

ARTICLE

# Analyzing age-specific genetic effects on human extreme age survival in cohort-based longitudinal studies

Qihua Tan<sup>\*,1,2</sup>, Rune Jacobsen<sup>1</sup>, Mette Sørensen<sup>1</sup>, Lene Christiansen<sup>1,2</sup>, Torben A Kruse<sup>2</sup> and Kaare Christensen<sup>1,2,3</sup>

The analysis of age-specific genetic effects on human survival over extreme ages is confronted with a deceleration pattern in mortality that deviates from traditional survival models and sparse genetic data available. As human late life is a distinct phase of life history, exploring the genetic effects on extreme age survival can be of special interest to evolutionary biology and health science. We introduce a non-parametric survival analysis approach that combines population survival information with individual genotype data in assessing the genetic effects in cohort-based longitudinal studies. Our approach is characterized by non-parametric analysis of late age survival to capture the observed pattern of mortality deceleration and frailty modeling to account for individual heterogeneity in unobserved frailty. The method is applied to ApoE genotype data in the Danish 1905 birth cohort to estimate effect of the e4 allele. Our results revealed an age-specific relative risk of the allele that increases nonlinearly with age and non-proportional patterns in hazard of death for carriers and non-carriers of the allele, suggesting that the e4 mutation preserves its deleterious effect that progressively affect human survival even at extreme ages.

*European Journal of Human Genetics* (2013) **21**, 451–454; doi:10.1038/ejhg.2012.182; published online 15 August 2012

**Keywords:** genetic effect; extreme age survival; cohort design; population data; frailty modeling

## INTRODUCTION

The evolutionary theory of aging assumes that the effect of a gene could change over an individual's life course as genetic mutation functioning at late ages are subject to weaker selection than early-acting mutations.<sup>1</sup> The age-specific genetic effects have been shown to affect fitness traits in animal models.<sup>2</sup> In humans, age-specific effects of genetic variations have been reported to influence body mass index,<sup>3</sup> blood pressure<sup>4,5</sup> and survival.<sup>6</sup> At late life, the force of natural selection during the reproductive period stops. In term of survival, mortality deviates significantly from the popular Gompertz model with a reliable attribute characterized by deceleration in age-specific mortality rates.<sup>7</sup> The paradoxical 'plateaued' mortality pattern implies that late life is a distinct phase of life history<sup>8</sup> for which exploring the genetic effects can be of special interest to evolutionary biology and health science.

The estimation of an age-dependent genetic effect on survival can often be confounded by differential life course exposure to environmental factors or the birth cohort effect in age-structured populations.<sup>9</sup> For that reason, a good choice is to conduct a follow-up or longitudinal study on a birth cohort, which has only been feasible in animal experiments. In human studies, however, longitudinal analysis on genetic association with human longevity can be done with old-aged birth cohorts, for example, the Danish 1905 birth cohort,<sup>10</sup> to look for genes that affect extreme age survival.<sup>9,11,12</sup> Although of great interest, estimating genetic effects on late life survival is confronted with the distinct mortality pattern and sparse genetic data available. In the literature, different theories or models have been proposed to

explain the late life-mortality pattern,<sup>8</sup> among them the heterogeneity model,<sup>13</sup> which assumes individual heterogeneity in unobserved frailty that follows a gamma distribution. Jacobsen *et al*<sup>11</sup> applied a Cox regression model with gamma-distributed frailty to the Danish 1905 birth cohort data to estimate the age-dependent effect on extreme age survival for the ApoE gene, the only gene whose role on longevity has been consistently demonstrated.<sup>14</sup> This paper introduces a demographic heterogeneity model that combines sparse individual genotype data with population survival information to measure age-specific genetic effect on survival at advanced ages. The method is applied to ApoE genotype data from the Danish 1905 birth cohort<sup>10</sup> to illustrate the patterns of the age-specific effect of the e4 allele in affecting extreme age survival. Results with and without consideration of unobserved frailty will be compared and genotype-specific mortality patterns illustrated.

