Demography, Vol. 27, No. 2, May 1990

Age-Specific Death Rates With Tobacco Smoking and Occupational Activity: Sensitivity to Sample Length, Functional Form, and Unobserved Frailty

Jere R. Behrman
Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6297
Robin C. Sickles
Department of Economics, Rice University, Houston, Texas 77251
Paul Taubman
Department of Economics, University of Pennsylvania, 3718 Locust Walk, Philadelphia, Pennsylvania 19104-6297

In this article, we estimate accelerated time-to-failure and proportional-hazard functions with about 100,000 members of the Dorn sample, finding greater hazards associated with smoking and some dependence on occupational variables that measure risk and physical activity. We answer three questions: (1) How sensitive are the estimates to sample length, using monthly data for the periods 1954–1969 and 1954–1980? The results differ somewhat between these sample periods. (2) How sensitive are the estimates to alternative functions for the hazard? Within a given time period, the estimates are fairly robust to specification changes in the distribution of the hazard in the accelerated time-to-failure models. (3) How sensitive are the estimates to alternative controls for unobserved frailty? Within a given sample period, the estimates are fairly robust to the allowance for parametric or nonparametric heterogeneity in the proportional-hazard models.

Introduction

Social scientists have generally analyzed the determinants of mortality by using single cross-sections or short panels.[1] The probability of dying in a given time period in these studies is usually related to a limited number of variables, such as the person's education and age. Three possible problems in earlier mortality studies pertain to the length of samples, the functional form assumption for the underlying hazard function, and unobserved heterogeneity in individual frailties. We consider each of these in turn.

Sample Length

A priori, extending the sample length may both worsen estimates in some respects and improve them in others. It may worsen the estimates because many studies in this genre depend in part on prospective data about covariates, such as whether an individual smokes, that are recorded at the start of or early in the sample, but such covariates may change over time. Prominent examples of the use of such prospective data include studies using the

NHANES data in Madans (1987) and Madans et al. (1986) and the Kaplan, Seeman, Cohen, Knudsen, and Guralnik (1987) study using the Alameda County data. Extending the sample period may also worsen the estimates because of changes in the underlying parameters due, for instance, to technological developments (e.g., changes in cancer treatment). On the other hand, extending the sample gives more observations over time, which may lead to greater precision. Extending the sample also lessens the problems of right censoring (i.e., not everyone has died by the end of most samples). In recent studies, right censoring has generally been dealt with by making distributional assumptions about the form of the survivor function, using an unobserved baseline hazard and employing a partial likelihood estimator, or by using nonparametric techniques. The results may, however, be sensitive to the specific assumptions that are made in these techniques. The impact of extending the sample on mortality estimates, given these conflicting a priori advantages and disadvantages, of course, is an empirical question. To our knowledge, however, there are not many, if any, systematic explorations of the empirical impact of extending samples.

Functional Forms for Underlying Hazard Functions

As noted, right censoring is generally dealt with by making distributional assumptions about the form of the survivor function.[2] There is limited evidence, however, about the sensitivity of mortality function estimates to alternative assumptions about the underlying functional form.

Unobserved Heterogeneity in Individual Frailty

Such heterogeneity may cause bias in the estimated duration dependence, since those who are more frail tend to die relatively early, leaving a sample dominated by those who are less frail. Efforts to control for such unobserved frailty in the mortality context, once again, are relatively few.

In Behrman, Sickles, and Taubman (1988), we presented survivor functions for the Dorn sample for the period of 1954–1969, based on accelerated time-to-failure and proportional-hazard models.[3] In the present article, we make three additional contributions in order to explore the three issues just presented. (1) We extend the Dorn coverage to 1980 and compare the results for the 1954–1969 and 1954–1980 time periods to observe the impact of extending the sample period. (2) We explore the sensitivity of our estimates to four alternative functional form assumptions for the underlying hazard functions. (3) We allow for unobserved parametric and nonparametric heterogeneity in frailties, which may be important because we do not have an exhaustive set of observed covariates, and we compare the sensitivity of our estimates to alternative approaches.

The article is organized as follows: The next section describes the sample and the general model. The following section gives the estimators. Then the results are displayed and discussed, with emphasis in turn on the empirical importance of each of the three issues. The last section is a conclusion.

