs with an LRC ≥0.60 (dark blue), LRC_g2: IPs with an LRC [≥0.1 & <0.60]
774 (light blue), LRC_g3: partners with an LRC >0 (light green), and LRC_g4: partners with an LRC
775 =0 (dark green). The mean difference between the LRC_g1-3 and LRC_g4 line represents the
776 HR shown in Table 3. Vertical lines within the colored lines represent right censoring events.
777 The bottom column of panel A and B shows how many persons were still at risk of having a
778 metabolic disease at different ages. Survival curves are adjusted for left truncation and right
779 censoring.

780

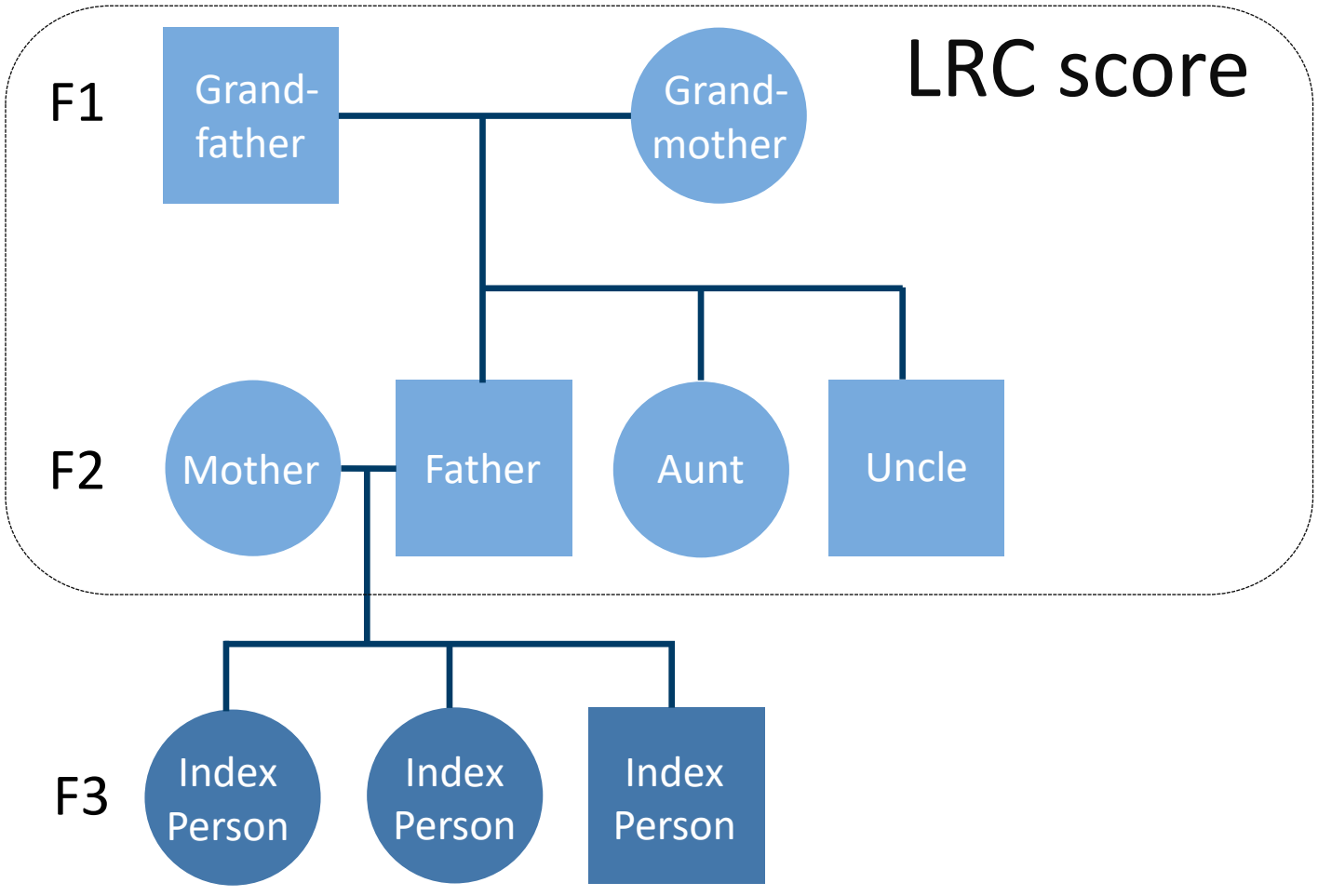
781 **Figure 4:** *MetaboHealth score differences for LRC groups at LLS study inclusion*

