

From Alzheimer's Disease to a Demography of Chronic Disease: The Development of Demographic Synthesis for Fitting Multistate Models

DOUGLAS EW BANK

This is the story of a process of discovery that led from a study of Alzheimer's disease to demographic models that incorporate genetics. This in turn led to a generalized demographic approach to estimating multistate models. This evolution is documented in four papers (Ewbank 1999, 2002a, 2002b, 2004). The first uses a basic demographic model to examine the claim that Alzheimer's disease is the fourth leading cause of death in the United States. The second expands this model by incorporating differences in the incidence of Alzheimer's disease by genotype. The next paper applies the demographic model incorporating genotype to all-cause mortality. And the fourth applies the results on mortality by genotype to examine the effects of one gene on the variation in mortality within and among populations.

The discussion of each paper, presented below, is prefaced by a description of how the research question originated and of previous research on the topic. This is followed by a brief description of the logic underlying the demographic model and the data used to answer the question. This description highlights a few elements of the models; they are described in full in the published papers. The main findings that relate to the development of the models are summarized and new insights are described. I make no claim for the originality of any of these ideas. They were merely ideas that were new to me or captured my attention and contributed to the research at hand.

Deaths attributable to Alzheimer's disease in the United States

In early 1992, Alzheimer's disease was labeled the fourth leading cause of death in the United States. This claim originated in an editorial in the *New*

England Journal of Medicine by Robert Katzman (1986). The method used to derive the conclusion was not described in detail; however, it was basically a back-of-the-envelope calculation of attributable risk. Although this claim was accepted by many Alzheimer's disease researchers, demographers were generally skeptical.

Alzheimer's disease is the most common cause of severe memory loss in the elderly. It is a chronic disease that causes a decline over a period of years in short-term memory, language, ability to recognize friends and family members, and other basic cognitive abilities. In 1991 Alzheimer's disease was reported as the underlying cause of death for 14,112 deaths in the United States, 13,768 of which involved people over age 65 years (Hoyert 1996). According to these data, Alzheimer's disease would rank as the eleventh most common cause of death among those over age 65—not the fourth leading cause at all ages. Including all deaths for which Alzheimer's disease was mentioned as a contributing cause more than doubles the reported number. Between 1979 and 1991, the reported age-adjusted death rate for Alzheimer's disease increased twelve-fold. Most of this increase was due to increased awareness of Alzheimer's disease and decreasing use of less specific diagnoses for dementia.

Numerous studies have documented excess mortality rates among patients with Alzheimer's disease (Barclay et al. 1985; Evans et al. 1991; Kukull et al. 1994; Larson et al. 2004). It is not clear, however, why this should be the case, at least until the last stages of the disease. It is clear from several studies that Alzheimer's disease is underreported on death certificates in the United States (Macera et al. 1992; Raiford et al. 1994). For example, a study in Rochester, Minnesota, of Alzheimer's disease cases diagnosed between 1960 and 1984 reported that only 11 percent of death certificates mentioned dementia (Beard et al. 1996).

Methods¹

To estimate deaths from Alzheimer's disease I followed Katzman's use of an attributed risk model. I produced two independent estimates by examining two sets of data using a different method for each set. The first estimate was based on published data from the East Boston Study, which was a large population-based study of the incidence and prevalence of Alzheimer's disease and the excess mortality associated with it (Evans et al. 1989; Evans et al. 1991). The second applied a demographic model and data from several different kinds of studies. It is this second approach that has led to much further research.

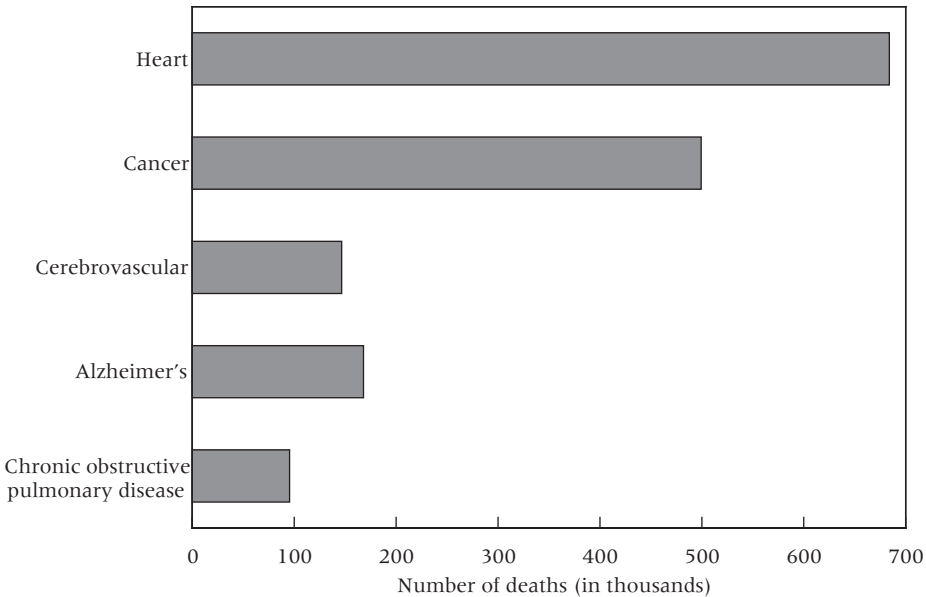
The demographic model underlying the second estimate is a relatively simple multistate model. It looks like a life table with additional columns to keep track of individuals by their disease status. Individuals at each age are divided into those who do not have Alzheimer's disease, $S_N(x)$, and those who do $S_A(x,d)$. The second index variable for Alzheimer's disease cases, d ,

tracks duration of disease. The incidence of disease is modeled using a Weibull distribution with unobserved heterogeneity (Manton, Stallard, and Vaupel 1986). This formulation allows the incidence rate to increase rapidly with age, but can allow the rates to level off (or even decline) at the oldest ages. It summarizes the age-specific incidence rates with three parameters, which are estimated using maximum likelihood as described below. Mortality rates for the model are based on a life table for the United States in 1990 and several published clinical studies of the relative risk of death for Alzheimer’s disease cases by duration of disease.

Findings

Although the two estimates were based on different models and different data, the results were surprisingly similar. The East Boston data led to an estimate of 173,000 deaths attributable to Alzheimer’s disease in 1995. The simulation led to an estimate of 163,000. These estimates are about four times greater than the number of death certificates that mentioned Alzheimer’s disease in 1991 (Hoyert 1996). It now appears that Alzheimer’s disease was responsible for about 7 percent of all deaths in 1995. Figure 1 compares the number of Alzheimer’s disease deaths with the other major causes of death in the United States in 1995. These estimates place Alzheimer’s disease on a

FIGURE 1 Estimated number of deaths from Alzheimer’s disease and reported numbers of deaths from heart disease, cancer, cerebrovascular disease, and chronic obstructive pulmonary disease, United States, 1995



SOURCES: Ewbank 1999; Rosenberg et al. 1996.

par with cerebrovascular diseases as the third leading cause of death in the United States.

Insights

The approach to estimating the parameters of the Weibull function for the incidence of Alzheimer's disease became the foundation for subsequent research described in this chapter using maximum likelihood to estimate the parameters of the multistate model from a variety of types of data.

