

Methodological problems in genetic association studies of longevity—the apolipoprotein E gene as an example

Sarah J Lewis and Eric J Brunner

Accepted 30 March 2004

Background Cross-sectional genetic association studies are now widely employed to look for genes which confer longevity. Such studies are based on two assumptions; (a) initial relative allele frequencies in the different age cohorts are similar, and (b) the risk of mortality conferred by genotypes does not depend on year of birth.

Methods We explored the validity of these assumptions and reviewed 15 cross-sectional studies of common *apolipoprotein E (APOE)* polymorphisms and longevity.

Results Higher relative $\epsilon 2$ frequencies, and lower relative $\epsilon 4$ allele frequencies were observed in elderly versus younger populations. If assumptions (a) and (b) were correct the estimates for $\epsilon 2$ and $\epsilon 4$ alleles respectively versus $\epsilon 3$ alleles would be 1.34 (95% CI: 1.19, 1.35) and 0.54 (95% CI: 0.46, 0.63) in elderly versus younger individuals. However, there was an association between relative $\epsilon 4$ allele frequency in controls and *APOE* $\epsilon 4$ effect ($\beta = -0.45$, 95% CI: -0.89 , 0.00). In relation to assumption (a) there is substantial variation in relative *APOE* allele frequencies (4–21%), with considerable heterogeneity evident within geographically proximate populations, population admixture is likely to have resulted in changes in allele frequency over time, and assumption (b) *APOE* related causes of death are context specific and have changed considerably over the last 30 years.

Conclusion The validity of case-control type studies of the genetic basis of longevity based on the above assumptions is questionable, especially when considerable differences exist in allele frequency by population and when the genes in question interact with environmental factors, which vary by time and place.

Keywords Longevity, ageing, genetic association, gene, *APOE*, methodology

Twin studies of longevity suggest that around 25% of the variation in lifespan in developed countries may be inherited.¹ Understanding the heritability and consequently the biological processes involved in ageing and longevity is one of the greatest scientific challenges of our time. In the post genomic era and with large DNA resources becoming available within epidemiological studies, we are better placed than ever before to begin to unravel this complex problem.

Epidemiological studies are widely used to look for genes conferring risk to common multifactorial diseases by comparing genotype distributions in cases and controls.² More recently

case-control designs have been used to identify differences in genotypes or allele frequencies of elderly compared with younger populations.^{3,4} Centenarians are typically studied, as these represent survivors, who may lack disease susceptibility genes and genes associated with premature mortality. However, specific methodological problems are inherent in studies comparing populations across time periods, which although alluded to in several reviews, have not been adequately addressed in study designs to date. We argue that previous claims resulting from these studies may be flawed because they are based on the following assumptions: (a) initial relative allele frequencies in the different age groups are similar, and (b) the risk of mortality conferred by genotype does not depend on year of birth.

Further, due to secular trends in common disease incidence and changes in patterns of exposure to environmental risk factors such as smoking, genetic associations found in such

International Centre for Health and Society, Department of Epidemiology and Public Health, University College London, London, UK.

Correspondence: Sarah J Lewis, Department of Social Medicine, University of Bristol, Canynge Hall, Whiteladies Road, Bristol BS8 2PR, UK. E-mail: s.j.lewis@bristol.ac.uk

studies are unlikely to contribute to a blueprint for future longevity, unless improvements are made to the methodology used. An indication of the type of changes needed is provided in the concluding section of this paper.

Approaches to studying the genetics of ageing

Centenarians are quite rare and linkage studies require a relatively large number of long-lived sib pairs together with DNA from sibs who died early to avoid ascertainment bias. Ideally DNA would also be available from more than one generation. Genetic association studies are therefore more feasible, and are becoming widely used to look for genes which are important in longevity.^{5,6}

It is likely that many genes contribute to longevity. Candidate genes can be selected from either: (1) those genes implicated in general processes of cellular maintenance and repair, such as *superoxide dismutase 2 (SOD2)*, which is involved in the scavenging of reactive oxygen species and apoptosis; or (2) genes which confer an increased susceptibility to age-related diseases such as *apolipoprotein E (APOE)*, a gene involved in lipoprotein metabolism and a proposed risk factor for coronary heart disease (CHD) and Alzheimers disease.^{7,8} Since most genetic studies of longevity to date have been of the second type, this paper will focus on studies of disease-risk genes.

Initial gene frequencies in the age cohorts represented in a longevity study are likely to be different and not necessarily represented by frequencies in the younger/control population

Population stratification

Genetic association studies have received a large amount of negative press recently, mainly due to the lack of repeatability of results and the opportunity for spurious results.^{2,9–12} One reason for these problems is population stratification and this phenomenon is pertinent to studies of longevity. Whilst genuine differences between elderly cohorts and younger individuals may exist due to the role of genes in longevity, artefactual results of genetic effects may also arise due to variations in initial relative allele frequency between the two comparison groups, as a result of population admixture. Although novel mutations are infrequent and diffuse slowly, because the interval between human generations is long, the gene pool of a given population can change considerably in a relatively short period as a result of geographical mobility, ethnic admixing, and selective mating.¹⁰

Migration

If we take Britain as an example, over the last 100 years there has been mass emigration to the USA, South Africa, Australia, and New Zealand, coupled with an influx of individuals from India, Pakistan, Bangladesh, China, the Caribbean and Africa. More recently migration between countries in Europe has become common as a result of constitutional rights established within the European Union. The scale of movement is substantial. Of respondents to the decennial census in England, 13% of the population of England gave their ethnic origin as other than white British in the 2001, compared with 11% in 1991: an increase of 18% in 10 years.¹³ Since migration among

young adults tends to be for work or education, net inflows are often concentrated in cities, where genetic association studies may be based. In England the highest levels of in-flow are in the South East.¹⁴ In central London 25.3% of residents classified themselves as white but not British or Irish in the 2001 census.¹³ Whilst genetic association studies usually collect data on ethnic group, they rarely distinguish between Caucasians and hence group together individuals with different countries and regions of origin.

Migration within countries, so that individuals are better placed for work and health care may also have played a role in population mixing over the last century. Movement between regions can result in changes in the genetic structure of regional populations. Approximately 20% of individuals born in England, Scotland and Wales during 3–9 March 1958 resided in a different region in 1981 to that of their birth; this resulted in a reduction in interregional variability in ABO blood groups.¹⁵ Migration patterns differ widely between countries and time periods, so the scale of problem will depend on the study population.

