10

Kindred Lifetimes: Frailty Models in Population Genetics

JAMES W. VAUPEL

How in a heterogeneous population do individual life-history traits that are theoretically important but largely unobservable, affect observed population dynamics? How can inferences be drawn about the underlying traits from the population patterns? As illustrated by Trussell and Rodríguez, Hoem, and Weiss (Chapters 8, 9 and 12), this pair of questions is of convergent interest to both geneticists and demographers; this chapter presents a new method for addressing some aspects of it.

Consider, for example, a life-history trait—lifetime—that is of fundamental interest to both geneticists and demographers. In studies of the duration of life, the data typically consist of a distribution of individual lifetimes and a corresponding age-specific survival curve. Geneticists ask: to what extent is the variation in lifetimes due to genetic versus environmental variation among individuals? Demographers ask: what is the shape of the survival curve for individuals and how does this trajectory differ across individuals? If individuals are classified by genotype (and if the effects of common environment are unimportant), then the demographers' question is the same as the geneticists', because variation across individuals is then genotypic variation and the variation in lifetimes implied by differing survival curves is the residual environmental variation.

Even with recent advances in mapping genes, genotypes are still largely unobserved and many of the details of environmental variation are also unobserved. Some environmental covariates, such as year of birth, caloric intake, and temperature may be measured in some empirical studies, but it is not practicable to measure all environmental and behavioral perturbations. Thus, as suggested by Trussell and Rodríguez in this volume, a metaphor for the geneticists' and demographers' question is the decomposition of a known quantity C into the sum of two unknown quantities A and B. Unless some further information is added, the equation $C = A + B$ has infinitely many solutions.

The tack taken by geneticists is to combine (1) theories of how genes are transmitted and (2) theories and assumptions about how genes interact with other genes and with environmental factors with (3) empirical data on related

individuals, such as twins, siblings, or parents and children. The observed population variation then can be uniquely decomposed into the two components of genetic and environmental variation. This classic and widely used method unfortunately has numerous limitations (e.g., Feldman and Lewontin, 1975; Falconer, 1981).

Demographers, in contrast, proceed by specifying functional forms for their A (the shape of survival curves for individuals) and B (the variation in survival curves across individuals) and for the relationship between A and B. This approach can also be severely criticized, especially when there is little ancillary evidence concerning the true nature of these functional forms [as discussed by Trussell and Rodríguez (Chapter 8) and by Hoem (Chapter 9) in this volume]. Following the geneticists' lead, this chapter develops a method that is grounded in genetic theory and that exploits data on related individuals.

Following the demographers' lead, the method is based on modern ideas of survival analysis and frailty modeling (rather than on the decomposition of variance). The hybrid method has some limitations and some advantages compared with existing methods. In some life-history applications, it may provide geneticists and demographers with a useful, convergent supplement to their current, disparate approaches.

For expository simplicity, the chapter focuses the analysis of lifetimes. As discussed in two companion articles (Vaupel, 1990; Larsen and Vaupel, 1989), the method developed for analyzing mortality is directly applicable to a variety of other life-history characteristics, including fertility, migration, marriage, and morbidity, and to data sets on repeated events as well as related individuals: instead of ages at death for relatives, the data might consist of times to successive conceptions. The illustrative examples used in the chapter pertain to human twins and to adopted children and their biological and adoptive parents, but applications to other sets of relatives and to other species can be developed. To facilitate extensions to various life-history traits, kin groupings, and species, mathematical results are presented in quite general terms. The mathematics, however, is not difficult and the only results presented are those of direct interest to geneticists and demographers who are analyzing survival or duration data.

A FRAILTY MODEL FOR GENOTYPES

Consider first a population of individuals who are classified into groups with identical genotypes. The data may pertain, for example, to monozygotic twins or to a set of inbred lines and the offspring derived from crosses between them (F_1 crosses). Suppose these data are of the kind typically studied in the branch of statistics known as survival analysis (Kalbfleisch and Prentice, 1980; Cox and Oakes, 1984). In particular, suppose the data include the lifetime X_{ij} of individual i in genetic group j, an indicator δ_{ij} equal to 1 if X_{ij} is a death time and 0 if X_{ij} is the oldest age when the individual was known to be alive prior to being censored (i.e., lost to further observation) in an uninformative way, and, perhaps, a vector of covariates v_{ij} that may vary with age or time.

Frailty models for analyzing survival data focus on estimating the age trajectory of the force of mortality (i.e., hazard of death). The force of mortality $\mu(x)$ at age x is related to the survival curve $s(x)$, which gives the probability

of surviving to age x, by

$$\mu(x) = \frac{-ds(x)/dx}{s(x)} \tag{10-1}$$

and

$$s(x) = e^{-H(x)}\,, \tag{10-2}$$

where the cumulative hazard $H(x)$ is given by

$$H(x) = \int_0^x \mu(t)dt\,. \tag{10-3}$$

The probability density function of age at death (i.e., lifetime) is given by

$$f(x) = \mu(x)s(x)\,. \tag{10-4}$$

In frailty models it is assumed that the force of mortality for an individual can be separated into two multiplicative components called frailty and the baseline force of mortality (Vaupel, Manton, and Stallard, 1979). This may not be entirely true in particular applications, but no model is a perfect representation of reality. The operative question is whether the model is useful: is it simple enough to be tractable and understandable but sophisticated enough to shed some new light on reality? The simplicity of the multiplicative frailty approach is analogous to the simplicity of linear regression; in a variety of theoretical and empirical applications, frailty models have provided useful insights (e.g., Vaupel, Manton, and Stallard, 1979; Manton, Stallard, and Vaupel, 1981 and 1986; Heckman and Singer, 1984; Vaupel and Yashin, 1985a,b; Hougaard, 1986a; Aalen, 1987, 1988).