The details of the Weibull function with heterogeneity are not important. The only important fact is that there are three parameters: the mean of the gamma distribution of risk, α (the level parameter); the parameter that determines the rate at which the risk increases with age, m ; and the coefficient of variation in the risk, which slows the rate of increase at the oldest ages, v . The task is to estimate the values of these three parameters for the incidence of Alzheimer's disease.

The maximum likelihood approach to estimation involves finding the values of the model parameters that maximize the probability of observing the recorded data. I started with published incidence rates from a longitudinal study of Alzheimer's disease. Incidence studies involve a baseline survey that tests everyone for Alzheimer's disease. A follow-up survey several years later picks up new cases of the disease among the survivors. The probability of observing $C(x)$ cases at age x among $N(x)$ individuals at risk is binomial with a rate of $\lambda(x|\alpha, m, v)$. The likelihood of observing the data from a single study for a given age-sex group is the probability of observing $C(x)$ and $N(x)$ given values of α , m , and v :

$$L(C_x, N_x | \alpha, m, v) = \lambda(x|\alpha, m, v)^{C_x} [1 - \lambda(x|\alpha, m, v)]^{(N_x - C_x)}.$$

Thinking of the model as a multistate life table, we can express $\lambda(x)$ in terms of the number of Alzheimer's disease cases at the appropriate age and various durations, $S_A(x, d)$, and the number who do not have Alzheimer's, $S_N(x)$. In a multiround survey, new cases recorded at the second round are used to estimate "incidence rates." However, these rates do not include new cases among individuals who died between survey rounds. Although this introduces little error, it is possible to match these data with model estimates. For example, if a follow-up survey at three years provided data by single years of age, the estimated rate at age x from the model is:

$$\lambda(x) = \frac{S_A(x, 0) + S_A(x, 1) + S_A(x, 2)}{S_A(x, 0) + S_A(x, 1) + S_A(x, 2) + S_N(x)}.$$

In other words, the observed rate is the number of survivors with new cases within the last three years (i.e., those at duration 0, 1, or 2 years), divided by the number of individuals at risk (i.e., the new cases plus those remaining without Alzheimer’s disease). Of course $S_N(x)$ and $S_A(x,d)$ are functions of the values assumed for the three parameters.²

The likelihood of observing all of the data is the product of the binomial probabilities for individual age groups from various studies:

$$L = L_{x,i}L_{y,i} \dots L_{z,j} \dots$$

where $L_{x,i}$ is the likelihood of observing the incidence rate reported at age x in study i . Then a simple search routine (available in spreadsheet programs) can find the values of the three parameters of the Weibull that maximize the total likelihood, L .

At the time I was doing this research there were very few data on the incidence of Alzheimer’s disease, although I found far more data on the prevalence of disease. The model also produces estimates of prevalence since it combines incidence rates with estimates of excess mortality among cases. Therefore, an alternative approach is to use prevalence data. This is possible because the incidence of Alzheimer’s disease has probably been relatively constant in the recent past.

The prevalence rate is also binomial where the model estimate of the probability is the ratio of the number of cases between ages x and $x+n$ to the total population at those ages:

$$\frac{\int_x^{x+n} S_A(z)dz}{\int_x^{x+n} S(z)dz}$$

The binomial probabilities for each age-specific rate in each of several studies would then be multiplied together to get the overall likelihood.

At some point it occurred to me that I did not need to limit the estimation to either incidence rates or prevalence rates. By simply multiplying together the likelihood values from the incidence and prevalence data, it is possible to use data from both types of studies. Therefore, I estimated the Weibull parameters from five prevalence studies and one incidence study.³

This approach to estimating the parameters formed the basis for what I later termed “demographic synthesis.” It is defined as the use of a multistate model to combine different types of functionally related data to produce estimates of quantities not adequately measured in any single study.

The model serves as a meta-analytic tool which is not limited to combining studies that all have the same research design and produce the same indexes. This is a significant improvement over standard meta-analytic meth-

ods that can only use a single type of data (e.g., incidence rates or prevalence rates). For example, Jorm and Jolley (1998) studied the incidence of dementia by combining data from numerous longitudinal studies, but they did not make use of the prevalence estimates that come from the baseline surveys.

This approach can be generalized to combine a wider variety of data. In the next paper I combined rates from longitudinal studies (incidence rates or mortality rates) with cross-sectional data (prevalence rates or genotype frequencies) and odds ratios from case-control studies.

A multistate model of the genetic risk of Alzheimer's disease

The attributable risk approach might overstate mortality from Alzheimer's disease if patients are much more likely than those without Alzheimer's disease to have another major risk factor for mortality. The next research question was whether the deaths attributed to Alzheimer's disease are really a result of another risk factor that is more common in Alzheimer's disease patients.

There are few well-established risk factors for Alzheimer's disease. One that seemed promising in this regard is the alleles of the gene for apolipoprotein-E (APOE). APOE is involved in the transport of fats (lipids) in the blood stream. Three variants of APOE have been shown to be associated with different serum levels of lipids. The most common variant is labeled $\epsilon 3$. The second most common is $\epsilon 4$, which is a risk factor for both Alzheimer's disease (Farrer et al. 1997) and ischemic heart disease (Eichner et al. 1993; Wilson et al. 1996). The $\epsilon 2$ variant is protective against Alzheimer's disease and may protect against heart disease. The three variants (alleles) of the gene that gives the code for APOE are termed $\epsilon 2$, $\epsilon 3$, and $\epsilon 4$.⁴ These alleles differ by single nucleic acids at positions 112 and 158: APOE $\epsilon 4$ has arginine at both sites, $\epsilon 3$ has cysteine at 112, and $\epsilon 2$ has cysteine at both sites. An individual's genotype is determined by the copies of the APOE alleles inherited from his or her parents. The most common genotype is $\epsilon 3/3$. The initial research question was whether many of the deaths attributed to Alzheimer's disease are really deaths from ischemic heart disease associated with the $\epsilon 4$ allele.

Methods⁵

Answering this question required several elaborations of the simple model of mortality and Alzheimer's disease. First, I had to expand to three causes of death: Alzheimer's disease, ischemic heart disease, and all others. Then I divided the population into groups by genotype. Each genotype would have different risks of Alzheimer's disease and ischemic heart disease. I built this model and tested it. However, I dropped this approach for two reasons. First,

the model suggested that the APOE–heart disease link was not responsible for many of the excess deaths among individuals with Alzheimer’s disease. Second, there were scant data on APOE and ischemic heart disease mortality, although there are plentiful data on APOE and heart disease *morbidity*. Therefore, I dropped back to a simpler model with only two causes of death (Alzheimer’s disease and all others) to study differences in the risk of Alzheimer’s disease by APOE genotype. After presenting the model, I applied this approach (2002a) to published data on males in populations of European origin.

Although the association of the APOE alleles with the risk of Alzheimer’s disease was widely accepted, the estimates of the risks associated with each genotype varied widely from study to study, largely because of sampling variability. In addition, no single study was large enough to produce useful estimates of the risks associated with the rarest genotypes: e2/2 and e2/4. Farrer et al. (1997) published an analysis of data on APOE and Alzheimer’s disease from 40 research teams. The sample size of 5,930 patients and 8,607 controls was large enough to examine the risks for e2/2 and e2/4 in addition to examining how the risks changed with age. The analysis relied on a logistic regression that included age and age-squared terms that differed by APOE genotype. Thus the odds ratios for the risk of Alzheimer’s disease by genotype were allowed to change with age. The problem with this approach is that it requires any changes with age in the odds ratio to be symmetrical. The mechanisms that might lead to changes with age, however, are very different at the youngest and oldest ages (Ewbank 1998).