The Data Set and Relation Estimated

This sample was initiated by Dorn (1958), was extended by Kahn (1966), Rogot (1974),[4] and Rogot and Murray (1980), and recently was updated through 1980 by Hrubec, Norman, and Rogot (personal communication from Z. Hrubec, February 1987). Dorn mailed a short questionnaire to 293,958 U.S. veterans, who in December 1953 held U.S. Government Life Insurance policies and had served in the armed forces between 1917 and 1940. The questionnaire (in Kahn 1966, app. E) asked how many times a day a person smoked cigarettes, cigars, and/or a pipe; how long ago he had stopped smoking; his occupation; his industry;

and his age. Since then the Veterans Administration (VA) has recorded deaths by month, year, and cause. The sample is not a random draw of arbitrary failure times because it contains veterans who were alive in 1954 and had VA insurance. The survival rates for veterans are slightly higher than for nonveterans (Behrman, Hrubec, Taubman, & Wales 1980). Therefore, inferences to a broader population of individuals should be qualified. Rogot (1974, p. 192) gave the age distribution as of 1954 and 1957, which is reproduced as Table 1.

We excluded people for whom there was no information on any of the variables studied, leaving a sample of nearly 200,000. Plots of the age-specific death rates for the 200,000 and the full sample are nearly identical. Dorn (1958) and Kahn (1966), however, showed that the sample has many fewer unskilled workers than the corresponding white male cohort.[5] We present estimates below for the overall sample and, at times, for those born before 1891 and between 1891 and 1899.[6] The estimates for these two cohorts give some insight regarding the stability of the parameters.

Epidemiologists have investigated the accuracy of the VA's information on date of death (see Beebe & Simon 1969; Cohen 1953; DeBakey & Beebe 1952). Details differ, but basically researchers took death certificates of men in the appropriate age range, matched them to military records to obtain military serial numbers, and then gave the names and numbers to the VA. In the VA population, roughly 95% of the deaths were recorded by the VA. Many of those not listed had been dishonorably discharged or were in the Army no more than four days during World War I. This high rate of coverage occurs because veterans draw pension benefits that cease at death and other benefits that commence at death (burial plots, a flag, and a burial allowance). In this sample of the VA population, the incentives to keep in touch with the VA were particularly strong, since all participants had VA life insurance in force in 1954.

The Dorn death data for 1970–1980 were compiled only recently and have not yet been analyzed for completeness, but the data through 1969 have been examined in detail. For the Dorn sample, Rogot (1974) reported that special efforts were made to check the 75,000 cases who terminated their VA insurance between 1963 and 1969. He stated, "The overall mortality follow-up, with respect to the fact of death and year of death is considered to be almost 100 percent complete" (p. 190).

The relation that we estimate may be interpreted as a health/mortality function, a central component in the standard economic analysis of individual health determinants (e.g., Anderson & Burkhauser 1985; Behrman & Wolfe 1987; Grossman 1972; Sickles & Taubman 1986; Taubman & Rosen 1982; Wolfe & Behrman 1987). We use the hazard of dying as an index of (otherwise unobserved) healthiness. It is dependent on observed health production "inputs" related to smoking and occupation. There are two possible problems with our estimates of this relation. First, though smoking and occupational characteristics

Table 1. Age and Response Distribution

	Responders			
Age	1954	1957	No reply	Total
30–34	7,421	43	2,148	9,612
35–44	16,735	7,156	4,037	27,928
45–54	10,317	1,242	2,232	13,791
55–64	137,820	26,579	31,468	195,867
65–74	25,002	13,683	5,603	44,288
75–84	1,525	523	424	2,472
30–84	198,820	49,226	45,912	293,958

are thought to be major factors in health/mortality determination, they constitute a limited representation of all health inputs; hence our estimates may be subject to omitted-variable biases if the omitted variables are correlated with the smoking and occupational variables. Second, we use the actual smoking and occupational variables as our right-side variables; but they may reflect individual choices, so their coefficient estimates may be subject to simultaneity bias.[7] Unfortunately, exogenous instruments are not available with which to explore controlling for these possible biases.[8] Good instruments must be both independent of the disturbance term and highly correlated with the variable being statistically refined. If they are not highly correlated with the variable being statistically refined, measurement bias just may replace simultaneity bias. Therefore, the gains from using instrumental variables may be less than often claimed. In any case, if simultaneity bias indeed is a major problem in our estimates, they still can be characterized as predictive relations with associations suggested by two major inputs in the health production functions rather than estimates of the structural health production function itself. We think that we make a contribution by exploring how such associations depend on the length of the sample period, the functional form, and the different ways in which we control for unobserved frailty.

