

NEWS AND COMMENTARY

Combining data sources

# Detecting genetic variants for extreme aging using multiple data sources

Jeanine J Houwing-Duistermaat

*European Journal of Human Genetics* (2013) **21**, 359–360; doi:10.1038/ejhg.2012.183; published online 15 August 2012

It has been long speculated that the effect of genetic variants on survival changes with age. Evidence for age-specific effects is, however, hard to obtain as most studies are underpowered or not suited to detect age-specific effects. Tan *et al*<sup>1</sup> in this issue present an elegant solution by borrowing information from population survival registries. They show age-specific effects of the e4 allele of APOE: the relative risks for carriers versus non-carriers are around 1.15 from age 92 to 99 years and then increase to around 1.22 up to age 104 years.

Demographic research has shown that age-specific mortality rates stabilize at very late ages. The plateaued mortality pattern implies that late life is a distinct phase of life history. As a consequence, the role of genetic variants may also change over time. Indeed, age-specific effects have been identified in animal models,<sup>2</sup> whereas for humans results have been lacking. Case-control studies have been used to assess differences in genotype frequencies among age categories. This approach has, however, a serious drawback, namely the identified effects may also be caused by differences between birth cohorts in exposure to environmental factors, such as introduction of penicillin or lack of food during World War I. An observational study on the other hand that follows the same birth cohort over time is more appropriate, but one has to be patient to collect enough follow-up data to be able to assess age-specific effects. Note that a retrospective approach cannot be used

in this case, because the genotypes of subjects are needed. Tan *et al*<sup>1</sup> have access to a unique follow-up study: the Danish 1905 birth cohort. From this cohort, 2662 individuals were genotyped at 92–93 years of age in 1998. Individual survival was collected, and the last update on survival information was in 2010 when 10 subjects were still alive. Although they have over 12 years of follow-up time, this data set does not contain sufficient information yet to estimate both the underlying hazards and the age-specific genotypic relative risks.

Previous work on this data set illustrates the challenge to detect age-specific effects. In 2006, they were not able to detect an age-specific effect for carriers of the e4 allele, due to limited follow up: at that moment the length of the follow-up period was around 5 years.<sup>3</sup> In 2010, reanalysis of the same data set complemented with data from the 1895–1896 birth cohort showed an increased hazard ratio for carriers of the e4 allele versus non-carriers after 98 years of age.<sup>4</sup> Even though the addition of the second cohort was needed to improve power for the age categories above 99 years of age, the introduction of possible bias due to heterogeneity between the two birth cohorts may be an issue. In the current paper,<sup>1</sup> the authors concentrate on the 1905 birth cohort and circumvent the estimation of baseline hazard by using cohort-specific survival information from the Human Mortality database. This database comprises information of the total of 3600 persons aged 92–93 years and still alive in 1998, which is the cohort to which the 2662 subjects of the Danish 1905 birth cohort belong. By adding the population survival information into the data analysis and using a constrained

likelihood, the estimates of the age-specific relative risks are more efficient. Indeed, significant age-specific associations between APOE genotypes and mortality were obtained using data from a single birth cohort.

Several studies have considered the use of additional information to improve statistical power. For example, for testing the presence of genetic linkage in selected samples, the data from registries was used to weight individuals according to current ages and or covariate patterns.<sup>5</sup> Typically, these weights are not optimal, because the additional information does not completely represent the analyzed data sample. Despite misspecification of the weights, the score statistic is valid as it is calculated under the null hypothesis, but model parameter estimates may be biased in such situations. The cohort-specific mortality rates used by Tan *et al*<sup>1</sup> are probably very similar to the mortality rates of the group of genotyped subjects. Small differences, however, may be present; for example, due to the fact that the genotyped participants are healthier and therefore have a slightly better survival prediction than the members of the birth cohort who do not participate in the study. A statistical solution for such a situation is to assume that the survival pattern of the genotyped subjects is similar to their birth cohort but not equal. Such a concept of similarity can be included by using a penalized likelihood instead of a constrained likelihood as the authors have done. Large penalties correspond to smaller differences between the survival pattern of the genotyped subjects and their birth cohort. Such an approach can also be viewed as a Bayesian approach where a prior is used for the underlying hazard.