For data classified by genotypic groups, the force of mortality for individual i in group j would be $z_j\mu_{ij}(x)$, where z_j denotes the frailty (or relative risk) of each of the individuals in group j and $\mu_{ij}(x)$ gives the baseline force of mortality. The value of z_j is not known; it is described by a probability density function $g_j(z)$. In many applications the same g will hold for all genotypic groups and this g may be interpreted as the distribution of genotypic frailty in the population. The key idea is that genotype determines frailty rather than the phenotypic trait (lifetime) *per se*. In genetics, the concept of liability is sometimes used, the notion being that an individual is susceptible to, say, some cause of death only if the individual's liability exceeds some threshold (Falconer, 1981). Frailty is fundamentally different from this kind of liability: frailty is a relative risk such that the greater an individual's frailty with regard to some cause of death (or death in general) the greater the individual's susceptibility to the cause of death. (See Vaupel, 1988 for further discussion of frailty with regard to overall mortality and see Weiss, Chapter 12, for some innovative ideas concerning frailty with respect to specific diseases).

The baseline force of mortality $\mu_{ij}(x)$ is a function of the individual's age x and any covariates v_{ij}. Often the log-linear form

$$\mu_{ij}(x) = e^{c'v_{ij}}\mu^0(x) \tag{10-5}$$

is used, where c is a vector of parameter values. The function $\mu^0(x)$, which describes the underlying age pattern of the force of mortality, is frequently represented by the Gompertz function ae^{bx} or the Weibull function ax^b. Because a

wide variety of other representations of $\mu_{ij}(x)$ may, however, be more reasonable in genetic and demographic research, throughout this chapter the general notation $\mu_{ij}(x)$ will be used to describe the baseline hazard faced by an individual with a given set of covariates. In the simplest case, no covariates are observed and $\mu_{ij}(x)$ is given by the same $\mu(x)$ for every individual. In another simple case, the only covariate is year of birth and the subscripts ij on μ merely indicate that different birth cohorts may suffer different levels and patterns of mortality.

In empirical applications of frailty models, the observed survival data are used to estimate the parameters of the distribution of frailty $g(z)$ and the baseline force of mortality $\mu_{ij}(x)$. Usually the parameters are estimated so as to maximize the likelihood of the observed data. The likelihood L of the survival data on a set of genotypes is the product of the likelihood L_j for each genotypic group:

$$L = \prod_{j=1}^{J} L_j \ . \tag{10-6}$$

A key mathematical result of this chapter is to derive a formula for the genotypic likelihoods. The formula is based on the theory of survival analysis as explained in such standard texts as Kalbfleisch and Prentice (1980) and Cox and Oakes (1984); it is related to a stream of biostatistical research, reviewed by Hougaard (1987), on so-called multivariate survival analysis. Multivariate, in this context, refers not to multiple covariates but to groupings of survival times; the terms "kindred-survival analysis" and "kindred-frailty models" are used in this chapter.

The required formula can be expressed as:

$$L_j = e^{h_j} g^{\dagger}(M_j, m_j) \ . \tag{10-7}$$

The formula involves three statistics that summarize the data. The first, h_j, is the total log hazard at observed death times:

$$h_j = \sum_{i=1}^{I_j} \delta_{ij} \log \mu_{ij}(X_{ij}) \ , \tag{10-8}$$

I_j being the number of individuals in genotypic group j. The second summary statistic, M_j, is the total cumulative hazard:

$$M_j = \sum_{i=1}^{I_j} H_{ij}(X_{ij}) \ , \tag{10-9}$$

with

$$H_{ij}(X_{ij}) = \int_0^{X_{ij}} \mu_{ij}(x)dx \ . \tag{10-10}$$

Finally, m_j is the number of deaths,

$$m_j = \sum_{i=1}^{I_j} \delta_{ij} \ . \tag{10-11}$$

The integral transform $g^\dagger$ is given by

$$g^\dagger(M,m) = \int_0^\infty z^m e^{-zM} g(z)dz \,. \tag{10-12}$$

Proof of (10-7) is straightforward. It follows from standard methods of survival analysis that the probability of the survival data for a genotypic group given the value of z is

$$L_{jz} = \prod_{i=1}^{I_j} \left[z\mu_{ij}(X_{ij})\right]^{\delta_{ij}} e^{-zH_{ij}(X_{ij})} \,. \tag{10-13}$$

Furthermore,

$$L_j = \int_0^\infty L_{jz} g(z)dz \,. \tag{10-14}$$

Rearranging terms yields (10-7).