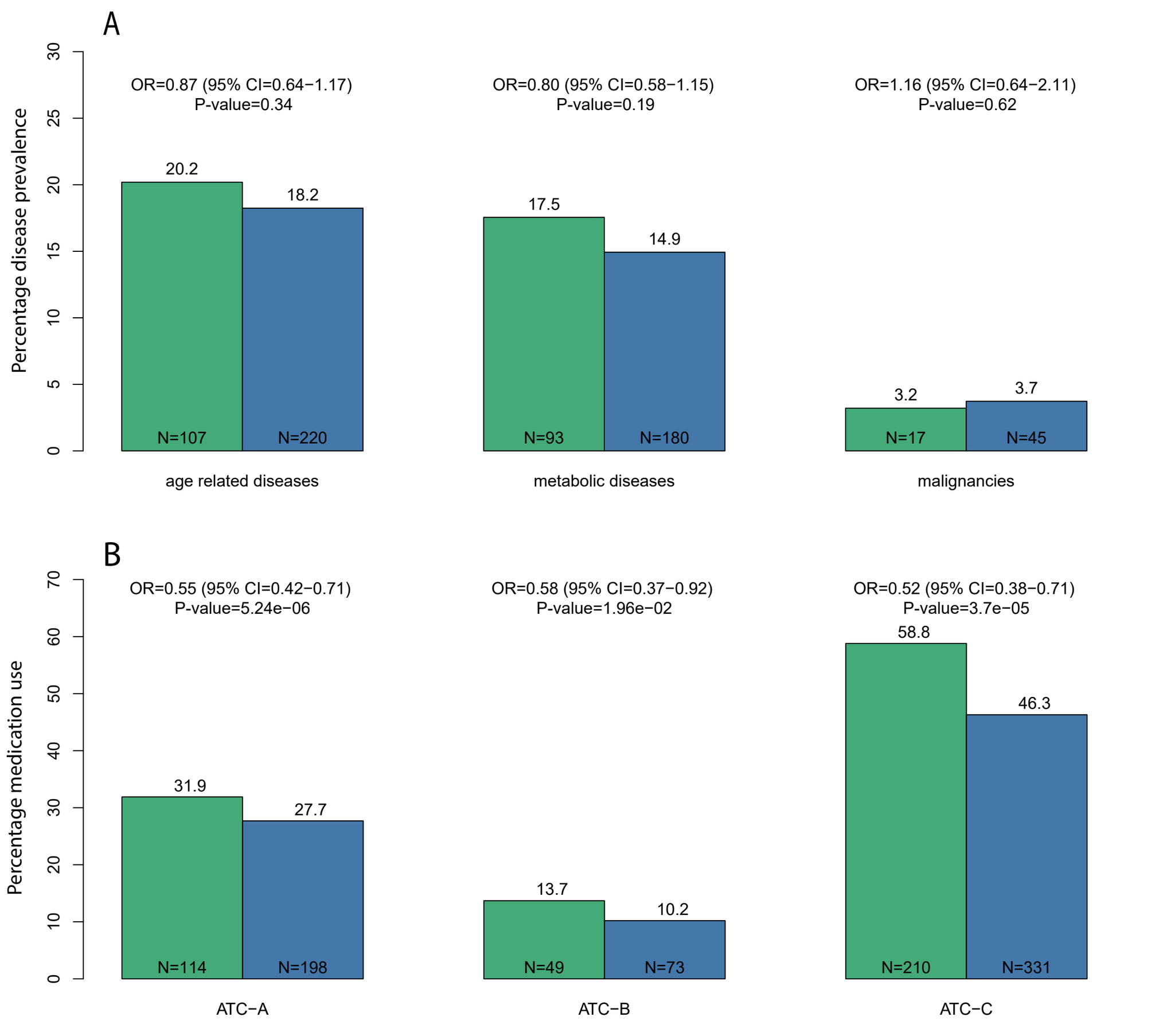
782 The x-axis depict three groups; LRC_g1: IPs with an LRC ≥0.60 (dark blue), LRC_g2: IPs with
783 an LRC [≥0.1 & <0.60] (light blue), and LRC_g3: partners with an LRC >0 (light green). The
784 dotted red line depicts the LRC_g4 group: partners with an LRC =0 (dark green). The y-axis
785 depict the MetaboHealth score. Higher MetaboHealth score values represent a less healthy
786 metabolomic profile as measured by the MetaboHealth score which represents 5/10 year
787 mortality risk (see methods for more details). Error bars represent confidence intervals. CI is
788 the abbreviation for confidence interval.