I used an approach that models the changes in the odds ratios at the youngest and oldest ages separately. The basic model follows the model used to estimate the deaths attributable to Alzheimer’s disease. It assumes that the risk of Alzheimer’s disease for each genotype follows a Weibull distribution with the same values of m (the exponent on age) and ν (the amount of heterogeneity) for each genotype. The differences in risk among genotypes are modeled by different values for α . After combining the rare e2/2 genotype with the e2/3 and the e2/4 with the e3/3, there are six main parameters to be estimated: the level of Alzheimer’s disease risk among the e3/3 at age 80 (which sets α for this reference group), the relative risks for the e2/3, e3/4, and e4/4 (which define the differences in α by genotype), and the values of m and ν shared by all four genotypes.

This basic model was modified to allow the odds ratios at the youngest ages to reflect changes in the relative importance of rare genetic forms of Alzheimer’s disease (Levy-Lahad and Bird 1996). Several genotypes have been associated with Alzheimer’s disease at ages as young as 35. There is some evidence that the risks associated with these rare genotypes are not modified by APOE genotype. Therefore, at the youngest ages APOE genotype has little or no effect on the risk of Alzheimer’s disease. As the importance of these rare genotypes declines with age, the effect of APOE becomes apparent.

Extending the model to include genotype was not difficult. Estimating the parameters presented a challenge, however. There are almost no data on the incidence of Alzheimer's disease by genotype and few data on prevalence by genotype. The problem is especially serious for the rarer genotypes. Therefore, we have little or no data from population-based surveys with which to estimate the relative risks.

The solution is to use clinical case-control data on the relative differences in risk among genotypes. The odds ratios from these studies are closely related to the relative risk parameters we need to estimate. Comparable odds ratios can easily be calculated from the demographic model. For example, the odds ratio for $\epsilon 3/4$ relative to $\epsilon 3/3$ is given by the standard equation for an odds ratio:

$$OR_{3/4} = \frac{P_{3/4}(x) / [1 - P_{3/4}(x)]}{P_{3/3}(x) / [1 - P_{3/3}(x)]}$$

where $p_{3/4}(x)$ is the proportion with Alzheimer's disease among the observed $\epsilon 3/4$:

$$P_{3/4}(x) = \frac{S_A^{3/4}(x)}{S_A^{3/4}(x) + S_N^{3/4}(x)}.$$

Similar odds ratios can be calculated using incidence rates to match clinical studies of the onset of Alzheimer's disease.

The model was fitted to published data on the incidence and prevalence of Alzheimer's disease and relative risks of Alzheimer's disease by APOE genotype.

Findings

The first finding was that the APOE connection with ischemic heart disease does not explain much of the excess risk previously attributed to Alzheimer's disease. Because I dropped ischemic heart disease mortality from the model, my published findings relate to the age pattern of incidence of Alzheimer's disease and differences in incidence among APOE genotypes.

Among men aged 80, the relative risks of Alzheimer's disease relative to $\epsilon 3/3$ are: $R_{3/4}$: 3.4 (95 percent CI: 2.5–4.4), $R_{4/4}$: 9.4 (3.8–26.6), and $R_{2/3}$: 0.43 (0.24–0.72). The differences between genotypes are slightly larger at age 65, change rapidly after age 85, and essentially disappear by age 100. The heterogeneity model assumes that this declining importance of APOE holds only for the average survivor. For example, if you could change a

centenarian's genotype from $e3/4$ to $e3/3$, you would reduce his risk of Alzheimer's disease by a factor of about 3.4. However, centenarians with the $e3/4$ genotype who have not developed Alzheimer's disease probably have numerous unknown characteristics that protect them against the disease. It is only because of these other protective characteristics that centenarians with the $e3/4$ genotype do not on average have higher risks than those with the $e3/3$ genotype.

These effects of heterogeneity are probably responsible for much of the leveling off of the incidence levels at the oldest ages. Some of this is caused by the fact that the riskier genotypes ($e3/4$ and $e4/4$) form a smaller proportion of the population at risk at the older ages. Similarly, unobserved heterogeneity (the ν parameter) causes the incidence rate among the riskier genotypes to level off. This is most evident in the incidence rates for the $e4/4$ genotype (see below). The model also estimated that about 0.20 percent of the population have rare genotypes that are associated with a risk of Alzheimer's disease of about 4 percent per year.

Insights

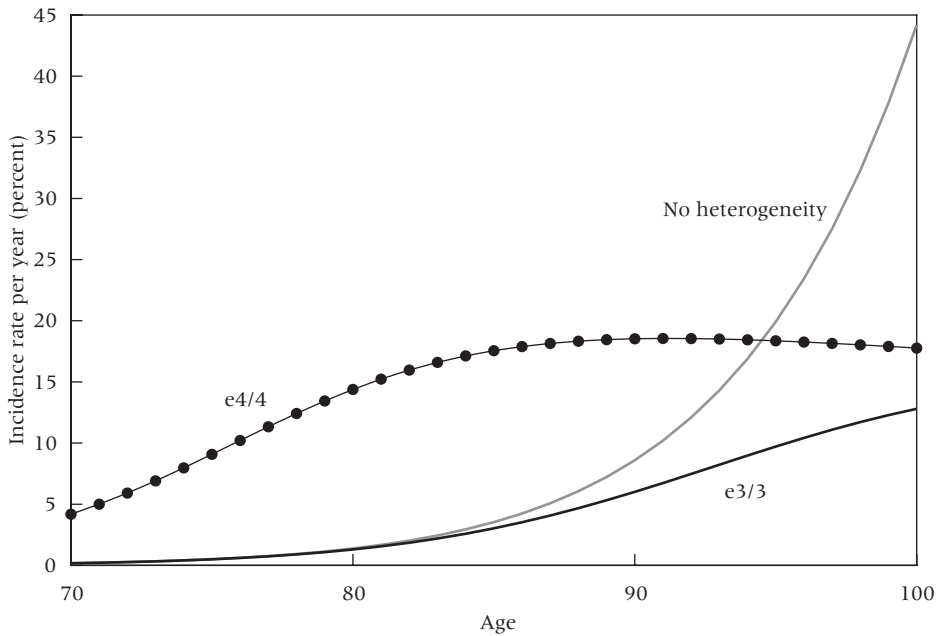
The first insight is that we can combine estimates of the odds ratios from case-control studies with incidence and prevalence rates from population-based surveys to estimate incidence rates by age, sex, and genotype. It is not feasible to estimate these relative risks from a single large longitudinal study.

The second insight relates to the estimate of ν , the coefficient of variation of unobserved heterogeneity. Unobserved heterogeneity arises because we cannot control for every factor affecting the risk of Alzheimer's disease. The effects of unobserved heterogeneity are generally described in terms of the age pattern of mortality. The solid line in Figure 2 shows the estimated incidence of Alzheimer's disease at each age among males with APOE $e3/3$. The gray line shows what the rates would be for this group with the same level of risk in the Weibull distribution (α) and the same rate of increase in the absence of heterogeneity (m) but without any heterogeneity (ν set to 0).