We now describe our occupation and smoking variables, both of which are based on questions asked in the mid-1950s and thus are prospective variables, as are those used in the related studies referred to in the Introduction. Occupation is represented by an index for the health riskiness of the workplace and by the extent of physical activity. We have information in the sample on industry and occupation at the three-digit level. We have replicated the risk classes attached to these industry and occupation classifications when insurance premiums are set, as in the *Underwriters' Handbook* used by the life insurance industry. The risk index ranges from 1 to 7 as the riskiness of the occupation rises and has a mean of 2.2 with some variation across age cohorts. The riskiest occupations are jobs such as fire fighting and police work, whereas the least risky are jobs such as teaching. The index for the physical activity of the occupation, which was developed from the same data, ranges from 1, for sedentary, to 4, for heavy construction jobs.[9] Table 2 gives sample means and standard deviations for all of the variables we used.

The Dorn sample contains information on smoking in the form of how much tobacco the respondent used, how many years he used it, and the manner in which he used it. Since the questionnaire was administered when the respondents' ages differed, we have divided the years smoked variables by (age $-$ 10). These normalized variables have much higher t statistics than their nonnormalized counterparts in otherwise identical equations. This adjustment, of course, is subject to measurement error, since age 10 is arbitrary; but the results are not very sensitive to the exact age used for the adjustment because most respondents were at least 50 years old at the time this information was gathered. We use two tobacco-use variables. One is the number of years of occasional tobacco usage, which in this sample averages to 3% of the years individuals were more than 10 years old. The second is the number of years of regular tobacco usage, an average of 54% of the years individuals were more than 10 years old.

Estimators

We consider two classes of continuous time duration models: proportional hazard and accelerated failure time.[10] The nonnegative random variable T, time until death, has density $f(t)$ and distribution $F(t)$. Both are assumed to be absolutely continuous. Right censoring occurs in the Dorn sample, as in virtually all samples used for mortality analysis, because some of the respondents were still alive at the end of the sample frame, though of course fewer for the extended 1954–1980 sample than for the 1954–1969 sample. Let $\delta = 1$ if the duration is right censored and 0 otherwise. The Dirac censoring distribution associated

Table 2. Sample Characteristics, Dorn Sample: Summary Statistics for Different Cohorts

	Cohort					
	Born before 1890		Born 1890–1899		All	
Variable	Mean	S.D.	Mean	S.D.	Mean	S.D.
Smoking						
Proportion of years used tobacco occasionally	.0330	.142	.0318	.143	.0307	.138
Proportion of years used tobacco regularly	.495	.327	.546	.353	.543	.343
Occupation						
Activity index	2.17	.848	2.13	.830	2.19	.841
Risk index	2.01	.158	1.19	.156	2.20	.170
Region						
South	.258	.438	.240	.425	.249	.430
Northeast	.300	.458	.307	.461	.290	.452
North Central	.264	.441	.289	.452	.271	.442
No. of observations	12,822		69,991		101,511	
% censored						
Through 1969	40.4		63.7		65.5	
Through 1980	8.5		22.9		33.3	

Note: Number of years of occasional and regular tobacco use is divided by age at survey date − 10 years.

with realizations on δ is assumed to be independent of the survival time and is functionally independent of the survival distribution.

In the *proportional-hazard model*, covariates are allowed to affect the hazard in a rather restrictive fashion, that is,

$$\lambda(t \mid x) = \lambda_0(t) \exp(x'\beta), \tag{1}$$

where $\lambda_0(t)$ is the baseline hazard.