The risk of mortality conferred by genotypes does depend on the birth year of the cohort

Secular trends in disease risk

Migration from different gene pools may change genetic risk within a given population and influence mortality rates over time periods of the order of decades. By addressing the first assumption we are in effect also addressing this dynamic. Aside from changes in population genetic structure, environmental risk factors for disease have changed considerably during the last 100 years and two populations who were born only a few decades apart are likely to have been exposed to a very different pattern of stressors. For this reason such populations will have different susceptibilities to disease regardless of genotype. Environmental risk factors are unlikely to change the population distribution of genetic variants; unless they offer strong selection against a particular allele in childhood and reproductive years. Therefore secular trends in genetic susceptibility to diseases of the elderly are not confounded by environment, but rather the environment acts as an effect modifier.

Life expectancy doubled in the first half of the 20th century, largely due to a reduction in malnutrition and childhood infection. Chronic diseases such as heart disease and cancer, which occur later in life, then became the major causes of death in developed countries.¹⁶ This shift in causes of mortality (the epidemiological transition) occurred over a relatively short period. It has been suggested that the rise in CHD and cancer incidence in Western countries is a consequence of ageing populations. However, the increase in rates of mortality from cancer which took place among men in Britain from 1921 to 1960 can be largely attributed to lung cancer, for which smoking is the main environmental cause.¹⁷ Levels of smoking in Britain reached a peak in the 1960s, followed by a steady decline, although approximately 30% of the adult population continue to smoke (Figure 1)¹⁶. Similarly, national CHD mortality rates rose and fell during the past century. Deaths from CHD reached a peak in many westernized countries in the 1970s and are now declining (Figure 2).¹⁸ Hence disease risks have changed over time, because levels of exposure to major environmental causes have changed.

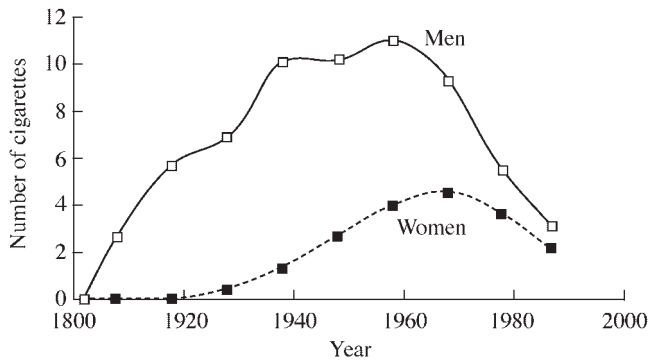


Figure 1 Number of cigarettes smoked per person per day in Britain¹⁶

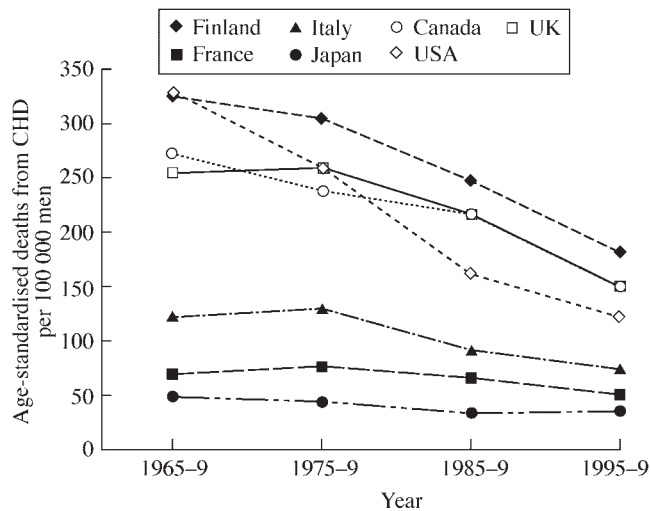


Figure 2 Mortality from cardiovascular disease in seven different countries¹⁸

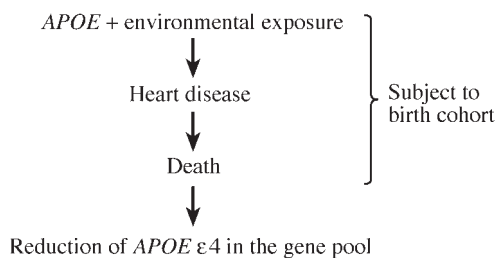


Figure 3 Schematic illustration of *APOE* and Longevity

Smoking, nutrition, physical activity, sexual behaviour, poverty, occupation, stress, education, social contacts, health care, pollution, war, and infectious disease are all factors which play a part in disease risk and longevity at the individual or societal level. Patterns of availability and exposure to many of them have changed enormously, with secular socio-economic and cultural trends, over the last 100 years, as we have seen in the case of smoking. Treatment for disease may also have an impact on mortality rates and will depend on the disease and the year that it is contracted.

***APOE* and longevity**

Possibly the most extensively studied gene in relation to longevity is *APOE*. The protein encoded by this gene is a ligand for the low density lipoprotein (LDL) receptor, other receptors in the LDL receptor gene family, and cell matrix-associated heparin sulphate proteoglycans. The gene plays an important role in lipoprotein metabolism, and has been shown to be associated with CHD risk.^{7,19} Three common polymorphic forms of *APOE* are known to exist: ε2, ε3, and ε4. The various *APOE* isoforms produced as a result of this polymorphism interact differently with specific lipoprotein receptors, ultimately altering circulating levels of cholesterol. High LDL cholesterol levels are associated with *APOE* ε4 and low levels with *APOE* ε2, with *APOE* ε3 levels being intermediate between the two.¹⁹ The *APOE* ε4 allele has also been associated with an increase in the risk of Alzheimer's disease.⁸ Although the mechanism is not entirely clear, this association could be due to: (1) effects on amyloid-β metabolism, (2) effects on neurons or glial cells, including neuronal survival and neurite extension, or (3) effects related to atherosclerosis, cerebral blood flow, or the blood-brain barrier.¹⁹ Since CHD and Alzheimer's disease are common disease of old age, *APOE* has become a candidate gene for studies of longevity. Several studies comparing *APOE* genotypes in elderly versus younger populations have been carried out²⁰⁻³⁴ and an overview of the literature by Gerdes *et al.*²⁸ found *APOE* ε4 to be associated with premature death.