Figure 2 also shows the estimated incidence rates for the $e4/4$ genotype. Heterogeneity affects the $e3/3$ and the $e4/4$ genotypes differently even though they are based on the same amount of heterogeneity (i.e., the same value of ν). The reason is that heterogeneity selects out the most frail faster when the incidence rates are high; the combination of the high risks from $e4/4$ and other unobserved risk factors is quite powerful. The incidence rates for the $e4/4$ level off by about age 90, but the rates for the $e3/3$ are still climbing. Therefore, the relative risks (the ratio of the values of the two curves at the same age) decline at the oldest ages (Figure 3).

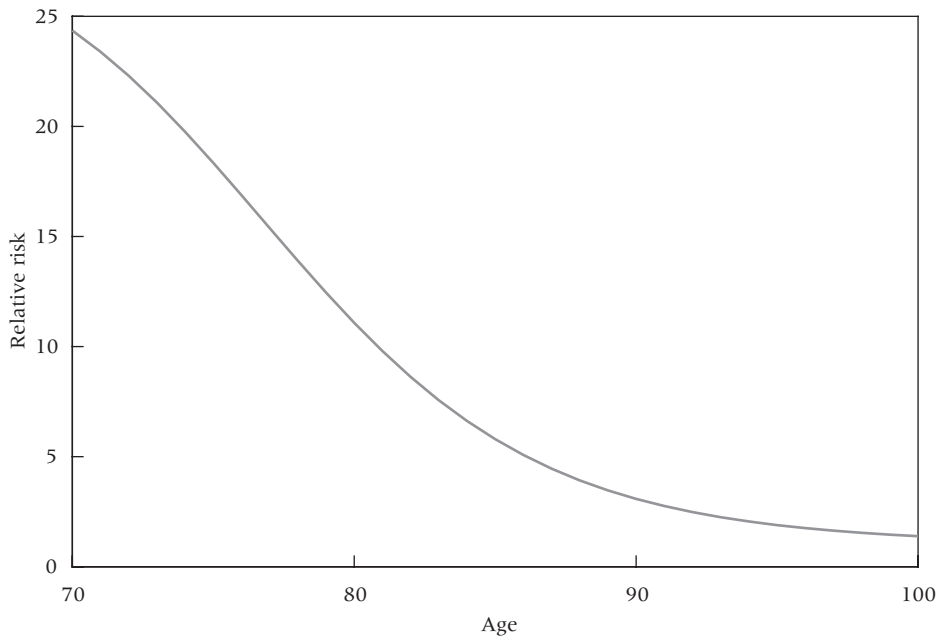
This leads to the second insight: changes in the relative risks with age provide information about the amount of heterogeneity. This is very important because the confidence intervals for the reported incidence rates

FIGURE 2 Estimated incidence rates for Alzheimer's disease, APOE genotypes e3/3 and e4/4 and the curve for e3/3 without unobserved heterogeneity



SOURCE: Ewbank 2002a.

FIGURE 3 Estimated risk of Alzheimer's disease, APOE e4/4 genotype relative to the e3/3 genotype



SOURCE: Based on parameter estimates in Ewbank 2002a.

become very large at the oldest ages and do not provide much information about heterogeneity. The confidence intervals on the relative risks at the oldest ages are also large. However, the combination of numerous studies of each kind can lead to more precise estimates of heterogeneity.

Mortality differentials by APOE genotype

There are good reasons to believe that the risk of overall mortality differs substantially by APOE genotype since it affects the risks of two major causes of death: ischemic heart disease (IHD) and Alzheimer's disease. The previous paper modeled the differentials in mortality associated with Alzheimer's disease, but there were not enough data relating APOE and IHD mortality. Therefore, the mortality differences by genotype were understated. Thus, my next paper (2002b) addressed two questions: 1) how much do all-cause mortality rates differ by APOE genotype, and 2) do the effects of APOE genotype on mortality differ by sex or area of residence?

A number of longitudinal studies have examined mortality rates by APOE genotype. Contrary to expectations, not all of them have found significant differences. For example, three of four prospective studies did not find a significant difference in mortality by APOE genotype (Tilvis et al. 1998; Vogt et al. 1997; Skoog et al. 1998; Juva et al. 2000). All four failed to find (or did not test for) differences associated with the $\epsilon 2$ alleles. Traditional meta-analytic methods are not well suited to this issue. A logistic regression with proper controls for differences in sample size and length of observation could address this issue with appropriate controls for an interaction between the gene effect and age. It is not clear, however, what form that interaction should take.⁶ In addition, there are no standard age groups used in published studies, which makes it difficult to properly control for age differences using logistic regression.

Cross-sectional studies of genotype frequencies by age provide additional information on mortality by genotype. A lower frequency of a genotype at the oldest ages suggests that individuals with that genotype died out at a faster rate. Traditional meta-analytic approaches are not capable of combining cross-sectional and longitudinal data. Genotype frequencies by age have been used by Yashin et al. (2000) and Toupance et al. (1998) to estimate excess mortality from cross-sectional surveys, but they did not include data from longitudinal studies. I used demographic synthesis to combine these two types of data from five European countries and American whites.

Methods⁷

The estimation included longitudinal data from six studies that followed almost 6,000 individuals for an average of 5.6 years of observation.⁸ The genotype frequencies come from 12 cross-sectional studies of APOE geno-

type frequencies by age. These studies include data from young or middle-aged adults and nonagenarians or centenarians. The total sample size for cross-sectional data is 7,264 (including the baseline surveys from the longitudinal studies).

The model used to estimate mortality by APOE genotype is a simplified version of the model of Alzheimer's disease incidence and prevalence. It does not specify causes of death and there is no onset of disease.

An important new element in this study was the use of cohort life tables. The estimation used 56 cohort life tables for different countries by sex and year of birth. The model forced the mortality rates by genotype to add up to the rates in the life table for an appropriate cohort. This was done using the equation:

$$\mu(x) = \frac{R_{2/3}(x)\mu_{3/3}(x)S_{2/3}(x) + \mu_{3/3}(x)S_{3/3}(x) + R_{3/4}(x)\mu_{3/3}(x)S_{3/4}(x) + \dots}{S_{2/3}(x) + S_{3/3}(x) + S_{3/4}(x) + \dots}$$

which merely states that $\mu(x)$, the reported mortality rate for the cohort at age x , is equal to the weighted average of the mortality rates for each genotype. The weights are the proportion of each genotype that survived to age x , for example $S_{2/3}(x)$. It is easy to solve this equation to derive $\mu_{3/3}(x)$ given values of $S(x)$ and the $R(x)$.

$R(x)$ can be derived from equations like the following:

$$R_{3/4}(x) = R_{3/4}(0) \left(\frac{S_{3/4}(x)}{S_{3/3}(x)} \right)^{\nu^2}$$

which relates the relative risk at every age to the rate at birth if the variation in risk follows a Gamma distribution. This can be rewritten to use any age as the reference rate in place of age 0. I used age 60 in order to get confidence intervals at a useful age. By using a cohort life table and parameterizing the $R(x)$, we do not have to impose a functional form (for example, a Gompertz curve) on the age-specific mortality rates. This removes one assumption and two parameters that are usually necessary for this type of analysis.