The frailty transform $g^\dagger(M,m)$ has some mathematically interesting properties. Furthermore, it appears frequently in the probability distributions and likelihoods used in kindred-frailty analysis, as illustrated by several formulas in this chapter, and it serves as a bridge between ordinary survival analysis and frailty modeling (on unrelated individuals and events) and kindred-frailty analysis. These properties and features are discussed in Vaupel (1990). For the purposes of statistical estimation, what is most important is that closed-form expressions can be derived for the transform for a variety of frailty distributions $g(z)$.

Suppose, for instance, that the distribution of frailty follows a gamma distribution, as assumed by Beard (1963); Clayton (1978); Vaupel, Manton and Stallard (1979); Oakes (1982); Wild (1983); Clayton and Cuzick (1985); and others:

$$g(z|\lambda,\kappa) = \lambda^\kappa z^{\kappa-1} e^{-\lambda z}/\Gamma(\kappa) \,. \tag{10-15}$$

Then it is readily shown that the frailty transform is

$$g^\dagger(M,m) = \frac{\Gamma(\kappa+m)}{\Gamma(\kappa)} \cdot \frac{\lambda^\kappa}{(\lambda+M)^{\kappa+m}} \,. \tag{10-16}$$

Alternatively, suppose that frailty follows a two-point distribution, such that individuals are either frail or robust (or either movers or stayers), as assumed in analyses of hidden heterogeneity by Blumen, Kogan, and McCarthy (1955), Shepard and Zeckhauser (1980), Keyfitz and Littman (1980), Trussell and Richards (1985), Vaupel and Yashin (1985b), and others. For this simple discrete distribution,

$$g(z_1) = p_1 \,, \qquad 0 < p_1 < 1 \,, \tag{10-17}$$

and

$$g(z_2) = 1 - p_1 = p_2 \,. \tag{10-18}$$

The likelihood formula (10-7) straightforwardly generalizes to discrete distributions, with a summation replacing the integrals and a probability mass function replacing the probability density function in the frailty transform. In particular, for the two-point distribution,

$$g^\dagger(M,m) = p_1 z_1^m e^{-z_1 M} + p_2 z_2^m e^{-z_2 M} \,. \tag{10-19}$$

This result immediately generalizes to N-point distributions. Expressions for other distributional forms of $g(z)$ are given by Vaupel (1990).

APPLICATION TO DANISH MONOZYGOTIC TWINS

In studies of the genetic and early environmental components of the longevity of monozygotic (MZ) twins, the frailty z_j of a twin pair might be defined as the relative risk the two twins share (Hougaard, 1986b, Vaupel, 1988). Data are available on the day, month, and year of birth and death of Danish twins born from 1870 through 1930 (Hauge *et al.*, 1968, Holm, 1983) and a proposal to computerize and analyze these data has been prepared by Vaupel, Holm and others. To explore the estimability of frailty models applied to the Danish twin data, 5 mortality data sets were generated that might resemble the actual data set for Danish male MZ twins. It was assumed that for twin pairs unbroken at age 35, mortality rates were given by the Gompertz trajectory

$$\mu(x) = ae^{-ry+bx} , \tag{10-20}$$

where x is age, y is the birth cohort (varying from zero in 1870 to 60 in 1930), a determines the level of mortality, r is the rate of progress in reducing this level, and b determines how quickly mortality rates increase with age. In the simulation, a was 0.0002, b was 0.1 and r was 0.01. Frailty was assumed to be Gamma distributed with a mean of 1 and a variance of .25; in the simulation the inverse of the variance, k, was used and set equal to 4. The data set generated consisted of 15 twin pairs in the 1870 cohort, gradually increasing to 45 twin pairs in the 1930 cohort. The last year of observation was 1991 (the final year of the proposed data updating and computerizing); all survivors were censored at this time.

Parameter values were then estimated from the simulated data using the likelihood function in (10-7). The results are given in Table 10-1. Reassuringly, the parameter estimates are close to actual values, with no evidence of important bias, and the estimated standard deviations are consistent with the standard deviations of the estimates.

INTERPRETATION OF THE PARAMETERS OF A FRAILTY MODEL

In a frailty model like the one described above, the parameters of a hazard function and a frailty distribution are estimated. Given an observed distribution of lifetimes, the parameters are linked in the following way. As the variance of the distribution of frailty increases, the variance of the distribution of lifetimes for each frailty group decreases. Equivalently, as the distribution of frailty spreads out, the baseline hazard function becomes steeper. In the limit, as the variance in frailty approaches infinity, the hazard function becomes vertical, implying that the level of frailty precisely determines age at death. At the other extreme, when the variance in frailty is zero, the population is homogeneous and the hazard function for the various, equivalent genotypic groups is also the hazard function for the entire population.