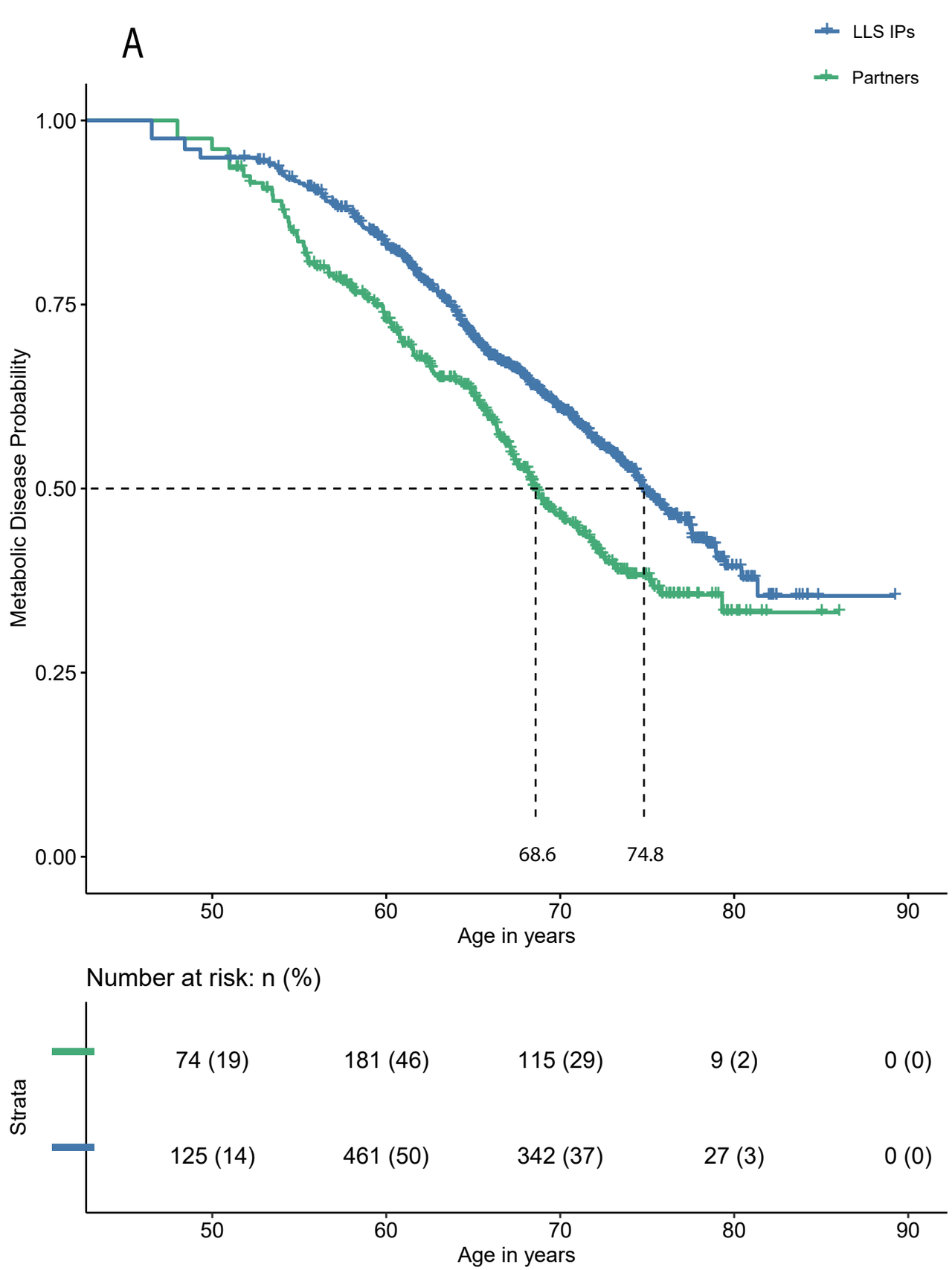
A

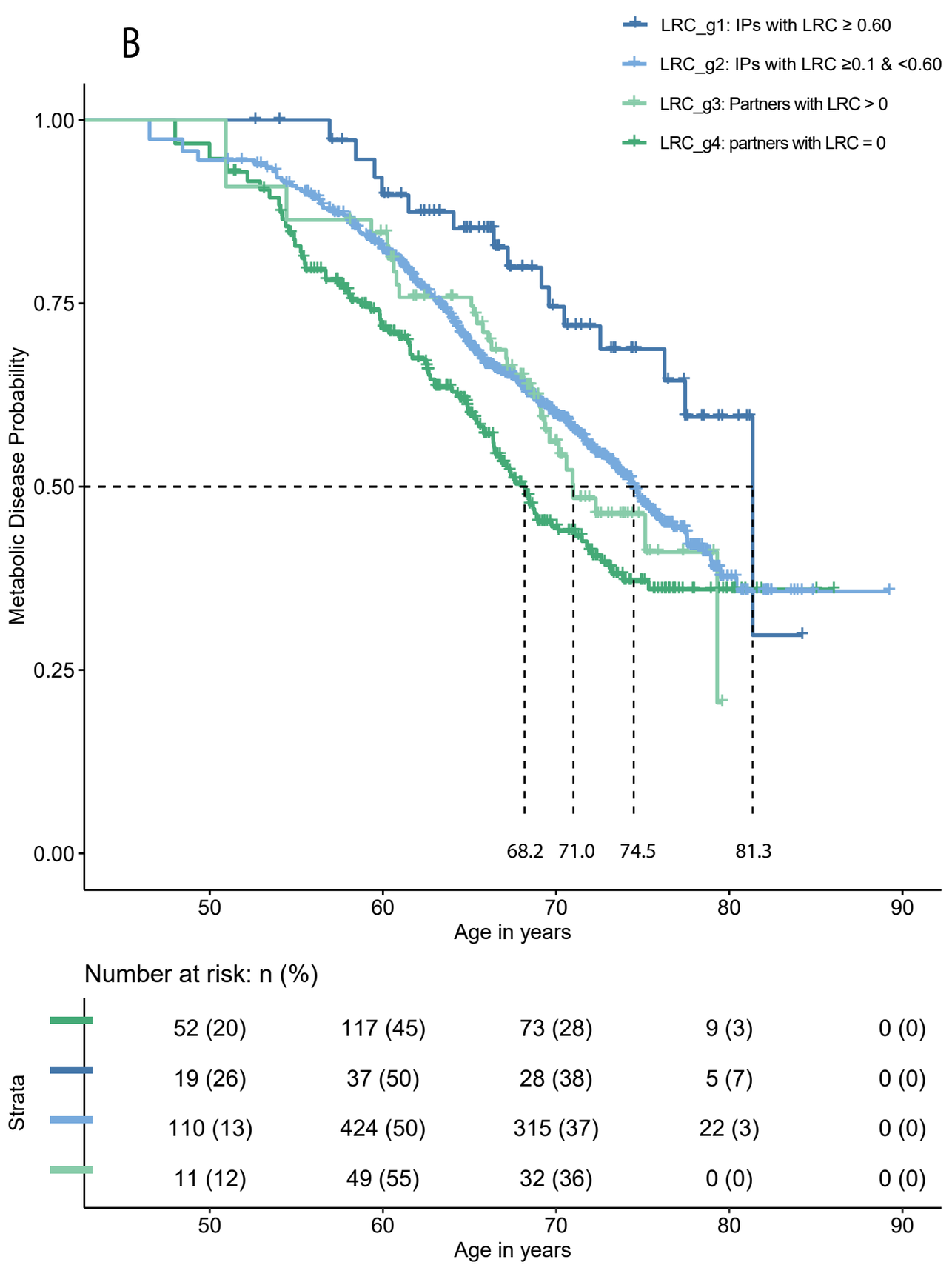


B




A
Percentage disease prevalence
0
5
10
15
20
25
30
OR=0.87 (95% CI=0.64−1.17)
P-value=0.34
20.2
18.2
N=107
N=220
age related diseases
OR=0.80 (95% CI=0.58−1.15)
P-value=0.19
17.5
14.9
N=93
N=180
metabolic diseases
OR=1.16 (95% CI=0.64−2.11)
P-value=0.62
3.2
3.7
N=17
N=45
malignancies
B
Percentage medication use
0
10
20
30
40
50
60
70
OR=0.55 (95% CI=0.42−0.71)
P-value=5.24e−06
31.9
27.7
N=114
N=198
ATC−A
OR=0.58 (95% CI=0.37−0.92)
P-value=1.96e−02
13.7
10.2
N=49
N=73
ATC−B
OR=0.52 (95% CI=0.38−0.71)
P-value=3.7e−05
58.8
46.3
N=210
N=331
ATC−C


A
LLS IPs
Partners
1.00
0.75
0.50
0.25
0.00
Metabolic Disease Probability
68.6
74.8
50
60
70
80
90
Age in years
Number at risk: n (%)
Strata
74 (19)
181 (46)
115 (29)
9 (2)
0 (0)
125 (14)
461 (50)
342 (37)
27 (3)
0 (0)