## METHODS

For a given genetic variation, for example, a SNP, individuals can be grouped according to their genotypes for a certain allele as non-carriers (0 allele), heterozygous (1 allele) and homozygous (2 alleles) carriers based on which effect of the allele can be assumed to be additive, dominant or recessive. For simplicity, we divide individuals as carriers and non-carriers of the allele, which is equivalent to a dominant assumption. In term of survival, the population survival rate in a birth cohort is the weighted mean for allele carriers ( $\geq 1$  allele) and non-carriers (0 allele),<sup>15</sup>

$$\bar{s}(x) = ps_1(x) + (1-p)s_0(x) \quad (1)$$

<sup>1</sup>Epidemiology, Institute of Public Health, University of Southern Denmark, Odense C, Denmark; <sup>2</sup>Department of Clinical Genetics, Odense University Hospital, Odense C, Denmark; <sup>3</sup>Department of Biochemistry and Pharmacology, Odense University Hospital, Odense C, Denmark

\*Correspondence: Dr Q Tan, Epidemiology, Institute of Public Health, University of Southern Denmark, J. B. Winsløvs Vej 9B, Odense C DK-5000, Denmark. Tel: +45 65503536; Fax: +45 65411911; E-mail: qtan@health.sdu.dk

Received 30 March 2012; revised 21 May 2012; accepted 31 May 2012; published online 15 August 2012

Here,  $\bar{s}(x)$  is the mean survival rate in the birth cohort at age  $x$ ,  $p$  is frequency of carriers of the allele,  $s_1(x)$  and  $s_0(x)$  are survival rates for carriers and non-carriers of the allele. The relationship between  $s_1(x)$  and  $s_0(x)$  reflects relative risk of the allele on survival. In a simple proportional hazard model, the hazards of death corresponding to  $s_1(x)$  and  $s_0(x)$  are related as  $\mu_1(x) = r\mu_0(x)$  such that

$$s_1(x) = e^{-\int_0^x \mu_1(t) dt} = e^{-\int_0^x r\mu_0(t) dt} = \left( e^{-\int_0^x \mu_0(t) dt} \right)^r = s_0(x)^r. \quad (2)$$

The relationship above is based on the assumption that individuals are homogenous except for their genotypes of the allele. However, in reality, individuals are heterogeneous in their unobserved factors or frailty, including genetic make-ups, which serves as the basis for existing theories that explain mortality deceleration at advanced ages, among which is the demographic heterogeneity theory by Vaupel *et al.*<sup>13</sup> It follows that, when an individual's unobserved frailty designated as  $z$  is gamma-distributed with mean 1 and variance  $\sigma^2$ , instead of (2), the relationship between  $s_1(x)$  and  $s_0(x)$  becomes

$$\begin{aligned} s_1(x) &= (1 - r\sigma^2 \ln s(x))^{-1/\sigma^2} \\ s_0(x) &= (1 - \sigma^2 \ln s(x))^{-1/\sigma^2}. \end{aligned} \quad (3)$$

$s'(x)$  is a homogenous baseline survival function. Note that the integration of (3) with (1) combines population survival with genotype frequency and relative risk parameters, which allows assessment of genetic effect on survival.