Table 10-1 Comparison of actual and estimated parameters of five simulated data sets generated by the frailty model described in the text

		Parameters				
		a	b	ρ	k	σ^2
Actual Values:		.0002	.1	.01	4.	.25
Estimated Values and (S.D.'s):						
Data Set	1	.00019 (.00002)	.101 (.002)	.0107 (.0015)	3.87 (.49)	.26
	2	.00019 (.00002)	.098 (.002)	.0081 (.0015)	4.70 (.69)	.21
	3	.00023 (.00003)	.102 (.002)	.0130 (.0015)	4.10 (.57)	.24
	4	.00022 (.00003)	.102 (.002)	.0125 (.0016)	3.50 (.43)	.29
	5	.00019 (.00002)	.099 (.002)	.0090 (.0015)	4.00 (.53)	.25

Table 10-2 Comparison of variances

Variance in Lifetimes for Entire Population	Variance in Lifetimes for Subpopulation with Frailty Equal to One	Variance in Frailty
161.4	161.4	0.
161.4	142.6	0.25
161.4	114.1	1.
161.4	52.3	10.
161.4	17.1	100.
161.4	0	∞

A simple numerical example provides an illustration. Suppose that the observed distribution of lifetimes is that implied by the Gompertz hazard function ae^{bx}. Further suppose that frailty is Gamma distributed with mean 1 and variance c. Finally, suppose that there is a common baseline hazard $\mu(x)$. Following Vaupel, Manton, and Stallard (1979) it can then be shown that this hazard function has the form

$$\mu(x) = ae^{bx+(ac/b)(e^{bx}-1)} . \tag{10-21}$$

Specifically, suppose a is .00005 and b is .1; using numerical methods it can be calculated that the population life expectancy (mean lifetime) is 70.3 with a variance of 161.4. If various values are specified for c, then numerical methods can be used to calculate the variance in lifetimes for "standard" individuals with frailty 1. Some results are shown in Table 10-2. Vaupel (1988) presents some additional results on the relationship between variance in lifetimes and variance in frailty.

COMPETING RISKS

The analysis of genetic factors in various causes of death is an active research frontier in genetics and epidemiology; Weiss' chapter (Chapter 12) in this volume provides a stimulating example. To extend frailty modeling to this area of research, suppose the force of mortality for individual i in genotypic group j from cause k is $z_j^k \mu_{ij}^k$. If the same value of z governs frailty with regard to two or more causes, these causes can be collapsed in a frailty model into a single, combined cause. Suppose, on the other hand, that different, independent values of z determine frailty with regard to different causes. If the causes of death are observed, then if follows from the standard methods of survival analysis and from (10-7) that the likelihood of the data for a genotypic group j can be expressed as:

$$L_j = \prod_{k=1}^{K} L_j^k \ , \tag{10-22}$$

where

$$L_j^k = e^{h_j^k} \cdot g_k^{\dagger}(M_j^k, m_j^k) \ . \tag{10-23}$$

The transform and three summary statistics in this formula are analogous to those used earlier. The total log hazard is given by:

$$h_j^k = \sum_{i=1}^{I_j} \delta_{ij}^k \ \log \ \mu_{ij}^k(X_{ij}) \ , \tag{10-24}$$

where the indicator δ_{ij}^k is one when the individual is known to have died of cause k and zero otherwise. The total cumulative hazard from cause k is given by:

$$M_j^k = \sum_{i=1}^{I_j} \int_0^{X_{ij}} \mu_{ij}^k(x)dx \ , \tag{10-25}$$

and the number of deaths from cause k is

$$m_j^k = \sum_{i=1}^{I_j} \delta_{ij}^k \ . \tag{10-26}$$

The frailty transform is taken with respect to the probability density function of z^k:

$$g_k^{\dagger}(M, m) = \int_0^{\infty} z^m e^{-zM} g^k(z)dz \ . \tag{10-27}$$

The parameters pertaining to cause k, that is the parameters of the distribution of z^k and of the hazard function $\mu_{ij}^k(x)$, that maximize the likelihood of the data can, in the case of independent causes of death with independent frailties, be estimated by maximizing

$$L^k = \prod_{j=1}^{J} L_j^k \ . \tag{10-28}$$

This convenient result implies that as in the case of cause-of-death data on unrelated individuals, data on independent causes of death for genotypic groups can be analyzed separately without reference to other, competing causes.

HIDDEN COMPETING RISKS

In intermediate cases where cause-specific frailty values are neither perfectly dependent nor independent, more structure is required. One approach is to assume that there are generalized frailty factors that affect two or more causes of death as well as specific factors that affect a single cause. In the simple case of two causes of death (perhaps the cause of interest and a group of other causes), the model might be that a genotype's hazard from cause 1 is

$$(z_j^0 + z_j^1)\mu_{ij}^1(x)\,, \tag{10-29}$$

and from cause 2 is

$$(z_j^0 + z_j^2)\mu_{ij}^2(x)\,. \tag{10-30}$$

This blending of risks can be viewed, at least mathematically, in the usual context of competing risks. However, if an individual dies, say, from cause 1, it is not known whether the operative hazard was the $z^0\mu^1(x)$ or the $z^1\mu^1(x)$ component of the risk. Thus the problem can be interpreted as one of hidden competing risks.

It turns out that methods for analyzing kindred-survival data with hidden competing risks are useful in several other applications of convergent interest to geneticists and demographers. Several specific examples are given subsequently; they involve changes in the impact of genotypic frailty over age, the analysis of premature vs. senescent death, and the analysis of survival data on relatives other than MZ twins.

Suppose K causes of death are known to exist but are not observed. Assume that each frailty group j consists of I_j individuals who share independent frailties $z_j^1, \ldots, z_j^K$ with respect to these causes. Let k_{ij} denote the unobserved cause of death for the i^{th} individual in the j^{th} group; if the individual is lost to follow-up let k_{ij} be zero. Let the vector $(k_{1j}, \ldots, k_{Ij})$ represent a possible set of causes of death and let $L_j^{(k_1,\ldots,k_I)}$ denote the likelihood of this set. This likelihood can be calculated by (10-22) with δ_{ij}^k equal to 1 when k equals k_{ij} and zero otherwise. The situation here is exactly the same as with observed causes of death because the possible set of causes is assumed to be the actual set.