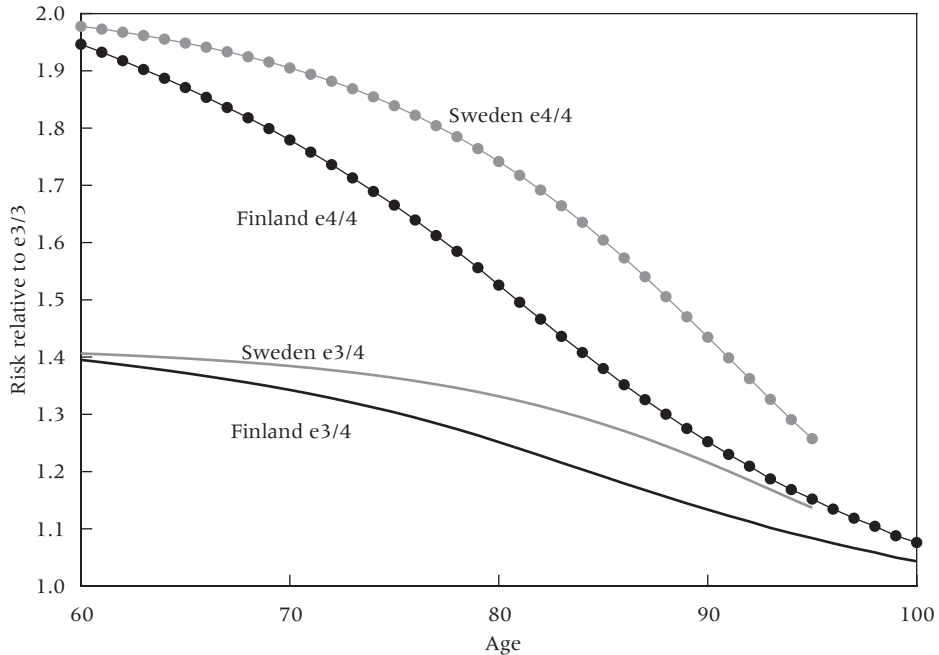
The use of a life table and the assumption about the relative risks are similar to some of the methods proposed by Yashin et al. (1999) for the analysis of cross-sectional data on genotype frequencies. There are a few differences. First, the life tables used here are cohort life tables. Second, instead of assuming that the relative risk is the same at all ages, I assume that they change with age in a way that is consistent with a Gamma distribution of frailty. In addition, multiple types of data are used to estimate the model.

Findings

The differences in mortality by genotype are all significant. The relative risks at death for the main APOE genotypes relative to e3/3 at age 60 are 1.4 for e3/4 and 0.81 for e2/3. The relative risk for e4/4 is not significantly different from the square of the risk for e3/4, 1.96, which means the effects of the e4 allele follow a dose–response pattern. There were no significant differences in the relative risks of death between men and women. In addition, there were no significant differences between northern and southern Europe. This is an interesting finding. Since APOE plays an important role in managing the transport of fats in the blood stream, we might expect the effect of the e4 allele on IHD mortality to be greater in populations whose diet is higher in cholesterol and fats. This effect is not discernible in the available data.

The differences by genotype become small after age 100 (the three relative risks approach 1.0) because of the effects of unobserved heterogeneity. The changes in the relative risks with age differ substantially from cohort to cohort because of large differences in mortality. For example, Figure 4 shows the relative risks by age for the e3/4 and e4/4 genotypes for males in Finland born in 1886–90 and for Swedish females born in 1905–09. Only 7.5 percent of the men in the Finnish cohort survived from age 60 to age 90, whereas

FIGURE 4 Risk of death, APOE genotypes e3/4 and e4/4 relative to e3/3, Swedish females born 1905–09 and Finnish males born 1886–90



25.1 percent of the women in the Swedish cohort survived to that age. With the higher mortality rates among the Finnish males, those with numerous risk factors die out very quickly and those $\epsilon 4$ carriers who survive to the oldest ages probably carry numerous unobserved protective factors.

Insights

The most important methodological finding is that it is possible to introduce heterogeneity into the model without imposing a Weibull or a Gompertz curve on the mortality rates. Instead, we can use an existing life table to define overall mortality rates and simply assume a distribution of frailty that determines changes in the relative risks by age.

Second, the use of cohort life tables does make a difference in the estimates because the relative risks approach 1.0 faster with higher mortality. However, it is differences in mortality at the older ages that are important. Therefore, it is not necessary to have very precise mortality rates for the earliest cohorts at young ages.

The APOE gene and differences in life expectancy in Europe

My fourth paper (Ewbank 2004) addresses two questions of great interest to demographers and others concerned with population-level effects: what are the implications of differences by APOE genotype for mortality differences within and among European countries? Since the three APOE alleles are relatively common and are associated with different risks of mortality, it is likely that APOE genotype frequencies may play an important role. Therefore, we can ask how much of the difference in life span (age at death) among individuals in a population is explained by APOE genotype. In addition, the frequencies of the APOE alleles differ among European countries: for example, the $\epsilon 4$ allele frequency varies from 9.4 percent in Italy to 21.6 percent in Sweden. In general, the $\epsilon 4$ allele is much more common in northern Europe (here represented by Denmark, Finland, and Sweden) than in southern Europe (Italy and France). This north–south gradient has been noted by several authors (Lucotte, Loirat, and Hazout 1997; Panza et al. 1999). The $\epsilon 2$ frequencies vary from 3.9 percent in Finland to 9.5 percent in Sweden and do not follow any simple geographic pattern in Europe.

Some evidence on this comes from a study by Stengård, Weiss, and Sing (1998). They examined data on middle-aged males from nine European populations to study the relationship between APOE and coronary heart disease (CHD). They found that differences in the frequency of the $\epsilon 4$ allele explained 75 percent of the variance in CHD mortality. Although this is an impressive result, the ecological fallacy states that associations

found at the population level do not necessarily reflect differences at the individual level.

An alternative approach is to cumulate data at the individual level (e.g., longitudinal studies on APOE and mortality) to estimate differences at the population level.

Methods⁹

The methods used here are simple analysis of variance and standardization. To study the effects of APOE on the life span of individuals in a cohort, I examine the sum of squared deviations in age at death around the mean age at death, $e(x)$. These calculations are based on the ${}_1d_x$ column of the life table (in survival analysis terms: $D(x) = S[x]-S[x+1]$). To calculate the proportion of the variation in life span that is attributable to differences by genotype, we compare the sum of squares around $e(x)$ with the sum of squares around the genotype-specific mean ages at death, e.g., $e^{3/4}(x)$ for genotype $e^{3/4}$.

The effect of differences in genotype frequencies on mortality differences among countries is also based on an analysis of variance. The variance among countries in a mortality measure (for example, $e(0)$ or a mortality rate) is compared to the variance among values standardized to a common set of genotype frequencies. For example, to examine the effect on life expectancy at birth, $e(0)$, I standardize life expectancy using a standard set of gene frequencies at birth. The life expectancy at birth in Denmark standardized to the gene frequencies in Italy is:

$$e_D^I(0) = f_{2/3}^I e_{2/3}^D(0) + f_{3/3}^I e_{3/3}^D(0) + \dots + f_{4/4}^I e_{4/4}^D(0)$$

where $f_{2/3}^I$ is the proportion with genotype $e^{2/3}$ at birth in Italy and $e_{2/3}^D(0)$ is the life expectancy at birth for the $e^{2/3}$ genotype in Denmark. Similarly, for $e(65)$ I weight the genotype-specific life expectancies at age 65 in Denmark by the genotype frequencies at age 65 in Italy.