To determine how inaccurate assumptions of initial *APOE* allele frequencies among elderly populations may influence the outcome of gene association studies, and whether any increased risk of mortality conferred by the *APOE* gene depends on year of birth, and to determine whether published studies have addressed these problems, we carried out a review of all studies comparing *APOE* allele frequencies in elderly versus younger populations.

Methods

Search strategy and selection criteria

Papers published prior to the end of April 2003 were identified through a search of *Medline* (www.ncbi.nlm.nih.gov), using the following search terms: 'longevity' or 'centenarian' or 'ageing' or 'ageing' and 'apolipoprotein E' or 'APO E' or 'APOE'. Publications were also identified by review of the bibliographies of retrieved articles. Studies were included if there was a control group from the same country, and where the elderly group was at least 80 years old. There were no restrictions on publications based on health of participants. Where results from the same study population were published more than once, the most recent report was used. Only studies published in English, or with an abstract in English were included.

Statistical methods

Unadjusted odds ratios (OR) were based on published allele frequencies and were calculated, as the proportion of ε2 alleles and ε4 alleles, in elderly versus control populations using relative ε3 allele frequencies as a reference. The effect of the *APOE* allele has been shown to be dependent upon genotype.^{35,36} However, the expected relative frequency of ε2ε2 and ε4ε4 homozygotes is low, around 1% and 2.5% respectively and stable estimates of the effect of these

homozygotes on longevity are not possible in small studies. Relative frequencies of $\epsilon 2\epsilon 2$ and $\epsilon 4\epsilon 4$ homozygotes were not available in 3 of the 15 studies reviewed here.^{24,29,32} Most studies compare allele frequencies, or combine $\epsilon 2\epsilon 2$ homozygotes with $\epsilon 3\epsilon 2$ heterozygotes and combine $\epsilon 4\epsilon 4$ homozygotes with $\epsilon 3\epsilon 4$ heterozygotes to give estimates for $\epsilon 2$ carriers versus $\epsilon 3\epsilon 3$ homozygotes and $\epsilon 4$ carriers versus $\epsilon 3\epsilon 3$ homozygotes. In using allele frequencies we have weighted the effect size by genotype frequency. However, an analysis by genotype instead of allele frequency gave similar results (not shown).

Random effects models were used to calculate summary OR and standard errors. Three studies provided genotype/allele frequencies for more than one age group in their control population;^{22,23,27} data from the different age groups in these studies was combined to give an overall estimate for age less than 80. Combined relative risks for males and females were presented. Forest plots were constructed with logarithmic scales. To assess the association between the frequency of the $\epsilon 2$ and $\epsilon 4$ alleles in the control group and the log OR of allele frequency in elderly populations versus controls, we used a Bayesian random-effects meta-regression model fitted using BUGS (WinBUGS 1.4, www.mrc-bsu.cam.ac.uk/bugs).³⁷ This avoids the bias that can be caused by the association between allele frequency and the OR that arises due to sampling error. All other statistical analysis was carried out in STATA version 8 (STATA corporation, Texas). Tests for funnel-plot asymmetry^{38,39} were used to investigate publication bias. Studies were grouped by geographical location: North America included studies from USA and Canada; Asia included studies from China, Japan, and Korea; Northern Europe included Finland, Denmark, and the UK; Central Europe included only France, and Southern Europe included Italy.

Results

Fifteen studies that met the inclusion criteria and characteristics of these studies are summarized in Table 1. The first study was carried out in Canada, one study was carried in the USA, four in Asia, and the remaining nine in Europe. OR in elderly versus control populations ranged from 0.56 (95% CI: 0.23, 1.39) to 2.44 (95% CI: 1.18, 5.06) for the presence of the $\epsilon 2$ allele, and 0.32 (95% CI: 0.10, 1.01) to 2.12 (95% CI: 0.79, 5.66) for the presence of the $\epsilon 4$ allele. The summary effect estimate was 1.34 (95% CI: 1.19, 1.35) for the presence of the $\epsilon 2$ allele and 0.54 (95% CI: 0.46, 0.63) for the presence of the $\epsilon 4$ allele in elderly versus younger individuals.

Large differences in relative *APOE* allele frequencies existed between control populations for $\epsilon 2$ (4–10%) and $\epsilon 4$ (4–21%). There were regional variations in *APOE* allele frequencies with high relative *APOE* $\epsilon 4$ in Northern Europe (15–21%), intermediate frequencies in Central Europe (11–13%), Asia (8–13%) and North America (12–15%), and low frequencies in Southern Europe (4–6%). In elderly people the $\epsilon 4$ allele frequencies converged to be more homogeneous (2–12%). Despite these differences, only 2 out of 15 studies reportedly matched on geographic origin and only one study matched on ethnicity. A meta-analysis including only these studies, gave a result closer to the null (RR = 0.74, 95% CI: 0.34, 1.56) for the presence of $\epsilon 4$ in elderly versus younger populations. A further

four studies stated that controls were of the same nationality,^{21,23,27,31} although this does not necessarily exclude migrants particularly second generation migrants.

The regression coefficient for the association between log odds of $\epsilon 4$ in the control group and the log OR of $\epsilon 4$ allele frequency in elderly versus controls was -0.45 (95% CI: $-0.89, 0.00$), providing some evidence of an association between the two. In a Finnish study²³ where the frequency of the $\epsilon 4$ allele was 20% an OR of 0.37 (95% CI: 0.25, 0.54) was obtained, whereas in an Italian study²⁶ where the $\epsilon 4$ frequency was 4% the OR was 2.12 (95% CI: 0.79, 5.66). There was not a significant association between $\epsilon 2$ effect size and relative $\epsilon 2$ allele frequencies in the control group ($\beta = -0.44$, 95% CI: $-1.07, 0.17$).

A meta-analysis considering only studies of centenarians gave a point estimate of 1.50 (95% CI: 1.27, 1.78) for $\epsilon 2$ and 0.49 (95% CI: 0.41, 0.58) for $\epsilon 4$ frequency among centenarians versus controls.