The likelihood of the actual data on the genotypic group is simply the sum of these cause-specific likelihoods over all possible sets of causes of death:

$$\begin{aligned} L_j &= \sum_1^K \cdots \sum_1^K L_j^{(k_1,\ldots,k_I)} \\ &= \sum_1^K \cdots \sum_1^K e^{h_j^{(k_1,\ldots,k_I)}} g_1^\dagger(M_j^1, m_j^1) \ldots g_1^\dagger(M_j^K, m_j^K)\,, \end{aligned} \tag{10-31}$$

where M_j^k and m_i^k are given by (10-25) and (10-26) as before, and where the summations are taken over all individuals whose age at death is known. For censored individuals, lost to follow-up, the value of k is zero.

As a simple example of this formula, consider the case of two hidden causes of death and two MZ twins with known lifetimes X_{1j} and X_{2j}. Then

$$L_j = L_j^{(1,1)} + L_j^{(1,2)} + L_j^{(2,1)} + L_j^{(2,2)}\,, \tag{10-32}$$

where, e.g.,

$$L_j^{(1,1)} = \mu_{1j}^1(X_{1j})\mu_{2j}^1(X_{2j})g_1^\dagger(M_1,2)g_2^\dagger(M_2,0) \tag{10-33}$$

and

$$L_j^{(1,2)} = \mu_{1j}^1(X_{1j})\mu_{2j}^2(X_{2j})g_1^\dagger(M_1,1)g_2^\dagger(M_2,1)\ . \tag{10-34}$$

Proof of (10-31) is straightforward. It follows from standard methods of survival analysis that

$$L_j = \int_0^\infty \cdots \int_0^\infty \prod_{i=1}^{I_j} \left\{ \sum_{k=1}^{K} z_j^k \mu_{ij}^k(X_{ij}) \right\}^{\delta_{ij}} \cdot e^{-\int_0^{X_{ij}} \sum_{k=1}^{K} z_j^k \mu_{ij}^k(x)dx} dz^1 \ldots dz^K\ . \tag{10-35}$$

Rearranging terms and substituting h, M, and m, yields (10-31).

CHANGING FRAILTY AND THE GERONTOLOGICAL PARADIGM

In some analyses it may be appropriate to assume that an individual's frailty changes with age (e.g., Yashin, Manton, and Vaupel, 1985 and Vaupel, Yashin, and Manton, 1988). In the case of the frailty shared by MZ twins, for instance, it may be plausible that the twins' shared frailty, due to common genotype and early environment, becomes less significant as the twins age and cumulatively experience different environmental influences. That is, it might be hypothesized that as they become older, twins become more like unrelated individuals.

This hypothesis could be modeled by setting the force of mortality for twins equal to

$$\{w(x)z_j + [1 - w(x)]\}\mu_{ij}(x)\ , \tag{10-36}$$

with $w(x)$ being a weighting function between zero and one that starts off at one and declines with age. For instance $w(x)$ might be given by e^{-bx}. The parameters of this model can be estimated using the method given above for hidden competing risks, with μ_{ij}^1 equal to $w(x)\mu_{ij}(x)$ and $\mu_{ij}^2(x)$ equal to $[1 - w(x)]\mu_{ij}(x)$ and with z_j^1 equivalent to z_j and z_j^2 equal to one for all genotypes. Because the population is homogeneous with regard to the second "cause of death," the frailty transform with respect to cause 2 reduces to $\exp(-M_j^2)$.

A second example of changing frailty (and of hidden competing risks) is provided by the hypothesis that there are two theoretically-important kinds of death at older ages that may be impossible to distinguish in practice: premature death due to some disease or mishap and genetically-predetermined death due to senescence. This hypothesis, which many gerontologists believe to be correct, was popularized by Fries and Crapo (1981). Two MZ twins might share two frailties with regard to these two broad categories of death and their force of mortality could be modeled by

$$z_j^1\mu_{ij}^1(x) + z_j^2\mu_{ij}^2(x)\ . \tag{10-37}$$

The force of mortality from senescent death is thought by many gerontologists to be an inexorable consequence of aging and essentially independent of envi-

ronmental influences, including personal behavior and medical interventions. Hence, the model might be reduced to

$$z_j^1 \mu_{ij}^1(x) + z_j^2 \mu^2(x) , \tag{10-38}$$

with $\mu^2(x)$ being close to zero until old age and then rising precipitously so that by age 85 or so it becomes the dominant cause of death. As noted earlier, if $\mu^2(x)$ rises sharply, then the distribution of z^2 has a very large variance. Also note that the model implies that the twins' overall frailty starts off at z_j^1 and then moves toward z_j^2 at older ages, with z^2 being more important in determining age at death than z^1 (because it has a large variance and its associated hazard function is much steeper). Thus, in contrast to the previous model, this gerontological model postulates that genotypic factors become more important at advanced ages in determining mortality.