Findings

The differences in age at death among genotypes are quite large. For example, in the cohort born in Denmark 1895–99 the estimates of $e(15)$ for both sexes combined for the relatively common $e^{2/3}$ and $e^{3/4}$ genotypes differ by 3.1 years. Although this is a substantial difference, it is relatively small compared to the standard deviation in life span after age 15, 17.6 years. Calculating the sum of squared differences around the genotype-specific $e(15)$, I find that the differences by APOE genotype explain 0.9 percent of the variation in life span over age 15 in Denmark. For survivors to

age 65, the proportion of the variance in remaining life span explained by APOE genotype is 1.8 percent.

No single factor explains a large proportion of the variation in age at death. We can put these percentages into perspective by comparing them to the proportions explained by sex and an estimate of the proportion explained by all gene effects. Differences in adult mortality by sex are substantial. For example, for the cohort born in Denmark in 1895–99 the female life expectancy at age 15 exceeds the male value by 3.3 years. These differences by sex explain the same proportion of the variation in life span as APOE genotype: 0.9 percent. At age 65, sex explains about twice as much as APOE genotype: 3.5 percent compared to 1.8 percent.

The proportion of the variance in age at death attributable to APOE depends on the gene frequencies. For example, the frequencies of the e2 and e4 alleles in France are smaller than in Denmark. Therefore, for the same birth cohort in France APOE explains less of the variation in life span after age 65 than in Denmark: 1.2 percent compared to 1.8 percent. In addition, the differences in e(65) by sex are much larger in France than in Denmark. Sex explains 5.5 percent of the variance after age 65 compared to only 3.5 percent in Denmark. In general, in Western Europe the effect of APOE on life span after age 65 is comparable to, but generally less than, the effect of sex.

The proportion of the variation in life span estimated for the Danish cohort can also be compared to an estimate of the proportion of the variance explained by all genetic effects. Herskind et al. (1996) used data on Danish twins born in 1870–1900 to estimate the proportion of life span after age 15 that is explained by genetic differences. They estimated that about 25 percent of the variation in adult life span is attributable to genetics. Therefore, in this Danish cohort, differences in mortality by APOE genotype explain about 3.5 percent (i.e., 0.009/0.25) of the contribution of genetics.

The differences in APOE genotype frequencies explain a large share of the differences among European countries. Table 1 provides the data re-

TABLE 1 Estimated life expectancy at age 65 for males by APOE genotype for Denmark in the mid-1990s and APOE genotype frequencies at age 65 in Denmark and Italy

	e2/2 and e2/3	e3/3 and e2/4	e3/4	e 4/4
Life expectancy at age 65, Denmark (years)	15.9	14.6	12.7	10.9
APOE genotype frequencies at age 65 (%)				
Denmark	14.1	59.1	24.3	2.5
Italy	13.3	71.4	14.5	0.7

SOURCE: Based on Ewbank 2004.

quired to adjust the $e(65)$ for males in Denmark to the APOE genotype frequencies in Italy. In the mid-1990s, Italian males had an $e(65)$ of 15.5 years compared to 14.3 for Danish males. The $\epsilon 3/4$ and $\epsilon 4/4$ genotypes—which have the lowest values of $e(65)$ —are much less common in Italy than in Denmark (Table 1). Combining the $e(65)$ values from Denmark with the genotype frequencies in Italy leads to a standardized $e(65)$ of 14.5, 0.25 years higher than the actual value in Denmark. Therefore, differences in APOE genotype frequencies explain 20 percent of the observed difference between these two countries in $e(65)$ in males. Standardizing the values of $e(65)$ for Denmark, Finland, France, the Netherlands, and Sweden to the genotype frequencies in Italy, I find that differences in APOE genotype frequencies explain 16 percent of the differences in $e(65)$ for males and 17 percent of the differences for females. Therefore, APOE explains a substantial share of the differences in mortality after age 65 among these countries.

Insights

These results demonstrate one of the advantages of demographic synthesis over other meta-analytic techniques. By estimating a multistate life table, demographic synthesis produces estimates of quantities that cannot be measured reliably from longitudinal studies with feasible sample sizes. In this case, the model provides estimates of $e(x)$ by genotype that are not available from any of the published studies. These estimates can be used for comparisons within and among populations. They can also be used to simulate the effects of reducing or eliminating the differences by genotype. For example, we could estimate what would happen if we had a therapy that reduced the excess mortality associated with the $\epsilon 4/4$ genotype. This approach can be used to study the effects of any fixed factors on differences in life span.

A demographic approach to synthesizing data on risk factors for chronic disease

This approach to modeling chronic disease relies on one of the fundamental concepts in epidemiology: the relationship between prevalence and cumulated incidence rates is determined by the duration of disease and differences in survival rates by disease status. In the case of chronic diseases, the duration is the time to death. Epidemiologists have generally not exploited this relationship in the study of chronic diseases. Instead this link between incidence and prevalence is central to demographic thinking. For example, the prevalence of people at each age (i.e., the age distribution) is determined by the history of birth, death, and migration incidence rates. This difference between epidemiology and demography reflects a fundamental difference between the two disciplines.

Most epidemiologists are primarily interested in determining factors associated with the onset of disease. Research on the progression of disease is usually left to clinical researchers, biostatisticians, and clinical epidemiologists. Prevalence studies are generally used to demonstrate the population burden of a disease or they are used as proxies for incidence rates in association studies. In general, epidemiologists model the relationship between incidence and prevalence for infectious diseases only where the prevalence of infectious cases in the population drives the incidence of new cases (e.g., HIV/AIDS, malaria, and measles). However, it is rarely necessary for epidemiologists to pay much attention to the linkage between incidence and prevalence.

Demographers, on the other hand, start with an interest in both prevalence and entrances to and exits from the population (births, deaths, and migrations). The study of the incidence or onset of various states (for example, marriage rates) is often driven by an interest in the prevalence of the state (for example, family structure or the prevalence of unmarried young adults). Therefore, demographers start with an interest in how dynamic processes determine the distribution of the population (the prevalence) by various characteristics. This perspective is the foundation for the way in which demographic synthesis estimates multistate models.

Multistate models have been used by demographers to study disease processes (e.g., Crimmins, Saito, and Ingegneri 1997; Al Mamun 2003) and social processes (e.g., Hayward and Grady 1990; Schoen and Standish 2001). The applications have often been based on transition rates (denoted as ${}_n m_x$ or $\mu(u)$) or on transition probabilities (${}_n q_x$ or $S(x+1)/S(x)$) from multiround surveys or vital statistics (Keyfitz 1985). In the absence of data on transitions, researchers have used current status data (i.e., prevalence rates) to estimate measures such as years of life free of chronic disease or disability (Crimmins, Saito, and Ingegneri 1989). However, demographic synthesis can combine a wide range of inputs including incidence, prevalence and odds ratios. In this way, the models that I have used to study Alzheimer's disease and the risks associated with APOE genotypes emphasize the fundamentally demographic nature of multistate models.¹⁰

By combining data from numerous studies, demographic synthesis performs a meta-analytic function. Unlike typical meta-analytic methods, however, this approach can integrate information from a wide range of study designs. It also produces estimates of quantities that cannot be, or have not been, adequately measured directly (e.g., life expectancy). The process is one of combining disparate pieces of a whole (synthesis) rather than the more typical social science activity of identifying the components or causes of a socially defined entity (analysis).