On the whole approaches for validating laboratory results were not stated in the papers. Although Blanche *et al.*³⁰ reported re-genotyping the 1994 cohort published by Schachter (1994),²¹ finding a discordance rate of 5.8%. A Begg's funnel plot (Figure 5) and an Egger's test of standardized effect versus precision did not provide evidence for publication bias ($P = 0.40$ Eggers and $P = 0.66$ Begg's).^{38,39}

Discussion

We have reviewed 15 cross-sectional studies of *APOE* polymorphisms and longevity to determine whether two key methodological assumptions have been addressed. The studies provide direct evidence that size of the genetic effect of interest is sensitive to the relative allele frequency among controls, and that relative *APOE* $\epsilon 4$ allele frequency exhibits wide variation between and within proximate populations. Our inference is that initial allele frequencies are likely to differ between the age cohorts from which the compared groups are drawn, and that assumption (a) is generally invalid. Further, the biology and epidemiology of presumed *APOE*-related diseases, in particular the great variation in coronary death rates by time and place, provides indirect evidence that the mortality risk conferred by *APOE* genotype depends on interaction with environmental and behavioural risk factors over the life course. The pattern of these exposures is closely related to the year of birth, and thus assumption (b) is likely to be invalid when gene-environment interaction is important, as is the case for *APOE*.

Since genetic studies of common polygenic disease are a recent phenomenon it is impossible to determine directly whether initial allele frequencies in elderly individuals are represented by younger controls and similarly whether any risk of mortality associated with common genetic variants depends on year of birth. However, in the case of *APOE* $\epsilon 4$ it is clear that allele frequencies differ widely by population, even among neighbouring countries and between geographically separated populations within the same country, for example between Lapps and Finns.⁴⁰ While migration patterns differ by country, there has been substantial internal and international migration in many countries within the last 100 years.⁴¹ These factors are likely to have produced differences in *APOE* allele frequency by age group within a given geographical area over recent decades. Migration is particularly prevalent among younger populations

Table 1 Cross-sectional studies of *apolipoprotein E (APOE)* polymorphisms and longevity

Reference	Place	No./age elderly	No./age Controls	Allele frequency (%) $\epsilon_2:\epsilon_3:\epsilon_4$		Adjustments	Odds ratio	Notes
				Elderly	Controls			
Davignon <i>et al.</i> , 1987 ²⁰	Canada	236/≥80	102/mean 36.1 ± 8.5	9: 82: 9	8: 77: 15	None	1.06 (0.58, 1.94) ϵ_2 0.53 (0.32, 0.88) ϵ_4	Octogenarians were able to answer a questionnaire, free of debilitating disease, mental illness, and CVD, with no recent major surgery. Normolipidemic controls.
Schachter <i>et al.</i> , 1994 ²¹	France	325/≥100	161/20–70	13: 82: 5	7: 82: 11	None	1.87 (1.14, 3.06) 0.47 (0.29, 0.76)	All centenarians were eligible.
Kervinen <i>et al.</i> , 1994 ²²	Oulu, Finland	95/≥90	260/21–64	6: 83: 12	4: 75: 21	None	1.37 (0.64, 2.93) ϵ_2 0.43 (0.26, 0.73) ϵ_4	All centenarians were eligible. Elderly included 12 diabetics. 87/95 elderly were women.
Louhija <i>et al.</i> , 1994 ²³	Finland	179/≥100	2192/3–55	7: 84: 8	4: 73: 20	None	1.65 (1.08, 2.54) ϵ_2 0.37 (0.25, 0.54) ϵ_4	All centenarians were included.
Galinsky <i>et al.</i> , 1997 ²⁴	Cambridge, UK	282/≥84	200/<17	8: 80: 12	10: 75: 15	None	0.78 (0.50, 1.22) ϵ_2 0.75 (0.51, 1.09) ϵ_4	All elderly individuals were eligible.
Hirose <i>et al.</i> , 1997 ²⁵	Tokyo, Japan	54/≥100	973/—	8: 83: 8	4: 85: 11	None	2.44 (1.18, 5.06) ϵ_2 0.78 (0.39, 1.58) ϵ_4	Reference in Japanese, figures taken from abstract.
Bader <i>et al.</i> , 1998 ³⁶	Abruzzo, Italy	93/≥80	84/20–70	4: 88: 8	8: 89: 4	Matched for origin	0.56 (0.23, 1.39) ϵ_2 2.12 (0.79, 5.66) ϵ_4	Elderly group were free-living with no history of CHD, respiratory or neurological disease, neoplasm, diabetes mellitus, dementia, severe arthritis, major depression, or alcoholism.
Jian-Gang <i>et al.</i> , 1998 ²⁷	Han Chinese, China	59/≥85	1503/20–84	8: 89: 2	7: 84: 8	None	1.17 (0.60, 2.27) ϵ_2 0.32 (0.10, 1.01) ϵ_4	Good physical and mental health.
Gerdes <i>et al.</i> , 2000 ²⁸	Denmark	177/≥100	466/40 Men only	13: 77: 10	8: 74: 17	None	1.44 (0.97, 2.13) ϵ_2 0.54 (0.37, 0.81) ϵ_4	All centenarians were eligible.
Rea <i>et al.</i> , 2001 ²⁹	Belfast, Ireland	114/>90	2071/30–65	12: 80: 8	8: 76: 16	None	1.58 (1.04, 2.40) ϵ_2 0.50 (0.30, 0.83) ϵ_4	Mentally competent, SENIUR status, live in same area as controls
Blanche <i>et al.</i> , 2001 ³⁰	France	560/≥100	560/18–70	11: 83: 6	7: 80: 13	Matched for sex & geographic origin	1.47 (1.10, 1.97) ϵ_2 0.46 (0.34, 0.62) ϵ_4	All centenarians were eligible.
Wang <i>et al.</i> , 2001 ³¹	Uyгур, China	35/≥90	71/20–35	7: 87: 6	7: 80: 13	None	0.93 (0.30, 2.83) ϵ_2 0.39 (0.13, 1.20) ϵ_4	Controls were men, sex of elderly not stated.
Zubenko <i>et al.</i> , 2002 ³²	Pennsylvania, USA	100/≥90	100/18–25	13: 81: 7	7: 82: 12	Matched for sex, ethnicity & location	1.88 (0.95, 3.73) ϵ_2 0.59 (0.30, 1.18) ϵ_4	Elderly patients had no evidence of cognitive impairment.
Choi <i>et al.</i> , 2003 ³³	Korea	103/≥100	6435/30–79	6: 87: 7	6: 84: 9	None	0.96 (0.54, 1.69) ϵ_2 0.72 (0.42, 1.24) ϵ_4	62% of individuals tested had dementia. ϵ_4 frequency was significantly higher among those with dementia than among those without.
Panza <i>et al.</i> , 2003 ³⁴	Southern Italy	52/≥100	72/mean 31.8	12: 86: 3	10: 84: 6	Matched for sex and geographic location	1.17 (0.51, 2.64) ϵ_2 0.45 (0.12, 1.72) ϵ_4	All centenarians were eligible.