MONZYGOTIC VS. DIZYGOTIC TWINS

As another example of the use of (10-31), consider the analysis of survival data on MZ and dizygotic (DZ) twins. Such data is sometimes used by human geneticists to try to separate the observed variance in, say, lifetimes into the three components of genetic variance, variance due to common environment, and variance due to other environmental factors (Falconer, 1981). In a corresponding frailty model, the force of mortality for MZ twins might be assumed to be given by

$$w z_j^1 \mu_{ij}^1(x) + (1 - w) z_j^2 \mu_{ij}^2(x) , \tag{10-39}$$

whereas the force of mortality for DZ twins might be given by

$$\frac{1}{2} w [z_j^1 + z_{ij}^1] \, \mu_{ij}^1(x) + (1 - w) z_j^2 \mu_{ij}^2(x) , \tag{10-40}$$

with w being a weight between zero and one.

The model can be interpreted as implying that there are two causes of death for MZ twins, due to common genotype and common early environment, but that there are three kinds of death for DZ twins. The first kind is due to common genes and the third kind to different genes, the weight of $\frac{1}{2}$ reflecting the fact that DZ twins share half their genes. The assumption here is that the genetic determinants of frailty are additive; if there are important dominance or interaction effects, or if the effects are additive on some other scale, such as log frailty, then the model needs more structure. Also note that the two twins have different frailties for the third cause of death: under the usual assumption of random mating, these two z's can be assumed to be independently drawn from the population distribution of genotypic frailty.

The likelihood function for the MZ twins is identical to (10-32): there are two causes of death and two individuals per genotypic group. The likelihood function for the DZ twins is somewhat more complicated because there are three causes of death and because the twins differ from each other in their frailty with respect to the third cause. That there are three causes of death implies that

$$\begin{aligned} L_j = L_j^{(1,1)} + L_j^{(1,2)} + L_j^{(1,3)} + L_j^{(2,1)} + L_j^{(2,2)} + L_j^{(2,3)} \\ + L_j^{(3,1)} + L_j^{(3,2)} + L_j^{(3,3)} . \end{aligned} \tag{10-41}$$

Deriving the formula for each of these eight terms requires a slight digression to understand the likelihood of survival data on unrelated individuals.

Consider a specific cause of death and suppose the survival data pertain to individuals with different frailties. Then there are really two groups, with only one member each, instead of one group with two members. The likelihood of the combined data is the product of the likelihoods for each of these single-individual groups:

$$\begin{aligned} L_j &= \{\mu_{1j}(X_{1j})^{\delta_{1j}} g^\dagger(H_{1j}, \delta_{1j})\}\{\mu_{2j}(X_{2j})^{\delta_{2j}} g^\dagger(H_{2j}, \delta_{2j})\} \\ &= e^{h_j} g^\dagger(M_{1j}, m_{1j}) g^\dagger(M_{2j}, m_{2j}) , \end{aligned} \tag{10-42}$$

where in this instance M_{ij} and m_{ij} are equivalent to H_{ij} and δ_{ij}. So instead of a single transform, it is necessary to use the product of two transforms. More generally, whenever individuals share a common frailty, their survival data should be included within the same frailty transform, whereas if they have different frailties, the data should be separated into different transforms.

Returning to the problem of DZ twins, it can now be seen that, e.g.,

$$L_j^{(1,1)} = e^{h_j^{(1,1)}} g^\dagger(M_j^1, 2) g^\dagger(M_j^2, 0) g^\dagger(H_{1j}^1, 0) g^\dagger(H_{2j}^1, 0) \tag{10-43}$$

and

$$L_j^{(2,3)} = e^{h_j^{(2,3)}} g^\dagger(M_j^1, 0) g^\dagger(M_j^2, 1) g^\dagger(H_{1j}^1, 0) g^\dagger(H_{2j}^1, 1) . \tag{10-44}$$

The model for MZ twins can be written as $z\mu(x)$, with z equal to $wz^1 + (1 - w)z^2$.

A geneticist might ask: what is the variance of the population distribution of z and how does this variance compare with the genotypic variance of z^1 vs. the common-environment variance of z^2. Since

$$\mathrm{Var}(z) = \alpha^2 \mathrm{Var}(z^1) + (1 - \alpha)^2 \mathrm{Var}(z^2) , \tag{10-45}$$

the proportion of the variance in overall frailty due to genotypic vs. common-environment variance is

$$\frac{\alpha^2 \mathrm{Var}(z^1)}{\alpha^2 \mathrm{Var}(z^1) + (1 - \alpha)^2 \mathrm{Var}(z^2)} . \tag{10-46}$$

As indicated earlier, the variance in lifetimes is a function of the variance in frailty and the variance in lifetimes among individuals with specific levels of frailty; Vaupel (1990) discusses this. The formulas are, in general, messy, but can be evaluated by numerical methods. Thus, the customary decomposition of variance can be retrieved from a frailty analysis. As discussed, however, by Vaupel (1990), frailty models provide a much richer description than that provided by a decomposition of variance, and this complexity can be used to gain a deeper, multifaceted understanding of the nature of genetic and environmental influences and their interaction.