Based on (1), the proportions of carriers and non-carriers of the allele at any age  $x$  can be estimated as  $p_1(x) = ps_1(x)/\bar{s}(x)$  and  $p_0(x) = (1-p)s_0(x)/\bar{s}(x) = 1 - p_1(x)$ , respectively. When genotype data is available for a random sample from the cohort, a likelihood function based on binomial distribution can be constructed at each age  $x$  as

$$\begin{aligned} LL(x) &\propto n_1(x) \log p_1(x) + n_0(x) \log p_0(x) \\ &\propto n_1(x) [\log p + \log s_1(x) - \log \bar{s}(x)] \\ &\quad + n_0(x) [\log(1-p) + \log s_0(x) - \log \bar{s}(x)] \end{aligned} \quad (4)$$

In (4),  $n_1(x)$  and  $n_0(x)$  are the number of counts for carriers and non-carriers of the allele at age  $x$ ,  $p$  is proportion of carriers in the population, which can be available for specific populations and  $\bar{s}(x)$  is population survival rate at age  $x$  obtainable from population statistics. With known  $\bar{s}(x)$  and  $p$ , (4) can be maximized to estimate the relative risk on survival for carrying the allele. In a longitudinal study on a birth cohort, (4) can be done for each age or year of follow-up so that age-specific effects can be estimated. The maximum likelihood estimation (MLE) is obtained by introducing a constraint as specified in (1) and optimization of (4) with numerical gradient and Hessian. Note that our MLE is free from specification of any parametric form for the survival function and is thus a non-parametric approach. In addition, it combines or makes use of population data in the analysis of genetic effect. Moreover, genotype-specific survival or mortality rates can be calculated at each age to further illustrate the genetic influence on mortality at advanced ages.

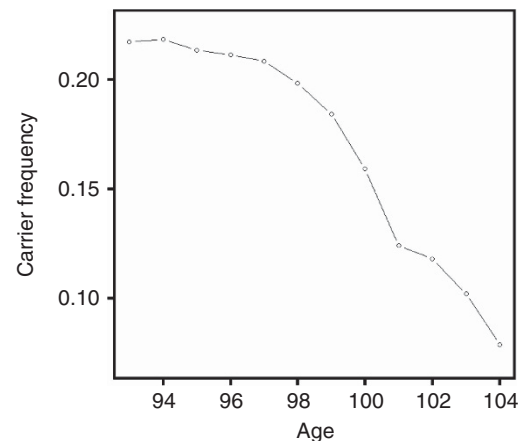
Finally, our model allows analysis of sex-specific effect<sup>16</sup> by simply replacing the mean cohort survival in (1) with survival rate for males or females and performing the analysis for each sex separately. However, because most of the survivors at extreme ages are females, insignificant results in males due to small sample sizes available may not necessarily mean sex-specific effect. In this case, a combined analysis should be preferred.

## RESULTS

We applied our method to the ApoE genotype data collected on 2662 individuals (584 males and 2078 females) from the Danish 1905 birth cohort<sup>10</sup> collected in a longitudinal survey initiated in 1998. All participants were genotyped at age 92–93 years. Individual survival information has been collected with the latest update at the end of 2010 when 10 subjects were still alive with their ages over 104 or 105. For the entire 1905 birth cohort, cohort-specific survival information is available from the Human Mortality Database at <http://www.mortality.org/> jointly hosted by University of California, Berkeley, and

the Max Planck Institute for Demographic Research, Rostock, Germany. For the ApoE gene, frequency for the allele of interest, that is, the e4 allele, was estimated to be 0.174 in the Danish population,<sup>17</sup> which corresponds to a carrier frequency of 0.318. As a 'thrifty' allele,<sup>18</sup> carriers of e4 have been shown to have a higher susceptibility to cardiovascular and Alzheimer's diseases, and are associated with higher mortality as compared with non-carriers under the contemporary environmental condition.<sup>17,19</sup> As such, frequency of the allele is expected to decrease with increasing age in a birth cohort. In our genotype data for the 1905 birth cohort, it is interesting to see that the decreasing pattern continues even at extreme ages starting with 21.7% at age 93 until 7.8% at age 104, a rapid decrease of about 14% in 11 years (Figure 1). The declining nonlinear pattern in e4 allele frequency that accelerates with age gives a clear indication of a deleterious effect of the allele on human extreme age survival, which needs to be characterized or measured by proper statistical models.