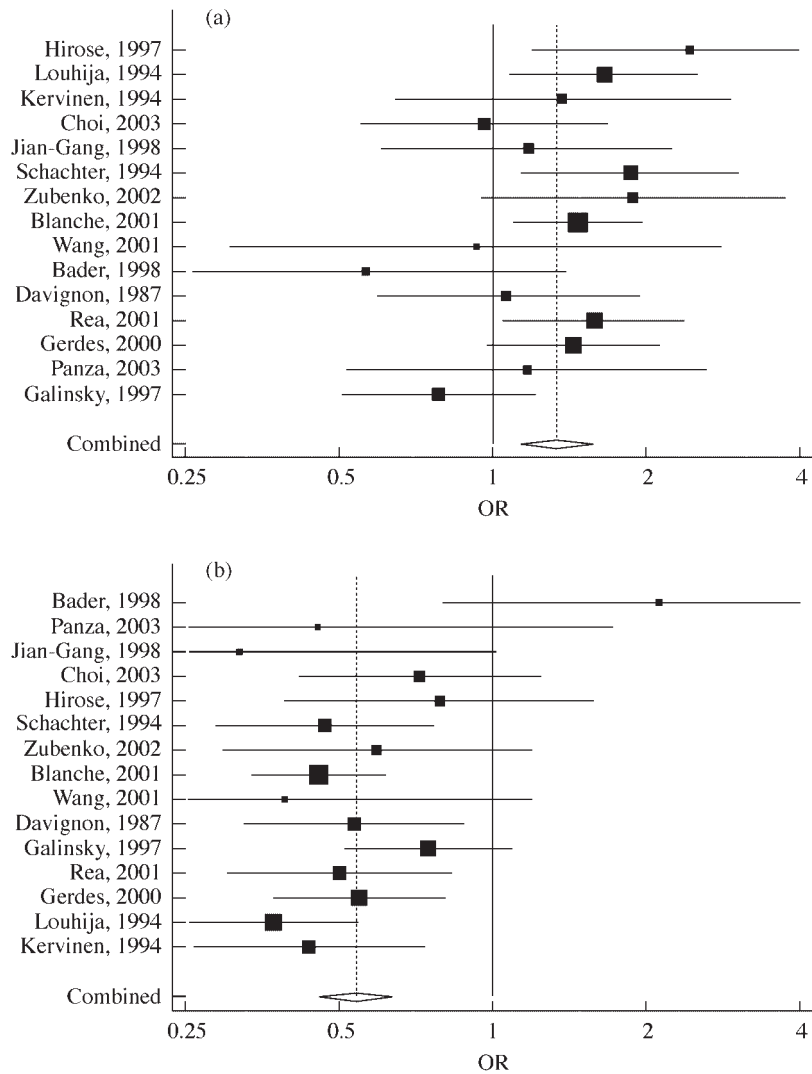


Figure 4 Forest plots representing case-control studies of *APOE* and longevity (ordered by allele frequency in controls)

(a) *APOE* ε2 allele and longevity^a

(b) *APOE* ε4 allele and longevity^a

^a Studies are in order of allele frequency in control population from low (top) to high (bottom)

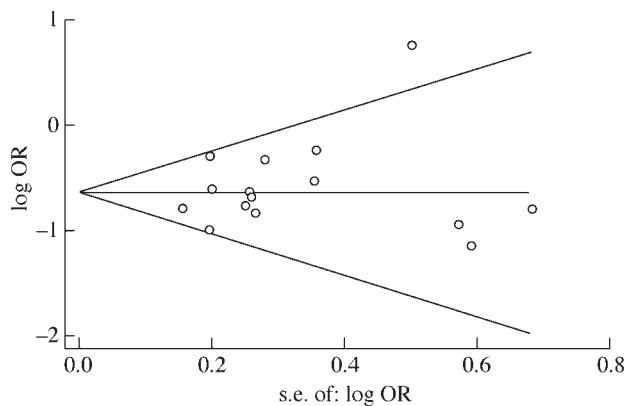


Figure 5 Begg's funnel plot of cross-sectional studies of *APOE* ε4 allele and longevity

and in large university cities,¹⁴ where most of these studies were carried out. Whilst almost all studies showed that there was a deficit of *APOE* ε4 alleles in elderly versus younger populations, heterogeneity in effect size remained, even when populations were stratified by region.

We are able to make inferences based on the biology and epidemiology of *APOE* which suggests that assumption (b) is invalid, and that any longevity effect may be subject to modification over periods of years or decades, corresponding to the age difference between elderly and control groups. *APOE* plays an important role in lipoprotein metabolism, therefore *APOE* genotype is one influence on the health effects of dietary fat intake. *APOE* is associated with CHD risk, and this is one mechanism proposed to explain the association of *APOE* with longevity. In the last 30 years CHD mortality has decreased substantially in many countries (Figure 2). Factors underlying

this decrease include changes in diet, smoking, and treatments for CHD. In the studies reviewed here the size of *APOE* $\epsilon 4$ effect differed by population, providing further evidence that $\epsilon 4$ effects are context dependant.

There appeared to be an association between $\epsilon 4$ allele frequency in controls and the effect of $\epsilon 4$ on longevity. Sampling variation is an important potential explanation. If by chance the estimated $\epsilon 4$ prevalence is above the true prevalence, the former will tend to be high within the distribution and the effect size will tend to be extreme, due to regression to the mean. However, relative *APOE* $\epsilon 4$ allele frequency is known to vary by population and the studies reviewed here are not anomalous in this respect. Among Europeans relative *APOE* $\epsilon 4$ frequencies exhibit a North to South cline.⁴² Accordingly, high relative *APOE* $\epsilon 4$ allele frequencies were observed in studies with Northern European participants (15–21%) and low frequencies in Southern European studies (4–6%). Furthermore, the Bayesian approach, which was used to calculate the association between control allele frequency and effect size, avoids the bias that can arise due to sampling error.