ADOPTED CHILDREN

As a final example of the use of frailty models with hidden competing risks, consider data on the lifetimes of adopted children and their adoptive and bio-

logical parents (Soerensen *et al.*, 1988). A simple, first-cut approach to frailty modeling of such data would be to use the data to construct three separate data sets, one on biological fathers and children, one on biological mothers and children, and the third on adoptive parents and children.

For the first two of these data sets, the hazard function for each parent/child pair might be

$$\frac{1}{2}z_j\mu_{ij}(x) + \frac{1}{2}z_{ij}\mu_{ij}(x)\,, \tag{10-47}$$

because parents and children, like DZ twins, share half their genes. Then the likelihood would be

$$L_j = e^{h_j}g^{\dagger}(M_j, m_j)g^{\dagger}(H_{1j}, \delta_{1j})g^{\dagger}(H_{2j}, \delta_{2j}) \tag{10-48}$$

For the trios of adoptive parents and children, the hazard function and corresponding likelihood might simply be

$$z_j^*\mu_{ij}^*(x) \tag{10-49}$$

and

$$L_j = e^{h_j}g^{\dagger}(M_j, m_j)\,, \tag{10-50}$$

where z^* would now be interpreted not as genetic frailty but as frailty due to common environment.

By comparing the distributions of z and z^* and the shapes of μ and μ^*, some insights might be gained into the interaction among nature, nurture, and subsequent environment in influencing the longevity of adopted children.

DISCUSSION

The examples in this chapter have concerned survival data on MZ and DZ twins and on adopted children and their adoptive and biological parents. Many other data sets on the lifetimes of related individuals exist for humans (e.g., the Utah genealogical data base described by Bean, Chapter 15) and various other species. Furthermore, as discussed in two companion articles (Vaupel, 1990 and Larsen and Vaupel, 1989), the frailty models developed in this chapter can be extended to other life-history traits and to data on such related events as an individual's waiting times to successive conceptions. Thus there are broad possibilities for research by geneticists and demographers in developing and applying appropriate frailty models to analyze various kinds of data on related individuals and events.

Geneticists have developed a large body of knowledge about how genes are transmitted and about how genes interact with each other and with environmental influences to produce phenotypic outcomes. Evolutionary theory places strong constraints on genetic properties and recent advances in mapping genes and in understanding the effects of specific genes are leading to detailed knowledge of the nature and influence of genetic factors. The theories and empirical findings of geneticists are crucial in constructing frailty models, both in the determining the general form of such models (as illustrated in this chapter) and in determining the functional forms to be used for frailty distributions and hazard functions. In this chapter, frailty was assumed to be, say, Gamma distributed and the force of mortality was assumed to follow, say, a Gompertz trajectory.

There is some evidence that such assumptions are reasonable for some kinds of analyses, but in other cases, as discussed by Trussell and Rodríguez in this volume, it is mathematical convenience more than biological reality that dictates the assumptions made. Consequently, as suggested by Weiss (Chapter 12), an important convergent area of research for geneticists and demographers is the study of the biological underpinnings of and constraints on the functional forms used in frailty models. An example of such research is a study in progress by J. W. Curtsinger and the author of the shape of the force of mortality function for several Drosophila genotypes: survival data are being gathered on four inbred lines and their six F_1 crosses, each population consisting of 5,000 individuals raised under similar conditions.

In addition to research on developing more powerful and appropriate methods and models for frailty analysis, geneticists and demographers can engage in research in applying the frailty methods that have been developed. Although existing methods have major weaknesses and shortcomings, they may lead to some different and perhaps deeper insights. In particular, the methods adumbrated in this chapter, which represent a hybridization of two very different approaches currently used by geneticists and demographers, may help researchers in both disciplines as well as providing a basis for productive cross-disciplinary research.

For geneticists, the methods of kindred-frailty modeling provide an alternative, in the analysis of life-history traits, to customary methods of decomposition of variance. A key strength of the frailty approach is that it provides a rich description of reality as summarized by hazard functions and frailty distributions. Furthermore, frailty models highlight the interaction between nature and nurture, because genotypic and environmental influences are fundamentally intertwined in the multiplicative relationship between frailty and the baseline hazard function.

For demographers, the methods of kindred-frailty modeling provide a means for taking advantage of life-history data on related individuals and events: nearly all demographic analyses to date have treated such data as if the individuals and events were unrelated. Furthermore, as indicated above, kindred-frailty models can help demographers clarify what is meant by "frailty." In kindred-frailty models, as illustrated in this chapter, the meaning of frailty is clear. In models of frailty for unrelated individuals, it is sometimes difficult to sort out the conceptual basis for dividing the heterogeneity among individuals into a frailty component and a residual component given by the distribution of lifetimes of individuals with the same frailty. Finally, from kindred-frailty models grounded in genetic theory and findings, demographers can gain a deeper understanding of the biological constraints on the distribution of frailty and on the shape of hazard trajectories.

ACKNOWLEDGMENTS

The author thanks Joanna H. Shih for her extensive statistical and mathematical assistance, and Julian Adams, Donald A. Berry, Morris L. Eaton, Philip Hougaard, Niels Kieding, David A. Lane, Ulla M. Larsen, Thomas A. Louis, James Trussell, Kenneth Weiss, John Wilmoth and Anatoli A. Yashin for their comments.

REFERENCES

Aalen, O. O., 1987 Two examples of modelling heterogeneity in survival analysis. Scand. J. Stat. 14: 19-25.