B
LRC_g1: IPs with LRC ≥ 0.60
LRC_g2: IPs with LRC ≥0.1 & <0.60
LRC_g3: Partners with LRC > 0
LRC_g4: partners with LRC = 0
1.00
0.75
0.50
0.25
0.00
Metabolic Disease Probability
68.2
71.0
74.5
81.3
50
60
70
80
90
Age in years
Number at risk: n (%)
Strata
52 (20)
117 (45)
73 (28)
9 (3)
0 (0)
19 (26)
37 (50)
28 (38)
5 (7)
0 (0)
110 (13)
424 (50)
315 (37)
22 (3)
0 (0)
11 (12)
49 (55)
32 (36)
0 (0)
0 (0)

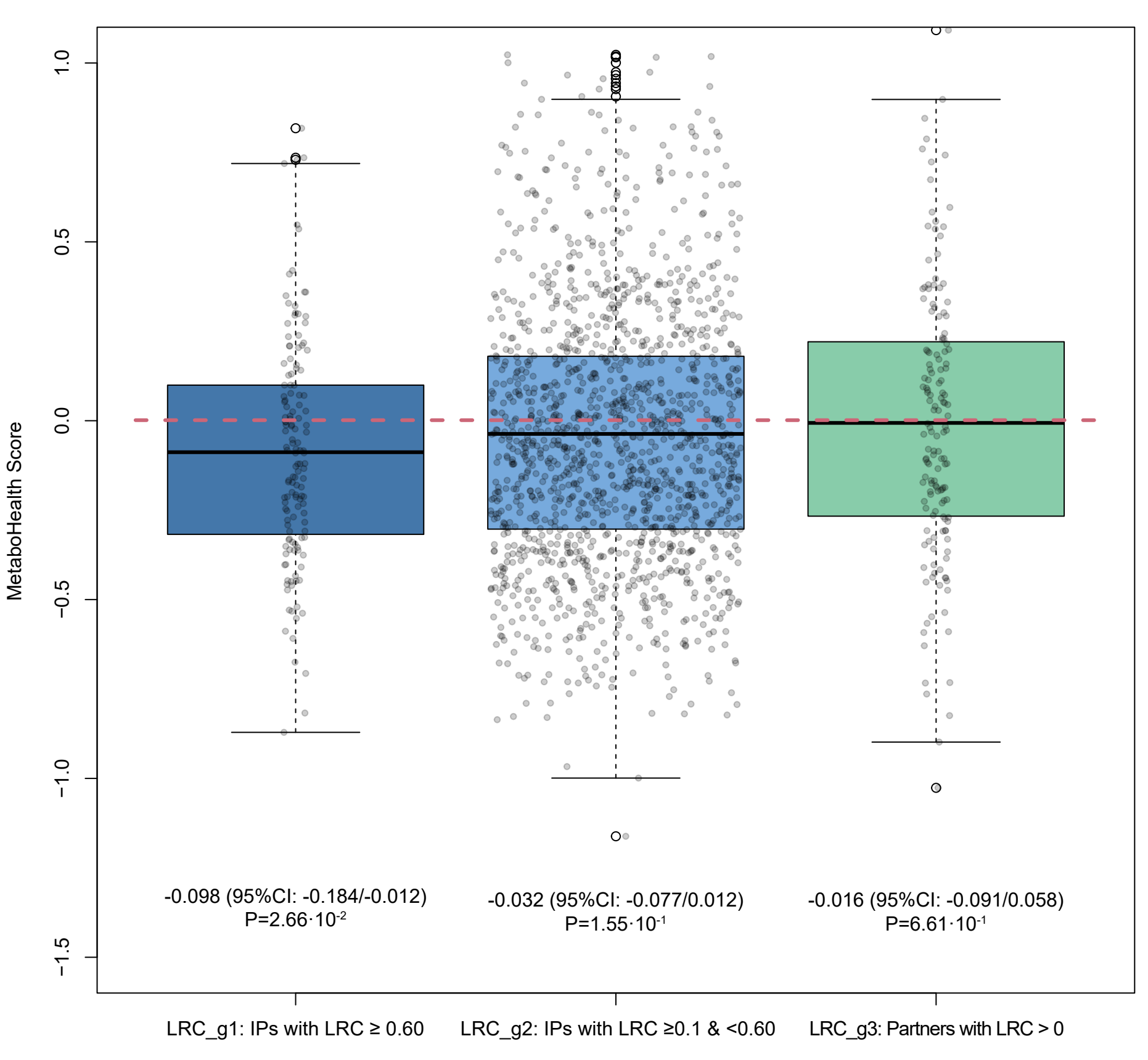

MetaboHealth Score
1.0
0.5
0.0
−0.5
−1.0
−1.5
-0.098 (95%CI: -0.184/-0.012)
P=2.66·10-2
-0.032 (95%CI: -0.077/0.012)
P=1.55·10-1
-0.016 (95%CI: -0.091/0.058)
P=6.61·10-1
LRC_g1: IPs with LRC ≥ 0.60
LRC_g2: IPs with LRC ≥0.1 & <0.60
LRC_g3: Partners with LRC > 0