It remains unclear whether sampling a younger control population produces valid estimates of initial allele frequencies in the corresponding elderly population. Migration and population substructure may confound such studies. If relative allele frequencies in control populations overestimate the initial relative allele frequencies in the elderly group a spurious estimate of the longevity effect will be obtained. Migration is likely to change relative allele frequencies substantially when the source and recipient populations have very different allele frequencies. Because changes in allele frequency in the short term are a function of the difference in frequency and migration rate, large differences can be seen in just one generation.⁴³ Net immigration from regions with relatively high allele frequency will overestimate the decrease in $\epsilon 4$ in the elderly relative to controls. Conversely net immigration from areas with relatively low allele frequency will underestimate the difference. For example in the Canadian study²⁰ the relative *APOE* $\epsilon 4$ allele frequency in the control group was 15.2%. If this represents the allele frequency in the elderly population at birth, 0.53 (95% CI: 0.32, 0.88) is a good estimate for the link between *APOE* $\epsilon 4$ and longevity. However, 19% of current Canadian residents were not born in Canada.⁴¹ If migrants came from countries with relatively high allele frequencies of 31% as reported among Lapps,⁴⁴ migration would result in an overestimation of the real initial relative allele frequency 11.5% and the true $\epsilon 4$ effect will be closer to the null 0.76 (95% CI: 0.44, 1.30). Alternatively, if migrants came from countries or areas of low *APOE* $\epsilon 4$ frequency (frequencies of 5% have been reported in Sardinians)⁴⁵ the real initial frequency would have been higher at around 17.6% and the $\epsilon 4$ effect will be around 0.45 (95% CI: 0.27, 0.72). This simple example of international migration occurring within one generation does not take account of second generation migrants born in Canada or internal migration. While it involves somewhat unrealistic assumptions about migration patterns, the example does offer some insight into how migration may influence results from genetic association studies of longevity.

Only 3 out of the 15 studies matched on ethnicity and/or place of birth. The remaining studies assumed a static population and an identical gene pool for elderly and younger

age groups. A meta-analysis including only studies matching on origin or ethnicity, gave a result closer to the null, consistent with an artefactual explanation for the studies failing to take account of genetic heterogeneity. The remaining studies effectively assumed a static population and an identical gene pool for elderly and younger age groups.

A further explanation for the apparent association between $\epsilon 4$ prevalence and magnitude of $\epsilon 4$ effect on longevity is that the impact is greater in populations with a higher underlying risk of CHD. International comparisons suggest that *APOE* $\epsilon 4$ allele frequency may explain some of the geographical variation in CHD risk.⁴⁶ In Japan (a low CHD risk population) the ratio for relative $\epsilon 4$ frequency in elderly versus younger populations is 0.78 (95% CI: 0.39, 1.58),²⁵ whereas in Finland (a high CHD risk population) this ratio was found to be 0.43 (95% CI: 0.26, 0.73)²² and 0.37 (95% CI: 0.25, 0.54).²³ The difference in CHD risk between Japan and Finland could be due to genetic susceptibility or environmental factors, but more likely a combination of both. For example, in Northern Europe high *APOE* $\epsilon 4$ relative allele frequencies have historically been accompanied by high levels of saturated fat intake.

The influence of genetic variation on longevity is *a priori* an important research question. Proponents of genetic epidemiology argue that 'predicting multifactorial disease outcomes without consideration of epigenetic networks is increasingly seen as naïve'.⁴⁷ The implication for methodology is that the genetic dimension should be included in risk factor models of chronic disease, and, by extension, of longevity. The papers reviewed here focus on a single, genetic effect (the main effect of *APOE* on longevity) and represent early attempts to develop a genetic perspective in a field that has been dominated by environmental explanations. For example, we know that environment, in the broad sense, including behavioural factors, is essential in understanding time trends in both CHD rates (Figure 2) and longevity. What is not known is whether single common genetic variations, such as the *APOE* polymorphism, are important factors that add to the explanation provided by conventional environmental risk factors. Studies of *APOE* and longevity reviewed here, which consistently showed an effect of $\epsilon 2$ and $\epsilon 4$ alleles on longevity, suggest that *APOE* may have a part to play. However, analyses taking simultaneous account of genetic and environmental factors are needed to address this question.

Conclusion

Our review leads us to the following conclusions in relation to the validity of assumptions (a) and (b)

Assumption (a) Initial relative allele frequencies in the different age groups are similar

This assumption may not hold when there is considerable variation in allele frequencies by population, and when migration has resulted in changes in population substructure.

Assumption (b) The risk of mortality conferred by genotype does not depend on year of birth.

This assumption may be invalid when mortality risk conferred by a genotype depends on interaction with environmental or behavioural risk factors, and when the pattern of exposure to such risk factors is closely related to year of birth.

Whilst the studies reviewed here do provide some evidence of a link between *APOE* and longevity, the problem is that the size of effect depends on the population in question and probably also the time period that the study spans. The link between the size of *APOE* effect and relative allele frequency seen in this review is likely to be a result of population differences in underlying disease susceptibility or changes in population structure over time and illustrates the impact of violating the above assumptions.

Future case-control type studies could be improved by at least including comprehensive data on birth origins and exposure to potential confounding factors. It is also possible to address concerns about population stratification in case-control studies, by carrying out genome analysis using methods such as genomic control and structured association.⁴⁸

Cohort studies are not subject to the above assumptions, and hence are preferable to a cross-sectional design. Participants can be selected and followed up over a given time period, during

which they will be subject to the same age-related mortality rate. The cohort studies carried out to date, which have looked at *APOE* and longevity, tend to have short follow-up periods (typically 5 years),^{49–56} hence they only provide a snapshot of overall survival. These studies have reported mixed results, with some documenting null findings for *APOE* $\epsilon 4$ and longevity,^{50,51} and others showing negative associations in subgroups only.^{54–56} To address this question adequately cohort studies, ideally birth cohorts, with a long follow-up are required.

Acknowledgements

We would like to thank Dr Meena Kumari and Dr Mark Pearce for their comments on the manuscript, and also Drs Jonathan Sterne and Nicola Welton for their help with statistical analysis using BUGS. This work was funded by a Co-op research grant from the Medical Research Council, UK.