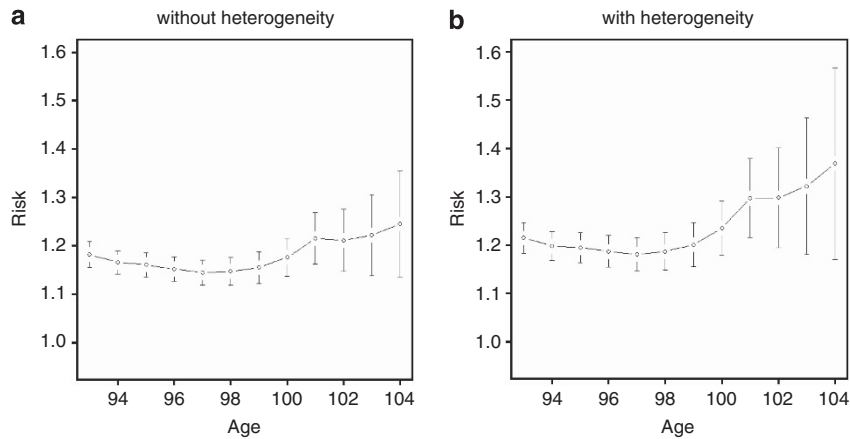
With known population survival for the entire 1905 birth cohort and frequency of e4 allele in the Danish population, we first fitted the likelihood function in (4) without frailty using genotype-specific survival as defined in (2). For each age  $x$ , our procedure estimated an age-specific relative risk on surviving from age  $x$  to  $x+1$  (Table 1). Our results showed that the estimated risks were all significantly different from one over all ages with a slight trend of increase at later



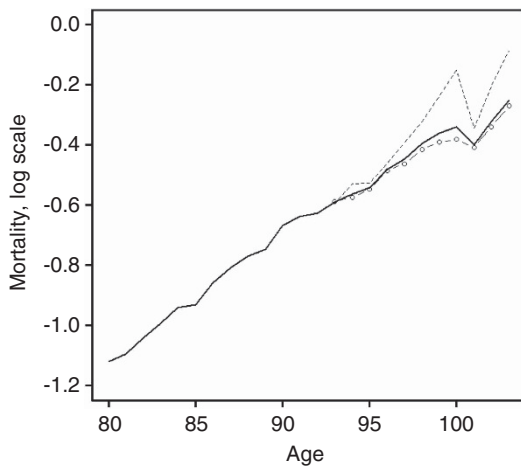
**Figure 1** Frequency of e4 allele carriers in the 2662 subject from the 1905 birth cohort starting from age 93 until age 104. There is a clear pattern of rapid decline as age increases, suggesting the increased risk of the allele on survival over advanced ages.

**Table 1** Estimated relative risk for e4 allele carriers with and without consideration of heterogeneity

Age	No frailty			Frailty		
	Relative risk	SE	P-value	Relative risk	SE	P-value
93	1.182	0.026	0.000	1.215	0.032	0.000
94	1.166	0.024	0.000	1.198	0.030	0.000
95	1.161	0.025	0.000	1.195	0.031	0.000
96	1.152	0.026	0.000	1.187	0.033	0.000
97	1.145	0.029	0.000	1.181	0.035	0.000
98	1.147	0.033	0.000	1.187	0.039	0.000
99	1.155	0.039	0.000	1.201	0.046	0.000
100	1.176	0.053	0.000	1.235	0.057	0.000
101	1.215	0.064	0.000	1.297	0.082	0.000
102	1.211	0.084	0.001	1.298	0.103	0.004
103	1.222	0.110	0.009	1.322	0.141	0.022
104	1.245	0.120	0.026	1.369	0.198	0.062