Demographic synthesis is ideally suited to the study of Alzheimer's disease and the effects of APOE genotypes. Alzheimer's is a chronic disease for which there is no cure. The treatments that are available only slow the pro-

cess for a short time. This makes it much easier to model than infectious or acute conditions that would require modeling of recovery and, possibly, reinfection. In addition, there is no evidence that the risk of Alzheimer's disease has changed in recent decades (although there is some evidence of a decline in dementia and impaired cognitive functioning). These characteristics of Alzheimer's disease simplify both the modeling and the data requirements. However, it is possible to use this approach for infectious diseases and in situations in which the incidence rates have been changing. For example, some of the basic concepts underlying demographic synthesis have been used by Heuveline for the study of HIV/AIDS, the prevalence of which has increased rapidly in the past two decades (Heuveline 2003).

The ability to generalize these models to other chronic diseases was demonstrated in a paper I wrote with Robbins (Robbins and Ewbank 2001). We used demographic synthesis to study the onset of Parkinson's disease and dementia among cases of Parkinson's. We combined data on the incidence and prevalence of Parkinson's with data on the onset of dementia by duration of Parkinson's and estimates of the relative risk of mortality among Parkinson's cases and cases of Parkinson's with dementia. This approach provides estimates of numerous useful summary statistics such as dementia-free life expectancy of patients with Parkinson's disease.

APOE is also well suited to demographic synthesis. First, it meets the three criteria for what I have termed a "demogene"—a gene that has noticeable effects at the population level (Ewbank 2000). Demogenes must have variants that are associated with sizable differences in the risk of common conditions, and these variants must be common in the population (e.g., carried by more than 5 percent of the population). The importance of a demogene is enhanced if the frequency of the variants differs substantially across populations. To date, the gene for APOE is the only gene proven to meet all three of these criteria. Demographic synthesis can be applied to genes that do not meet some or all of these characteristics if there is sufficient population-level data. However, the full benefits of demographic synthesis for population genetics are realized for demogenes.

The applicability of demographic synthesis to the study of other fixed traits has been demonstrated by Stone (2004). She has used this approach to study the mortality of individuals who have a sibling who survived to age 110. The advantage of this approach is that it is semi-parametric—it models the changes in the relative risks without imposing a parametric form for the full life table. In that regard it improves on the nonparametric Kaplan–Meier approach without imposing a Gompertz or Weibull distribution on the age pattern of risk.

The value of this approach is not limited to applications to chronic disease and mortality. It can be applied to estimate any multistate model. The modeling is easier if the risks (or the relative risks) have stayed constant. Even this assumption is not necessary, however.

Postscript

This chapter has described the evolution of demographic synthesis from a specific application (the study of Alzheimer's disease as a cause of death) to a generalized approach to studying chronic disease and the population-level effects of genetic variability. Demographic synthesis also offers an alternative way to fit multistate models. I illustrated the development of this approach through references to four previously published papers. However, this discussion makes the process of discovery look far more orderly than it was. When describing an intellectual process, the chronological approach can be at odds with the logical. The situation that led to a discovery is rarely the best example for explaining it. Over time, what began as a solution to a small problem at hand slowly becomes an approach with more general applications. Therefore, this retrospective look obscures what was really a process of rediscovery, refinement, and restatement.

This voyage of discovery is not complete. Currently, I am expanding the model of the risk of Alzheimer's disease by APOE genotype and applying it to the raw data from a large number of case-control studies. I am beginning a study of the effects of several fixed characteristics on mortality at the oldest ages. I also plan to modify the models of genetic effects to address questions that involve latent classes and to develop applications to the study of centenarians. Variants of these models can find wide applicability in both demography and epidemiology.

Notes

This research was supported by National Institute on Aging (NIA) grant R01-AG16683.

1 This section and the subsequent section, "Findings," are based on "Deaths attributable to Alzheimer's disease in the U.S." (Ewbank 1999).

2 The published data are actually for much larger age groups. Therefore, we need to weight the $\lambda(x)$ using an appropriate age distribution by single years of age.

3 The model also relied on estimates of excess mortality among Alzheimer's disease cases by duration of disease. These were estimated by averaging data from several clinic-based studies.

4 In many articles and in discourse, the distinction between the gene forms ($\epsilon 2$, $\epsilon 3$, and $\epsilon 4$) and the resulting gene products ($e 2$, $e 3$, and $e 4$) is often ignored.

5 This section and the subsequent section, "Findings," are based on "A multistate model of the genetic risk of Alzheimer's disease" (Ewbank 2002a).

6 For a brief discussion of some of these issues, see Ewbank 1998.

7 This section and the subsequent section, "Findings," are based on "Mortality differences by APOE genotype estimated from demographic synthesis" (Ewbank 2002b).

8 Since the time this paper was published in *Genetic Epidemiology*, I have updated the estimates using two new data sets from the Netherlands and two from the United States. Therefore, the sample sizes and the results given here are slightly different from those given in the paper. These more recent estimates are described briefly in the fourth paper discussed below.

9 This section and the subsequent section, "Findings," are based on "The APOE gene and differences in life expectancy in Europe" (Ewbank 2004).

10 One of the risks in using current status (prevalence) data is that they may be affected by past trends in incidence rates or relative

risks. The applications described here are all based on the assumption that the Alzheimer's disease incidence rates and all relative risks at young ages have remained constant. These assumptions can easily be relaxed if sufficient data are available.