KEY MESSAGES

- Cross-sectional studies of longevity assume that (a) relative allele frequencies are the same in different age cohorts, and (b) risk of mortality conferred by genotypes does not depend on year of birth.
- Population admixture may result in considerable population changes in allele frequency over time, if allele frequencies differ substantially by population and if migration is high.
- Mortality risk conferred by genotypes may differ if the gene in question interacts with environmental factors, which vary by time and place.
- Studies of *apolipoprotein E (APOE)* $\epsilon 4$ and longevity show that effect size estimates are associated with relative allele frequency in control populations, hence accurately predicting the impact of *APOE* on longevity depends on the validity of assumption (a).
- *APOE* is associated with coronary heart disease (CHD) risk, and this is one mechanism proposed to explain the association of *APOE* with longevity. However, CHD mortality has decreased mortality in the last 30 years, largely due to environmental and behavioural factors, hence the effect of *APOE* on longevity is context dependent.

References

- Cournil A, Kirkwood TBL. If you would live long, choose your parents well. *Trends Genet* 2001;**17**:233–35.
- Hirschhorn JN, Lohmueller K, Byrne E, Hirschhorn K. A comprehensive review of genetic association studies. *Genet Med* 2002;**4**:45–61.
- Heijmans BT, Westendorp RGJ, Slagboom PE. Common gene variants, mortality and extreme longevity in humans. *Exp Gerontol* 2000;**35**:865–77.
- De Benedictis G, Tan Q, Jeune B *et al*. Recent advances in human gene-longevity association studies. *Mech Ageing Dev* 2001;**122**: 909–20.
- Vijg J, van Orsouw N. Searching for genetic determinants of human aging and longevity: opportunities and challenges. *Mech Ageing Dev* 2002;**123**:195–205.
- Schachter F, Cohen D, Kirkwood T. Prospects for the genetics of human longevity. *Hum Genet* 1993;**91**:519–26.
- Wilson PWF, Schaefer EJ, Larson MG, Ordovas JM. Apolipoprotein E alleles and risk of coronary disease, a meta-analysis. *Arterioscler Thromb Vasc Biol* 1996;**16**:1250–55.
- Rubinsztein DC, Easton DF. Apolipoprotein E genetic variation and Alzheimer's disease. a meta-analysis. *Dement Geriatr Cogn Disord* 1999;**10**:199–209.
- Ioannidis JPA, Ntzani EE, Trikalinos TA, Contopoulos-Ioannidis DG. Replication validity of genetic-association studies. *Nat Genet* 2001;**29**:306–09.
- Cardon LR, Bell JI. Association study designs for complex diseases. *Nat Rev Genet* 2001;**2**:91–99.
- Colhoun HM, McKeigue, Davey Smith G. Problems of reporting genetic associations with complex outcomes. *Lancet* 2003;**361**:1–8.
- Cardon LR, Palmer LJ. Population stratification and spurious allelic association. *Lancet* 2003;**361**:598–604.
- Census 2001*. The Stationery Office, ONS, London, 2001.
- Dobson J, McLaughlan G. International migration to and from the United Kingdom 1975–1999: consistency, change and implications for the labour market. *Popul Trends* 2001;**106**:29–38.
- Mascie-Taylor CG, Lasker GW. Effect of interregional migration on geographic variability in biological and social trends in Britain. *Hum Biol* 1995;**67**:629–40.
- Office for National Statistics. *The Health of Adult Britain 1841–1994*. The Stationery Office, ONS, London 1998.
- Davey Smith G, Marmot MG. Trends in mortality in Britain: 1920–1986. *Ann Nutr Metab* 1991;**35**:53–63.
- Levi F, Lucchini F, Negri E, La Vecchia C. Trends in mortality from cardiovascular and cerebrovascular diseases in Europe and other areas of the world. *Heart* 2002;**88**:119–124.