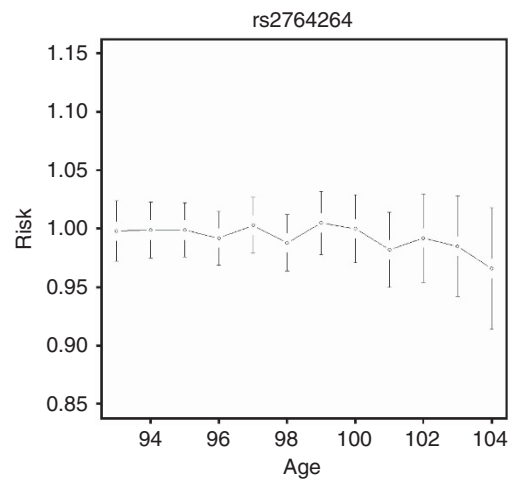
**Figure 2** Estimated age-specific relative risks for carrying the e4 allele over extreme age survival with 95% confidence intervals, which deviate from constant and increase slightly over ages. Risk estimates without (a) and with (b) consideration of unobserved heterogeneity show obvious underestimation by the former, suggesting the necessity of frailty modeling.



**Figure 3** Age-specific hazards of death for the whole 1905 birth cohort starting from age 80 (solid line) and e4 allele carriers (dashed line) and non-carriers (dash-dotted line) starting from age 93. The genotype-specific mortality deviates remarkably from proportional.

ages. Figure 2a plots the estimated age-specific relative risks together with their 95% confidence intervals. The figure clearly displays the increasing risk for the e4 allele in the oldest survivors. The highest risk of 1.23 ( $P = 0.026$ ) was obtained at the highest age of 104. We continued our analysis with frailty modeling by introducing gamma-distributed frailty with mean of 1 and variance of 0.1 (according to our experience in fitting frailty models to oldest-old mortality). From the estimated relative risks (Table 1), one could see that the frailty model gives higher risk estimates as compared with the no frailty model. In addition to the increased risk, the age-dependent increase in risk estimates is more clearly seen with frailty modeling, although the overall pattern of increase remained (Figure 2b).

Using the relationships in (2) and (3), age-specific survivals for carriers and non-carriers of the e4 allele can be calculated with the estimated relative risk and baseline survival rate. This allows calculation of age-specific hazard rate  $\mu(x)$  because  $\mu(x) = -d(\ln s(x))/dx$ . In Figure 3, we show the non-parametric age-specific hazard functions for the total population starting from age 80 (solid line) and the e4 allele carriers (dashed line) and non-carriers (dash-dotted line) starting from age 93. Although mortality patterns for carriers and non-carriers followed the main pattern of the whole cohort,



**Figure 4** Age-specific relative risks for carrying minor allele of SNP rs2764264 in the FOXO3A gene estimated with consideration of unobserved heterogeneity. No risk estimate reached statistical significance, although there is a slight trend of protective effect.

carriers had higher whereas non-carriers had lower instant probability of death than that for the mean population, and overall this deviation grew larger at later ages. Moreover, the population mortality pattern in Figure 3 also exhibits the mortality leveling-off at high ages, suggesting the necessity of frailty modeling.

Note that the calculated patterns of genotype-specific hazards were the same for both frailty and no frailty models as optimization of (4) was done for each age, however, the genetic risk was underestimated when unobserved heterogeneity in frailty is ignored.

Finally, we applied the frailty model to another example for SNP rs2764264 in the FOXO3A gene. The SNP was first reported to show association with human longevity in a case-control study conducted in the Italian population.<sup>20</sup> Recently, the SNP was tested in both case-control samples and the Danish 1905 cohort with the significant association replicated only in the case-control samples.<sup>21</sup> In Figure 4, we show age-specific risks estimated from our frailty modeling (frequency of carriers of minor allele set to 0.495 according to Soerensen *et al*<sup>21</sup>). Different from the e4 all of ApoE gene, no risk estimate in Figure 4 reached statistical significance, although there is a slight trend toward a protective effect similar to that reported in the literature in case-control studies.<sup>20,21</sup>