Another issue the authors consider in their paper is heterogeneity due to missing covariates. It is well known that effect sizes are attenuated when important covariates are not included in the model. It is likely that in addition to APOE other unknown genetic factors have a role, as it has been suggested that the genetic influence increases with age.<sup>6</sup> Frailty models is an approach to adjust for heterogeneity in samples due to omitted covariates such as genetic factors. In such a model, subjects with deleterious covariate profiles will have large frailties, whereas subjects with beneficial covariate patterns will have relatively small frailties. Tan *et al*<sup>1</sup> also consider a survival model including a gamma distributed frailty with mean equal to 1 and a variance of 0.1. Some caution is needed here: assumptions on the frailty distributions influence the parameter estimates. Keiding *et al*<sup>7</sup> illustrated the instability of

Dr JJ Houwing-Duistermaat is at the Department of Medical Statistics and Bioinformatics, Leiden University Medical Center, Albinusdreef 2, Postzone S5-P, Kamer T-05-058, Leiden 2333 ZA, The Netherlands.  
Tel: +31 71 5269707; +31 71 5269700; Fax: +31 71 5268280; E-mail: j.j.houwing@lumc.nl

parameter estimates in frailty models fitted to data sets with no replications per heterogeneity unit. Analogously to using additional information on population survival, information on frailty distributions should be included in the model while accounting for uncertainty in parameter values by using penalized likelihoods. Furthermore, the hope is that longitudinal family studies provide more information on the frailty distribution.<sup>8</sup>

To conclude, Tan *et al*<sup>1</sup> are the first to show that carriers of the e4 alleles have a higher mortality than non-carriers in a specific population. A method that accounts for uncertainty in overall survival and frailty parameter values should, however, be advocated for this type of data problem. The availability of follow-up studies in the oldest old, the feasibility of

genetic sequencing, together with advanced statistical modeling for various data sources provide the opportunity to further unravel the genetic basis of extreme aging in the near future ■

#### CONFLICT OF INTEREST

The author declares no conflict of interest.

- 1 Tan Q, Jacobsen R, Sorensen M, Christiansen L, Kruse TA, Christensen K: Analyzing age-specific genetic effects on human extreme age survival in cohort-based longitudinal studies. *Eur J Hum Genet* 2012; doi:10.1038/ejhg.2012.182.
- 2 Leips J, Gilligan P, Mackay TF: Quantitative trait loci with age-specific effects on fecundity in *Drosophila melanogaster*. *Genetics* 2006; **172**: 1595–1605.

- 3 Bathum L, Christiansen L, Jeune B, Vaupel J, McGue M, Christensen K: Apolipoprotein e genotypes: relationship to cognitive functioning, cognitive decline, and survival in nonagenarians. *J Am Geriatr Soc* 2006; **54**: 654–658.
- 4 Jacobsen R, Martinussen T, Christiansen L *et al*: Increased effect of the ApoE gene on survival at advanced age in healthy and long-lived Danes: two nationwide cohort studies. *Aging Cell* 2010; **9**: 1004–1009.
- 5 Callegaro A, van Houwelingen JC, Houwing-Duistermaat JJ: Score test for age at onset genetic linkage analysis in selected sibling pairs. *Stat Med* 2009; **28**: 1913–1926.
- 6 vB Hjelmborg J, Iachine I, Skytthe A *et al*: Genetic influence on human lifespan and longevity. *Hum Genet* 2006; **119**: 312–321.
- 7 Keiding N, Andersen PK, Klein JP: The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Stat Med* 1997; **16**: 215–224.
- 8 Hsu L, Gorfine M: Multivariate survival analysis for case-control family data. *Biostatistics* 2006; **7**: 387–398.