- ¹⁹ Smith JD. Apolipoproteins and aging: emerging mechanisms. *Ageing Res Rev* 2002;**1**:345–65.
- ²⁰ Davignon J, Bouthillier S, Nestruck C, Sing CF. Apolipoprotein E polymorphism and atherosclerosis: Insight from a study in octogenarians. *Trans Am Clin Climatol Assoc* 1987;**99**:100–10.
- ²¹ Schächter F, Faure-Delanef L, Guénot F *et al*. Genetic associations with human longevity at the APOE and ACE loci. *Nature Genet* 1994;**8**:29–32.
- ²² Kervinen K, Savolainen MJ, Salokannel J *et al*. Apolipoprotein E and B polymorphisms—Longevity factors assessed in nonagenarians. *Atherosclerosis* 1994;**126**:89–95.
- ²³ Louhija J, Miettinen K, Kontula K, Tikkanen MJ, Miettinen TA, Tivlis RS. Aging and genetic variation of plasma apolipoproteins. Relative loss of the apolipoprotein E4 phenotype in Centenarians. *Arterioscler Thromb* 1994;**14**:1084–89.
- ²⁴ Galinsky D, Tysoe C, Brayne CE *et al*. Atherosclerosis. Analysis of the apo E/ apo C-I, angiotensin converting enzyme and methylenetetrahydrofolate reductase genes as candidates affecting human longevity. *Atherosclerosis* 1997;**129**:177–83.
- ²⁵ Hirose N, Homma S, Arai Y *et al*. Tokyo Centenarian Study. 4. Apolipoprotein E phenotype in Japanese centenarians living in the Tokyo metropolitan area. *Nippon Ronen Igakkai Zasshi* 1997;**4**:267–72.
- ²⁶ Bader G, Zuliani G, Kostner GM, Fellin R. Apolipoprotein E polymorphism is not associated with longevity or disability in a sample of Italian octo- and nonagenarians. *Gerontology* 1998;**44**:293–99.
- ²⁷ Jian-Gang Z, Yong-Xing M, Chuan-Fu W *et al*. Apolipoprotein E and longevity among Han Chinese population. *Mech Ageing Dev* 1998;**104**:159–67.
- ²⁸ Gerdes LU, Jeune B, Ranberg KA, Nybo H, Vaupel JW. Estimation of apolipoprotein E genotype-specific relative mortality risks from the distribution of genotypes in centenarians and middle-aged men: Apolipoprotein E gene is a 'frailty gene', not a 'longevity gene'. *Genet Epidemiol* 2000;**19**:202–10.
- ²⁹ Rea IM, McDowell I, McMaster D, Smye M, Stout R, Evans A. Apolipoprotein E alleles in nonagenarian subjects in the Belfast elderly longitudinal free-living ageing study (BELFAST). *Mech Ageing Dev* 2001;**122**:1367–72.
- ³⁰ Blanche H, Cabanne L, Sahbatou M, Thomas G. A study of French Centenarians: are ACE and APOE associated with longevity? *C R Acad Sci III* 2001;**324**:129–35.
- ³¹ Wang X, Wang G, Yang C, Xuan L. Apolipoprotein E gene polymorphism and its association with human longevity in Uygar nationality in Xinjiang. *Chin Med J* 2001;**114**:817–20.
- ³² Zubenko GS, Stiffler S, Hughes HB, Fatigati MJ, Zubenko WN. Genome survey for loci that influence successful aging: sample characterization, method validation, and initial results for the Y chromosome. *Am J Geriatr Psychiatry* 2002;**10**:619–30.
- ³³ Choi YH, Kim JH, Kim DK *et al*. Distributions of ACE and APOE Polymorphisms and their relations with dementia status in Korean centenarians. *J Gerontol A Biol Sci Med Sci* 2003;**58**:227–31.
- ³⁴ Panza F, Solfrizzi V, Colacicco AM *et al*. Apolipoprotein E (APOE) polymorphism influences serum APOE levels in Alzheimer's disease patients and Centenarians. *Neuroreport* 2003;**14**:605–08.
- ³⁵ Schiele F, De Bacquer D, Vincent-Viry M *et al*. Apolipoprotein E serum concentration and polymorphism in six European countries: The ApoEurope project. *Atherosclerosis* 2000;**152**:475–88.
- ³⁶ Frikke-Schmidt R, Tybjaerg-Hansen A, Steffensen R, Jensen G, Nordestgaard BG. Apolipoprotein E genotype: epsilon 32 women are protected while epsilon 43 and epsilon 44 men are susceptible to ischemic heart disease. The Copenhagen city heart study. *J Am Coll Cardiol* 2000;**35**:1192–99.
- ³⁷ Thomson SG, Smith TC, Sharp SJ. Investigating underlying risk as a source of heterogeneity in meta-analysis. *Statist Med* 1997;**16**:2741–58.
- ³⁸ Begg CB, Mazumdar M. Operating characteristics of a rank correlation test for publication bias. *Biometrics* 1994;**50**:1088–101.
- ³⁹ Egger M, Davey Smith G, Schneider M, Minder C. Bias in meta-analysis detected by a simple, graphical test. *BMJ* 1997;**315**:629–34.
- ⁴⁰ Corbo RM, Scacchi R. Apolipoprotein E (APOE) allele distribution in the world. Is APOE*4 a 'thrifty' allele? *Ann Hum Genet* 1999;**63**:301–10.
- ⁴¹ United Nations Population Division, Department of Economic and Social Affairs. *International Migration 2002*. United Nations Publication 2002.
- ⁴² Haddy N, De Bacquer D, Chemaly MM *et al*. The importance of plasma apolipoprotein E concentration in addition to its common polymorphism on inter-individual variation in lipid levels: results from Apo Europe. *Eur J Hum Genet* 2002;**10**:841–50.
- ⁴³ Wang J, Whitlock MC. Estimating effective population size and migration rates from genetic samples over space and time. *Genetics* 2003;**163**:429–46.
- ⁴⁴ Lehtinen S, Luoma P, Lehtimäki T, Nayha S, Hassi J, Nikkari T. Differences in genetic variation of apolipoprotein in Lapps and Finns. *Atherosclerosis* 1994;**109**:263.
- ⁴⁵ Corbo RM, Scacchi R, Mureddu L, Mulas G, Alfano G. Apolipoprotein E polymorphism in Italy investigated in native plasma by a simple polyacrylamide gel isoelectric focusing technique. Comparison with frequency data of other European populations. *Ann Hum Genet* 1995;**59**:197–209.
- ⁴⁶ Stengård JH, Weiss KM, Sing CF. An ecological study of association between coronary heart disease mortality rates in men and the relative frequencies of common allelic variations in the gene coding for apolipoprotein E. *Hum Genet* 1998;**103**:234–41.
- ⁴⁷ Sing CF, Stengård JH, Kardia SLR. Genes, environment and cardiovascular disease. *Arterioscler Thromb Vasc Biol* 2003;**23**:1190–96.
- ⁴⁸ Ewbank DC. Mortality differences by APOE genotype estimated by APOE genotype estimated from demographic synthesis. *Genet Epidemiol* 2002;**22**:146–55.
- ⁴⁹ Tilvis RS, Strandberg TE, Juva K. Apolipoprotein E phenotypes, dementia and mortality in a prospective population sample. *J Am Geriatr Soc* 1998;**46**:712–15.
- ⁵⁰ Slooter AJC. Apolipoprotein E and longevity: the Rotterdam study. *J Am Geriatr Soc* 2001;**49**:1258–59.
- ⁵¹ Louhija J, Aguero-Torres H, Tilvis R. Survival in the Finnish centenarians in relation to apolipoprotein E polymorphism. *J Am Geriatr Soc* 2001;**49**:1007–08.
- ⁵² Stengard JH, Zerba KE, Pekkanen J, Ehnholm C, Nissinen A, Sing CF. Apolipoprotein E polymorphism predicts death from coronary heart disease in a longitudinal study of elderly Finnish men. *Circulation* 1995;**91**:265–69.
- ⁵³ Stengard JH, Pekkanen J, Ehnholm C, Nissinen A, Sing CF. Genotypes with the apolipoprotein epsilon 4 allele are predictors of coronary heart disease mortality in a longitudinal study of elderly Finnish men. *Hum Genet* 1996;**97**:677–84.
- ⁵⁴ Corder EH, Lannfelt L, Vitanen M *et al*. Apolipoprotein E genotype determines survival in the oldest old (85 years or older) who have good cognition. *Arch Neurol* 1996;**53**:418–22.
- ⁵⁵ Lee JH, Tang MX, Schupf N *et al*. Mortality and apolipoprotein E in Hispanic, African-American and Caucasian Elders. *Am J Med Genet* 2001;**103**:121–27.
- ⁵⁶ Vogt MT, Cauley JA, Kuller LH. Apolipoprotein E phenotype, arterial disease and mortality among older women: the study of osteoporotic fractures. *Genet Epidemiol* 1997;**14**:147–56.