**Table 2** Relative risks on survival for ApoE4 carriers estimated in Danish cohort studies

Authors	Samples	Methods	Results
Bathum <i>et al</i> <sup>22</sup>	Danish 1905 birth cohort, censoring rate 17%	Cox model	Overall risk of 1.1 ( $P=0.07$ ) from ages 92 to 97+
Jacobsen <i>et al</i> <sup>11</sup>	Danish 1905 birth cohort, censoring rate 4%, plus samples from Longitudinal Study of Danish Centenarians	Aalen's additive hazards model	Risk of 1.1 ( $P=0.26$ ) from ages 92 to 97 and 1.4 ( $P=0.007$ ) from 97+
Current study	Danish 1905 birth cohort, censoring rate 4%	Frailty model combining genotype with population survival data	Age-specific risk up to age 104

## DISCUSSION

The cohort study is deemed as the most ideal design for assessing risk factors that affect human longevity<sup>9</sup> and in characterizing their age-specific effects. In humans, longitudinal following up for survival analysis is only feasible in very old cohorts, such as the Danish 1905 birth cohort. However, at advanced ages, human survival is characterized by mortality deceleration, which challenges conventional survival models.<sup>8</sup> We introduced a non-parametric survival analysis that combines population survival information with individual genotype data in estimating the genetic effects on human longevity. Our method conducts frailty modeling by introducing the simple gamma frailty model. Our comparison with a model that ignores unobserved heterogeneity showed underestimated genetic effect by the latter, which emphasizes the importance of frailty modeling in genetic risk assessment at advanced ages. The constraint likelihood for parameter estimation integrates population data with individual genotype data and allows non-parametric estimation of genetic risk parameters and the baseline survival function to avoid specification of parametric survival models that deviate from the observed mortality pattern. In addition to parameter estimation, our procedure also calculates non-parametric genotype-specific hazard of death over the observed ages to allow comparison with population mean death rate (Figure 3).

Our likelihood-based procedure is made possible by restricting estimation on each age separately. As an advantage, this allows measurement of age-specific genetic effect. As shown by Figure 2, the age-specific pattern of the estimated genetic risk deviates clearly from being constant or linear, which contradicts to the proportional hazard assumption. From the hazard functions for carriers and non-carriers of e4 allele, one can easily see that they are far from proportional. Such a pattern will be missed by traditional survival analysis, such as the Cox's proportional hazard model. In Table 2, we compare the different analyses that have been applied to the ApoE genotype data in the 1905 cohort. The early analysis (with high censoring rate of 17%) by Bathum *et al*<sup>22</sup> (Table 2) obtained an overall risk for e4 carriers, which was only borderline significant. Jacobsen *et al*<sup>11</sup> introduced Aalen's additive hazards model,<sup>23</sup> an extended Cox model, to estimate age-dependent risk assuming additive risks over age intervals. It is interesting that, when applied to the same updated data set (censoring rate 4%), their analysis also reported the increased effect of the e4 allele on longevity, although their analysis was limited to three age intervals. In comparison, our combined analysis of population and individual data enabled estimation for each age until the age as high as 104 years such that patterns of the mean genetic effects and genotype-specific mortality at extreme ages can be examined (Table 2). It can be expected that, with the rapid development in the SNP genotyping and genome sequencing, more genetic data will be available for association analysis of human extreme age survival for which proper statistical models can contribute.

## CONFLICT OF INTEREST

The authors declare no conflict of interest.

## ACKNOWLEDGEMENTS

This work was partially supported by the EU Seventh Framework Programme (FP7/2007–2011) under grant agreement no 259679 and NIH/NIA grant P01 AG08761.