Aalen, O. O., 1988 Heterogeneity in survival analysis. Stat. Med. 7: 1121-1137.

Beard, R. E., 1963 A theory of mortality based on actuarial, biological, and medical consideration. In *Proceedings of the International Population Conference*. International Union for the Scientific Study of Population, London, UK.

Blumen, I., M. Kogan, and P. J. McCarthy, 1955 *The Industrial Mobility of Labor as a Probability Process*. Cornell University Press, Ithaca, NY.

Clayton, D., 1978 A model for association in bivariate life tables and its application in epidemiological studies of familial tendency in chronic disease incidence. Biometrika 65: 141-151.

Clayton, D., and J. Cuzick, 1985 Multivariate generalisations of the proportional hazards model (with discussion). J. R. Stat. Soc. 148: 82-117.

Cox, D. R., and D. Oakes, 1984 *Analysis of Survival Data*. Chapman Hall, London, UK.

Falconer, D. S., 1981 *Introduction to Quantitative Genetics*. Second Edition. Longman, New York.

Feldman, M. W., and R. C. Lewontin, 1975 The heritability hang-up. Science 190: 1163-1168.

Fries, J. F., and L. M. Crapo, 1981 *Vitality and Aging: Implications of the Rectangular Curve*. W. H. Freeman, San Francisco.

Heckman, J., and B. Singer, 1984. Econometric duration analysis. J. Econometrics 24: 63-132.

Hauge, M., B. Harvald, M. Fischer, K. Gotlieb Jensen, N. Juel-Nielsen, J. Raebild, R. Shapiro, and T. Videbech, 1968 The Danish twin register. Acta Genet. Med. Gemellol. 17: 315-332.

Holm, N. V., 1983 Tvillingstudiers anvendelse til belysning af årsagsforholdene for sygdomme af kompleks aetiologi med cancer som eksempel (The use of twin studies to investigate causes of diseases with complex etiology, with a focus on cancer). Ph.D. Thesis (in Danish), 236 pages. Odense University, Denmark.

Hougaard, P., 1986a Survival models for heterogeneous populations derived from stable distributions. Biometrika 73: 387-396.

Hougaard, P., 1986b A class of multivariate failure time distributions. Biometrika 73: 671-678.

Hougaard, P., 1987 Modelling multivariate survival. Scand. J. Stat. 14: 291-304.

Kalbfleisch, J. D., and R. L. Prentice, 1980 *The Statistical Analysis of Failure Time Data*. John Wiley, New York.

Keyfitz, N., and G. Littman, 1980 Mortality in a heterogeneous population. Pop. Studies 33: 333-343.

Larsen, U. M., and J. W. Vaupel, 1989 Frailty models of Hutterite fertility. Working paper WP-89-09-1, Center for Population Analysis and Policy, University of Minnesota.

Manton, K. G., E. Stallard, and J. W. Vaupel, 1981 Methods for comparing the mortality experience of heterogeneous populations. Demography 18: 389-410.

Manton, K. G., E. Stallard, and J. W. Vaupel, 1986 Alternative models for the heterogeneity of mortality risks among the aged, J. Am. Stat. Assoc. 81(395): 635-644.

Oakes, D., 1982 A model for association in bivariate survival data. J. R. Statis. Soc. B44: 414-422.

Shepard, D. S., and P. J. Zeckhauser, 1980 Long term effects of intervention to improve survival in mixed populations. J.Chron. Dis. 33: 413-433.

Soerensen, T. I. A., G. G. Nielsen, P. K. Andersen, and T. W. Teasdale, 1988 Genetic and environmental influences on premature death in adult adoptees. N. Eng. J. Med. 315: 727-732.

Trussell, J., and T. Richards, 1985 Correcting for unmeasured heterogeneity in hazard models using the Heckman-Singer procedure. pp. 242-276 in *Sociologicai Methodology*, edited by N. B. Tuma. Jossey-Bass, London, UK.

Vaupel, J. W., 1988 Inherited frailty and longevity. Demography 25(2): 227-287.

Vaupel, J. W., 1990 Relatives' risks: Frailty models of life history data. Theor. Pop. Biol. In press.

Vaupel, J. W., K. G. Manton, and E. Stallard, 1979 The impact of heterogeneity in individual frailty on the dynamics of mortality. Demography 16: 439-454.

Vaupel, J. W., and A. Yashin, 1985a The deviant dynamics of death in heterogeneous populations. pp. 179-211 in *Sociological Methodology*, edited by N. B. Tuma. Jossey-Bass, London, UK.

Vaupel, J. W., and A. Yashin, 1985b Heterogeneity's ruses: Some surprising effects of selection on population dynamics. Am. Statist. 39: 176-185.

Vaupel, J. W., A. I. Yashin, and K. G. Manton, 1988 Debilitation's aftermath: Stochastic process models of mortality. Math. Pop. Studies 1(1): 21-48.

Wild, C. J., 1983 Failure time models with matched data. Biometrika 70: 633-641.

Yashin, A. I., K. G. Manton, and J. W. Vaupel, 1985 Mortality and aging in a heterogeneous population: a stochastic process model with observed and unobserved variables. Theor. Pop. Biol. 27: 154-175.