The Cox (1975) estimator of the proportional-hazard model avoids distributional assumptions about the baseline hazard by maximizing the partial likelihood function. The resulting estimates of the covariate effects are still consistent and asymptotically normal. Unobserved heterogeneity is not controlled for explicitly in the Cox model but is of major interest to us. Heckman and Singer (1984) proposed a more general proportional-hazard model that allows for arbitrary duration dependence and heterogeneity. The model is

$$\lambda(t \mid x, \theta) = \exp\left[x'\beta + \frac{t^{\eta_1} - 1}{\eta_1}\gamma_1 + \frac{t^{\eta_2} - 1}{\eta_2}\gamma_2 + \theta\right], \tag{2}$$

where θ represents heterogeneity or frailty differences among individuals (Manton, Stallard, & Vaupel 1986). We make an assumption that is common in the random effects literature (Heckman & Singer 1984; Manton et al. 1986)—that individual specific frailty differences are not correlated with observed covariates. This assumption would be less tenable were our covariates time varying. Because the study is, however, both retrospective (how much one smoked in the past) and prospective (the respondents had not died yet at the time of collection of smoking and occupational information), with observations recorded only once and often

at midlife, the strength of the correlation between our occupational characteristics, smoking, and regional variables and frailty is certainly less than it would be if the observations were recorded, for example, every year. In our empirical implementation, we use both parametric and nonparametric representations of frailty and let $\eta_1 \rightarrow 0$ and set $\gamma_2 = 0$, which collapses this hazard to the Weibull model, with heterogeneity and duration affecting the hazard in a constant, proportional fashion.

The *accelerated failure time model* specifies the natural logorithm of date of death as a linear function of the covariates, $\log T = x\beta + \sigma\varepsilon$, where ε is a random disturbance and σ is a scale parameter. Failure time can be written as $T = \exp\{x\beta\}T_0^{\sigma}$, where T_0 is an event time drawn from a baseline duration distribution for which the covariates are zero.

For our maximum likelihood estimates of accelerated failure time models, we use four different specifications for the baseline duration distribution: Weibull, lognormal, log logistic, and generalized gamma.[11] The hazard functions corresponding to these different baseline duration distributions are, respectively,

$$\lambda(t \mid x) = \sigma^{-1} \exp(-x\beta/\sigma)t(\sigma^{-1} - 1); \tag{3}$$

$$\lambda(t \mid x) = \{\exp[-(\log t - x\beta)^2/2\sigma^2]/\sqrt{2\pi}\sigma t\} \Big/ \int_{\xi_0}^{\infty} f(\xi)d, \tag{4}$$

where $f(\xi)$ is the standard normal density function and $\xi_0 = (\ln t - x\beta)/\sigma$;

$$\lambda(t \mid x) = \frac{\sigma^{-1} \exp(-x\beta/\sigma)t^{\sigma^{-1}-1}}{1 + \exp(-x\beta/\sigma)t^{\sigma^{-1}-1}}; \tag{5}$$

$$\lambda(t \mid x) = \frac{|\gamma|/\sigma t \{[\exp(-x\beta)t]^{\gamma/\sigma}/\gamma^2\}^{1/\gamma^2} \exp\{-[\exp(-x\beta)t]^{\gamma/\sigma}/\gamma^2\}}{\delta\Gamma(1/\gamma^2) + (-1)^{\delta}\Gamma\{1/\gamma^2, [\exp(-x\beta)t]^{\gamma/\sigma}/\gamma^2\}}, \tag{6}$$

where $\delta = 1$ if $\gamma < 0$ and 0 if $\gamma > 0$, $\Gamma(\alpha_1)$ is the complete gamma function, $\Gamma(\alpha_1, \alpha_2)$ is the incomplete gamma function, and γ is the distribution's shape parameter. In the Weibull accelerated failure time model, the effect of a change in x on $\ln \lambda(t \mid x)$ is $-x\beta/\sigma$, which differs from the same effect in the Cox and Weibull proportional-hazard models by its sign and the scale parameter, σ.

Results

Sensitivity to Sample Length

The first question that we consider is, How stable are the results as the sample period lengthens from 1954–1969 to 1954–1980? With such an extension, right censoring becomes less important, but possible problems with using prospective data and technological and behavioral changes increase. Table 3 contains the Cox partial likelihood estimates for the period of 1954–1969 in a sample of 101,511 persons, nearly two-thirds of whom were alive at the start of 1970.[12] We present the results in terms of elasticities calculated at the sample means.[13] These results are described in detail in Behrman et al. (1988). The proportional-hazard function for all cohorts indicates a greater association with hazard of dying: the more one smokes, the riskier one's usual occupation, the less physical activity in one's longest occupation,[14] and the higher one's likelihood of residing outside of the West.