- Zwaan BJ: The evolutionary genetics of ageing and longevity. *Heredity* 1999; **82** (Part 6): 589–597.
- Leips J, Gilligan P, Mackay TF: Quantitative trait loci with age-specific effects on fecundity in *Drosophila melanogaster*. *Genetics* 2006; **172**: 1595–1605.
- Atwood LD, Heard-Costa NL, Fox CS, Jaquish CE, Cupples LA: Sex and age specific effects of chromosomal regions linked to body mass index in the Framingham Study. *BMC Genet* 2006; **7**: 7.
- Tambs K, Eaves LJ, Moum T *et al*: Age-specific genetic effects for blood pressure. *Hypertension* 1993; **22**: 789–795.
- Joubert BR, Diao G, Lin D, North KE, Franceschini N: Longitudinal age-dependent effect on systolic blood pressure. *BMC Proc* 2009; **3**(Suppl 7): S87.
- Tan Q, Bathum L, Christiansen L *et al*: Logistic regression models for polymorphic and antagonistic pleiotropic gene action on human aging and longevity. *Ann Hum Genet* 2003; **67**(Part 6): 598–607.
- Vaupel JW, Carey JR, Christensen K *et al*: Biodemographic trajectories of longevity 1998. *Science*, **280**: 855–860.
- Rausser CL, Mueller LD, Rose MR: The evolution of late life. *Ageing Res Rev* 2006; **5**: 14–32.
- Tan Q, Kruse TA, Christensen K: Design and analysis in genetic studies of human ageing and longevity. *Ageing Res Rev* 2006; **5**: 371–387.
- Nybo H, Gaist D, Jeune B *et al*: The Danish 1905 cohort: a genetic-epidemiological nationwide survey. *J Aging Health* 2001; **13**: 32–46.
- 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. *Ageing Cell* 2010; **9**: 1004–1009.
- Sørensen M, Thinggaard M, Nygaard M *et al*: Genetic variation in TERT and TERC and human leukocyte telomere length and longevity: a cross-sectional and longitudinal analysis. *Ageing Cell* 2012; **11**: 223–227.
- Vaupel JW, Manton KG, Stallard E: The impact of heterogeneity in individual frailty on the dynamics of mortality. *Demography* 1979; **16**: 439–454.
- Christensen K, Johnson TE, Vaupel JW: The quest for genetic determinants of human longevity: challenges and insights. *Nat Rev Genet* 2006; **7**: 436–448.
- Vaupel JW, Yashin AI: Heterogeneity's ruses: some surprising effects of selection on population dynamics. *Am Stat* 1985; **39**: 176–185.
- Rosvall L, Rizzuto D, Wang HX, Winblad B, Graff C, Fratiglioni L: APOE-related mortality: effect of dementia, cardiovascular disease and gender. *Neurobiol Aging* 2009; **30**: 1545–1551.
- Gerdes LU, Klausen IC, Sihm I, Faergeman O: Apolipoprotein E polymorphism in a Danish population compared to findings in 45 other study populations around the world. *Genet Epidemiol* 1992; **9**: 155–167.
- Corbo RM, Scacchi R: Apolipoprotein E (APOE) allele distribution in the world. Is APOE\*4 a 'thrifty' allele? *Ann Hum Genet* 1999; **63**: Part 4 301–310.
- Tan Q, Christiansen L, Christensen K, Kruse TA, Bathum L: Apolipoprotein E genotype frequency patterns in aged Danes as revealed by logistic regression models. *Eur J Epidemiol* 2004; **19**: 651–656.
- Willcox BJ, Donlon TA, He Q *et al*: FOXO3A genotype is strongly associated with human longevity. *Proc Natl Acad Sci USA* 2008; **105**: 13987–13992.
- Soerensen M, Dato S, Christensen K *et al*: Replication of an association of variation in the FOXO3A gene with human longevity using both case-control and longitudinal data. *Ageing Cell* 2010; **9**: 1010–1017.
- 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.
- Aalen OO: A linear regression model for the analysis of life times. *Stat Med* 1989; **8**: 907–925.