Table 1: Basic characteristics of LLS Index Persons, partners, and ancestral groups

	Analyses groups		Ancestral family groups			
	IPs F3	Partners of IPs F3	Parents of IPs F2	Aunts/uncles of IPs F2	Grandparents of IPs F1	Parents of partners F2
Panel A: LLS						
Number, N	1674	745	1295	2370	760	1237
Female, N (%)	905 (54)	429 (58)	646 (50)	1169 (49)	380 (5)	620 (5)
Range birth cohorts, years	1923-1971	1924-1974	1882-1928	1875-1951	1850-1894	1864-1947
Alive, N (%)	1409 (84)	619 (83)	116 (9)	368 (16)	0 (0)	254 (21)
Deceased, N (%)	227 (14)	113 (15)	1179 (91)	2001 (84)	760 (100)	1011 (82)
Missing age, N (%)	38 (2)	13 (2)	94 (7)	34 (1)	0 (0)	35 (3)
Mean ad or al, years (SD)	74.87 (6.67)	74.21 (7.22)	87.09 (14.13)	71.93 (28.03)	77.06 (14.15)	76.68 (13.86)
Disease free, N (%)	535 (32)	206 (28)	-	-	-	-
Diseased, N (%)	671 (40)	324 (43)	-	-	-	-
Missing disease, N (%)	468 (28)	215 (29)	-	-	-	-
Mean number of diseases, N, (SD)	0.82 (0.92)	0.94 (0.98)	-	-	-	-
Panel B: SEDD						
Number, N	2497	-	2969	5830	3028	-
Female, N (%)	1252 (50.1)	-	1480 (49.8)	2866 (49.2)	1532 (50.6)	-
Range birth cohorts, years	1930-1950	-	1853-1914	1860-1950	1833-1918	-
Alive, N (%)	1803 (72.2)	-	7 (0.2)	855 (14.7)	131 (4.3)	-
Deceased, N (%)	694 (27.8)	-	2962 (99.8)	4975 (85.3)	2897 (95.7)	-
Missing age, N (%)	-	-	-	-	-	-
Mean ad or al, years (SD)	74.6 (6.7)	-	76.0 (17.9)	57.4 (32.2)	64.0 (19.6)	-
Disease free, N (%)	1307	-	-	-	-	-
Diseased, N (%)	1190	-	-	-	-	-
Missing disease, N (%)	-	-	-	-	-	-
Mean number of diseases, N, (SD)	-	-	-	-	-	-

Index Persons (IPs) and their partners were included in the Leiden Longevity Study (LLS) between 2002 and 2006. The table is separated by analyses groups and ancestral family groups. The analyses group information represent the persons (IPs and their Partners) that are investigated in this study, as a complete group or in different subgroups. The ancestral family group information represent the ancestors of the IPs and their Partners which are used to calculate the Longevity Relatives count (LRC) score. Panel A: For all groups, except the Parents of partners (F2), mortality information was updated in January 2021. All mortality information was obtained from the official and verified Netherlands Population Registers. Mortality information for the Parents of partners (F2) was obtained from questionnaires filled in by the partners of IPs (F3). Disease (morbidity) information was updated in 2018 based input from the General Practitioners of the IPs and their Partners. The study covers a maximal mortality follow-up of 19 years (2002-2021) and maximal morbidity follow-up of 16 years (2002-2018). Panel B: SEDD IPs were followed over time from 1990, at an average age of 52 years, with a maximal mortality and morbidity (disease) follow-up of 25 years (1990-2015). All mortality and morbidity data was obtained from the SEDD which is linked to the National Swedish Registers (see methods for more details). Because register data was used, there was no missing data for ages and diseases. Moreover, we focused on first diseases and therefore no mean number of diseases can be provided for the SEDD data.