References

- Al Mamun, Abdullah. 2003. *Life History of Cardiovascular Disease and Its Risk Factors: Multistate Life Table Approach and Application to the Framingham Heart Study*. Amsterdam, Netherlands: Rozenberg Publishers.
- Barclay, Laurie L., Alexander Zemcov, John P. Blass, and Joseph Sansone. 1985. "Survival in Alzheimer's disease and vascular dementias," *Neurology* 35: 834–840.
- Beard, C. Mary, Emre Kokmen, Cathy Sigler, Glenn E. Smith, Tanya Petterson, and Peter C. O'Brien. 1996. "Cause of death in Alzheimer's disease," *Ann Epidemiology* 6(3): 195–200.
- Crimmins, Eileen M., Yasuhiko Saito, and Dominique Ingegneri. 1989. "Changes in life expectancy and disability-free life expectancy in the United States," *Population and Development Review* 15(2): 235–267.
- . 1997. "Trends in disability-free life expectancy in the United States, 1970–90," *Population and Development Review* 23(3): 555–572.
- Eichner, June E. et al. 1993. "Relation of apolipoprotein E phenotype to myocardial infarction and mortality from coronary artery disease," *Am J Cardiol* 71: 160–165.
- Evans, Denis A. et al. 1989. "Prevalence of Alzheimer's disease in a community population of older persons," *Journal of the American Medical Association* 262: 2551–2556.
- Evans, Denis A., Laurel A. Smith, Paul A. Scherr, Marilyn S. Albert, H. Harris Funkenstein, and Liese E. Hebert. 1991. "Risk of death from Alzheimer's disease in a community population of older persons," *Am J Epidemiol* 134(4): 403–412.
- Ewbank, Douglas C. 1998. "Effects of age and ethnicity on the link between APOE ϵ 4 and Alzheimer's disease," *Journal of the American Medical Association* 279(8): 580–581.
- . 1999. "Deaths attributable to Alzheimer's disease in the U.S.," *Am J Public Health* 89(1): 90–92.
- . 2000. "Demography in the age of genomics: A first look at the prospects," in Caleb E. Finch, James W. Vaupel, and Kevin Kinsella (eds.), *Cells and Surveys: Should Biological Measures Be Included in Social Science Research?* Washington, DC: National Academy Press.
- . 2002a. "A multistate model of the genetic risk of Alzheimer's disease," *Exp Aging Res* 28(4): 477–499.
- . 2002b. "Mortality differences by APOE genotype estimated from demographic synthesis," *Genet Epidemiol* 22(2): 146–155.
- . 2004. "The APOE gene and differences in life expectancy in Europe," *J Gerontol A Biol Sci Med Sci* 59A(1): 8–12.
- Farrer, Lindsay A. et al. 1997. "Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease," *Journal of the American Medical Association* 278: 1349–1356.
- Hayward, Mark D. and William R. Grady. 1990. "Work and retirement among a cohort of older men in the United States, 1966–1983," *Demography* 27(3): 337–356.
- Herskind, Anne Marie, Matthew McGue, Niels V. Holm, Thorkild I. A. Sørensen, Bent Harvald, and James W. Vaupel. 1996. "The heritability of human longevity: A population-based study of 2872 Danish twin pairs born 1870–1900," *Hum Genet* 97(3): 319–323.

- Heuveline, Patrick. 2003. "HIV and population dynamics: A general model and maximum-likelihood standards for East Africa," *Demography* 40(2): 217–245.
- Hoyert, Donna L. 1996. "Mortality trends for Alzheimer's disease, 1979–1991," *Vital Health Stat* 20(28).
- Jorm, A. F. and D. Jolley. 1998. "The incidence of dementia: A meta-analysis," *Neurology* 51(3): 728–733.
- Juva, K., A. Verkkoniemi, P. Viramo, T. Polvikoski, K. Kainulainen, K. Kontula, and R. Sulkava. 2000. "APOE epsilon4 does not predict mortality, cognitive decline, or dementia in the oldest old," *Neurology* 54(2): 412–415.
- Katzman, R. 1986. "Alzheimer's disease," *New England Journal of Medicine* 314: 964–973.
- Keyfitz, Nathan. 1985. *Applied Mathematical Demography*. New York : Springer-Verlag.
- Kukull, W. A., D. E. Brenner, C. E. Speck, D. Nochlin, J. Bowen, W. McCormick, L. Teri, M. L. Pfanschmidt, and E. B. Larson. 1994. "Causes of death associated with Alzheimer disease variation by level of cognitive impairment before death," *J Am Geriatr Soc* 42(7): 723–726.
- Larson, Eric B., Marie-Florence Shadlen, Li Wang, Wayne C. McCormick, James D. Bowen, Linda Teri, and Walter A. Kukull. 2004. "Survival after initial diagnosis of Alzheimer disease," *Ann Intern Med* 140(7): 501–509.
- Levy-Lahad, Ephrat and Thomas D. Bird. 1996. "Genetic factors in Alzheimer's disease: A review of recent advances," *Ann Neurol* 40(6): 829–840.
- Lucotte, G., F. Loirat, and S. Hazout. 1997. "Pattern of gradient of apolipoprotein E allele $\epsilon 4$ frequencies in Western Europe," *Hum Biol* 69 (2): 253–262.
- Macera, Caroline A., R. K. Sun, Kimberly K. Yeager, and Debra A. Brandes. 1992. "Sensitivity and specificity of death certificate diagnoses for dementing illnesses, 1988–1990," *J Am Geriatr Soc* 40(5): 479–481.
- Manton, Kenneth G., Eric Stallard, and James W. Vaupel. 1986. "Alternative models for the heterogeneity of mortality risks among the aged," *JASA* 81(395): 635–644.
- Panza, Francesco et al. 1999. "Decreased frequency of apolipoprotein E $\epsilon 4$ allele from Northern to Southern Europe in Alzheimer's disease patients and centenarians," *Neurosci Lett* 277: 53–56.
- Raiford, K., S. Anton-Johnson, Z. Haycox, K. Nolan, A. Schaffer, C. Caimano, G. Fillenbaum, and A. Heyman. 1994. "CERAD Part VII: Accuracy of reporting dementia on death certificates of patients with Alzheimer's disease," *Neurology* 44(11): 2208–2209.
- Robbins, Jessica M. and Douglas C. Ewbank. 2001. "Dementia in Parkinson's disease: Demographic models and estimates," paper presented at the Annual Meeting of the Population Association of America, Washington, DC, 29–31 March.
- Rosenberg H. et al. 1996. "Births and deaths: United States, 1995," *Monthly Vital Statistics Report* 45(3) suppl 2.
- Schoen, R. and N. Standish. 2001. "The retrenchment of marriage: Results from marital status life tables for the United States, 1995," *Population and Development Review* 27(3): 553–563.
- Skoog, I., C. Hesse, O. Aevarsson, S. Landahl, J. Wahlstrom, P. Fredman, and K. Blennow. 1998. "A population study of ApoE genotype at the age of 85: Relation to dementia, cerebrovascular disease, and mortality," *J Neurol Neurosurg Psychiatry* 64(1): 37–43.
- Stengård, Jari H., Kenneth M. Weiss, and Charles F. Sing. 1998. "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* 103(2): 234–241.
- Stone, Leslie. 2004. "Studies in the demography of supercentenarians in the United States." Unpublished doctoral dissertation, University of Pennsylvania.
- Tilvis, R. S., T. E. Strandberg, and K. Juva. 1998. "Apolipoprotein E phenotypes, dementia and mortality in a prospective population sample," *J Am Geriatr Soc* 46(6): 712–715.
- Toupance, Bruno, Bernard Godelle, Pierre-Henri Gouyon, and François Schächter. 1998.

- "A model for antagonistic pleiotropic gene action for mortality and advanced age," *Am J Hum Genet* 62: 1525–1534.
- Wilson, Peter W. F., Ernst J. Schaefer, Martin G. Larson, and Jose M. Ordovas. 1996. "Apolipoprotein E alleles and risk of coronary disease: A meta-analysis," *Arterioscler Thromb Vasc Biol* 16: 1250–1255.
- Vogt, M. T., J. A. Cauley, and L. H. Kuller. 1997. "Apolipoprotein E phenotype, arterial disease, and mortality among older women: The study of osteoporotic fractures," *Genet Epidemiol* 14(2): 147–156.
- Yashin, A. I., G. De Benedictis, J. W. Vaupel, Q. Tan, K. F. Andreev, I. A. Iachine, M. Bonafe, M. DeLuca, S. Valensin, L. Carotenuto, and C. Franceschi. 1999. "Genes, demography and life span: The contribution of demographic data in genetic studies of aging and longevity," *Am J Hum Genet* 65(1): 78–93.
- Yashin, A. et al. 2000. "Genes and longevity: Lessons from studies of centenarians," *J Gerontol* 55A(7): B319–B328.