Table 2: LLS disease incidence

	N (prop)	Events (prop)	HR (95% CI)	P-Value
	A: Time from inclusion to first disease			
Age related diseases				
LLS IPs	917 (0.70)	362 (0.39)	0.79 (0.65-0.97)	$2.32 \cdot 10^{-2}$
Partners (ref)	395 (0.30)	171 (0.43)		
Metabolic diseases				
LLS IPs	917 (0.70)	261 (0.23)	0.71 (0.55-0.90)	$5.18 \cdot 10^{-3}$
Partners (ref)	395 (0.30)	135 (0.34)		
Malignancies				
LLS IPs	917 (0.70)	130 (0.14)	0.95 (0.70-1.31)	$7.66 \cdot 10^{-1}$
Partners (ref)	395 (0.30)	56 (0.14)		
	B: Time from inclusion to 2 diseases			
Age related diseases				
LLS IPs	611 (0.70)	73 (0.12)	0.55 (0.36-0.85)	$6.83 \cdot 10^{-3}$
Partners (ref)	268 (0.30)	47 (0.18)		
Metabolic diseases				
LLS IPs	611 (0.70)	45 (0.07)	0.51 (0.26-0.97)	$3.96 \cdot 10^{-2}$
Partners (ref)	268 (0.30)	29 (0.11)		
Malignancies				
LLS IPs	611 (0.70)	7 (0.01)	1.39 (0.29-6.70)	$6.82 \cdot 10^{-1}$
Partners (ref)	268 (0.30)	2 (0.01)		
	C: Time from first disease to second disease			
Age related → age related diseases				
LLS IPs	500 (0.68)	79 (0.16)	0.46 (0.26-0.83)	$9.82 \cdot 10^{-3}$
Partners (ref)	237 (0.32)	55 (0.23)		
Age related → Metabolic diseases				
LLS IPs	500 (0.68)	62 (0.12)	0.33 (0.14-0.81)	$1.46 \cdot 10^{-2}$
Partners (ref)	237 (0.32)	44 (0.19)		
Age related → Malignant diseases				
LLS IPs	500 (0.68)	17 (0.03)	0.58 (0.27-1.25)	$1.66 \cdot 10^{-1}$
Partners (ref)	237 (0.32)	11 (0.05)		

Table shows the time from inclusion to first disease in panel A, the time from inclusion to having 2 diseases (panel B), and the time from first to second disease (panel C). N is the group size used for the analyses and prop. Is the proportion from the total. Events are the events of the specific diseases, for example age related diseases, and prop. indicates the proportion from the size of a specific group (LLS IPs or partners). For example, in panel A, there are 917 offspring and 395 partners. The total is 1312. 917 is 70% out of the total and 395 is 30% of the total. 362 events are then 39% out of the 917 offspring and 171 events are 50% out of the 395 partners. HR is the abbreviation for Hazard Ratio. Statistical testing was performed using Wald tests for the conditional log-hazard ratio estimated with a Cox-type frailty regression model. The analyses are adjusted for sex, different ages of study entry (left truncation) and right censoring. Survival curve details can be found in Supplementary Figure 2-4. To study time from inclusion to first disease, only persons without any disease at inclusion were studied. Note that the numbers in panel B are lower than those in panel A. This is because the censoring group reflects those for whom we have not observed any disease at the end of follow-up. As a results, persons with only one disease are excluded from the analyses. Moreover, the analyses in panel C are confined to persons who experienced a first disease. The numbers are larger than the events reported in panel A because we allowed for 1 disease prior to the start of follow-up (see methods section for more details).

Table 3: LLS disease incidence and medication use in LLS LRC groups

Panel A: disease incidence

	A: Time from inclusion to first disease			
	N (prop)	Events (prop)	HR (95% CI)	P-Value
Age related diseases				
LRC_g4: partners with LRC = 0 (ref)	262 (0.21)	113 (0.43)		
LRC_g3: Partners with LRC > 0	89 (0.07)	38 (0.43)	0.85 (0.58-1.26)	$4.29 \cdot 10^{-1}$
LRC_g2: IPs with LRC ≥0.1 & <0.60	843 (0.66)	339 (0.40)	0.80 (0.63-1.01)	$5.98 \cdot 10^{-2}$
LRC_g1: IPs with LRC ≥ 0.60	74 (0.06)	23 (0.31)	0.56 (0.34-0.92)	$2.17 \cdot 10^{-2}$
Metabolic diseases				
LRC_g4: partners with LRC = 0 (ref)	262 (0.21)	88 (0.34)		
LRC_g3: Partners with LRC > 0	89 (0.07)	30 (0.34)	0.82 (0.51-1.30)	$3.89 \cdot 10^{-1}$
LRC_g2: IPs with LRC ≥0.1 & <0.60	843 (0.66)	246 (0.30)	0.72 (0.54-0.95)	$2.17 \cdot 10^{-2}$
LRC_g1: IPs with LRC ≥ 0.60	74 (0.06)	15 (0.20)	0.47 (0.25-0.87)	$1.72 \cdot 10^{-2}$
Malignancies				
LRC_g4: partners with LRC = 0 (ref)	262 (0.21)	37 (0.14)		
LRC_g3: Partners with LRC > 0	89 (0.07)	14 (0.16)	1.04 (0.56-1.94)	$8.93 \cdot 10^{-1}$
LRC_g2: IPs with LRC ≥0.1 & <0.60	843 (0.66)	122 (0.13)	0.97 (0.64-1.46)	$8.89 \cdot 10^{-1}$
LRC_g1: IPs with LRC ≥ 0.60	74 (0.06)	8 (0.11)	0.69 (0.31-1.53)	$3.62 \cdot 10^{-1}$
	B: Time from inclusion to 2 diseases			
Age related diseases				
LRC_g4: partners with LRC = 0 (ref)	177 (0.21)	31 (0.18)		
LRC_g3: Partners with LRC > 0	60 (0.07)	9 (0.15)	0.66 (0.28-1.54)	$3.33 \cdot 10^{-1}$
LRC_g2: IPs with LRC ≥0.1 & <0.60	565 (0.67)	71 (0.13)	0.56 (0.34-0.93)	$2.50 \cdot 10^{-2}$
LRC_g1: IPs with LRC ≥ 0.60	46 (0.05)	2 (0.04)	0.14 (0.03-0.70)	$1.65 \cdot 10^{-2}$

Panel B: medication use

	N (prop)	Events (prop)	OR (95% CI)	P-Value
ATC-C (Metabolic medication)				
LRC_g4: partners with LRC = 0 (ref)	428 (0.21)	144 (0.35)		
LRC_g3: Partners with LRC > 0	149 (0.07)	43 (0.29)	0.70 (0.39-1.25)	$2.26 \cdot 10^{-1}$
LRC_g2: IPs with LRC ≥0.1 & <0.60	1363 (0.66)	306 (0.23)	0.43 (0.30-0.63)	$1.02 \cdot 10^{-5}$
LRC_g1: IPs with LRC ≥ 0.60	124 (0.06)	24 (0.19)	0.26 (0.12-0.57)	$7.16 \cdot 10^{-4}$

In this table the analyses with the largest sample size are repeated after defining new groups based on the LLS IPs and partner ancestral mortality information. LLS IPs were separated into two groups. Group 1 are those who have 60% of their ancestor belonging to the top 10% of their birth cohort (LRC_g1) and group 2 are the other, original LLS, offspring (LRC_g1). Partners were also separated into two groups. Group 3 are those who have at least one parent that belongs to the top 10% survivors of his/her birth cohort (LRC_g3) and group 4 are those without any parents who belong to the top 10% survivors of their birth cohort (LRC_g4). As some ancestors of LLS IPs were still alive, they were excluded from the LRC score (see methods for more details). 236 ancestors of partners were still alive. To not lose those who were still alive for our analyses, we used single imputation to predict the remaining life expectancy after having reached a specific age (last age of observation) based on the Dutch population life expectancy. Lifetables were provided by Statistics Netherlands from 1850 up to 2021 (see methods for more details). Statistical testing was performed using Wald tests for the conditional log-hazard ratio estimated with a Cox-type frailty regression model. The analyses are adjusted for sex, different ages of study entry (left truncation) and right censoring.

Table 4: Quantitative LRC analyses of time to first disease in LLS en SEDD

	N	Events (prop)	HR (95% CI)	P-Value
	A: LLS data			
Age related diseases	1312	533 (0.41)	0.95 (0.91-0.99)	$3.06 \cdot 10^{-2}$
Metabolic diseases	1312	396 (0.30)	0.93 (0.88-0.98)	$1.37 \cdot 10^{-2}$
Malignancies	1312	186 (0.14)	0.97 (0.91-1.04)	$4.98 \cdot 10^{-1}$
	B: SEDD data			
Age related diseases	2497	1,190 (0.48)	0.94 (0.89-0.98)	$1.22 \cdot 10^{-3}$
Metabolic diseases	2497	706 (0.28)	0.91 (0.87-0.96)	$4.84 \cdot 10^{-3}$
Malignancies	2497	671 (0.27)	0.95 (0.90-0.99)	$4.43 \cdot 10^{-2}$
Death	2497	694 (0.28)	0.92 (0.87-0.97)	$1.72 \cdot 10^{-3}$

Table shows the time from inclusion to first disease. N is the group size. Events are the events of the specific diseases, for example age related diseases, and prop. indicates the proportion from the size of a specific disease group. for example, 533 (41%) out of the 1312 persons had an age-related disease. sd indicates the standard deviation. HR is the abbreviation for Hazard Ratio. Statistical testing was performed using Wald tests for the conditional log-hazard ratio estimated with a Cox-type frailty regression model. The analyses are adjusted for sex, different ages of study entry (left truncation) and right censoring. Only persons without any disease at inclusion are studied.