Table 4 presents the corresponding results for data through 1980. With the longer sample, there are some changes in coefficient estimates and their t statistics. We cannot tell whether the coefficient changes reflect (a) greater measurement error for the longer time period because the right-side variables were measured prospectively in the mid-1950s,

Table 3. Dorn Sample, 1954–1969, Cox Partial Likelihood Proportional-Hazard Model

Variable	Cohort estimate (t)		
	Born before 1891	Born 1891–1899	All[a]
Smoking			
Proportion of years used tobacco occasionally	.532 (.562)	.287 (5.78)	.212 (4.44)
Proportion of years used tobacco regularly	.454 (11.68)	.624 (30.3)	.577 (32.8)
Occupation			
Activity index	−.013 (−.26)	−.022 (−2.21)	−.021 (−2.43)
Risk index	.0013 (.135)	.014 (2.69)	.020 (4.71)
Region			
South	.057 (1.63)	.078 (3.87)	.055 (3.30)
Northeast	.014 (.404)	.077 (3.96)	.027 (1.65)
North Central	.034 (.969)	.051 (2.58)	.0088 (.534)
Log likelihood	−69,998	−278,404	−387,619
χ^2	192.8	1,522.8	2,113.2

[a] Includes people born in the period 1900–1924 in addition to those born before 1900.

(b) the greater censoring problem in the shorter period for Table 3, or (c) secular changes over time, such as improvements in treatments for tobacco-usage–related disease. But the combination of such considerations does seem to make some difference. We first discuss the results for all cohorts and then summarize some of the results for the cohorts born before 1891 and during 1891–1899.

For all cohorts the estimated impact of occasional tobacco use shown in Table 4 is slightly less than that shown in Table 3, though not significantly so. The coefficient is estimated more precisely, however, and the hazard is still increased by about 56% by the regular use of tobacco. The usual occupation now yields an insignificant coefficient for physical activity. Its risk effect is somewhat stronger (though not significantly so), with a much larger t statistic. Those in the most risky occupations have 16.8% higher hazard than those in the least risky occupations.[15] The coefficients on the 1950s residence data also differ in Tables 3 and 4: in comparison with those from the West in Table 4, inhabitants of the South have a much higher hazard and those from the Northeast now have a statistically significant coefficient.

The accelerated time-to-failure models estimates through 1969 are in Table 5. For the gamma estimates, time to death is shorter: the more one smokes, if one lives in the South, the less one's physical activity and the riskier one's job.[16]

The longer sample findings are given in Table 6. These differ somewhat from those for the shorter period. The tobacco-use variables have smaller coefficient estimates in absolute value in Table 6 than in Table 5, and the differences in the coefficient estimates in these two tables for regular tobacco use are significant. For the regular tobacco-use variable, the

Table 4. Dorn Sample, 1954–1980, Cox Partial Likelihood Proportional-Hazard Model

Variable	Cohort estimate (t)		
	Born before 1891	Born 1891–1899	All[a]
Smoking			
Proportion of years used tobacco occasionally	.171 (2.34)	.189 (5.46)	.187 (6.14)
Proportion of years used tobacco regularly	.425 (13.6)	.550 (38.8)	.564 (45.0)
Occupation			
Activity index	−.0014 (−.099)	−.0062 (−.86)	−.0087 (−1.42)
Risk index	.012 (1.59)	.016 (4.28)	.028 (8.83)
Region			
South	.063 (2.24)	.090 (6.28)	.070 (5.73)
Northeast	.014 (.52)	.060 (4.35)	.026 (2.19)
North Central	.059 (2.10)	.028 (2.04)	.0072 (.60)
Log likelihood	−101,289	−530,670	−716,098
χ^2	203.9	1,699.5	2,329.5

[a] Includes people born in the period 1900–1924 in addition to those born before 1900.

elasticities are in the −.06 to −.07 range, with t statistics in the high 40s. The corresponding numbers in Table 5 are about −.08. The occupational variables, in contrast, have some larger and some smaller coefficients in Table 6 than in Table 5, with some of the differences statistically significant. The activity variable consistently has smaller coefficient estimates and smaller t values in Table 6. The risk variable consistently has larger t statistics and, except for the log logistic case, larger point estimates (in absolute values) in Table 6. This may indicate that the advantage of prior exercise atrophies quickly when one stops the activity, whereas the effects of work-related health hazards do not vanish as quickly. The regional controls have larger t values in two-thirds of the cases.

The accelerated failure-time estimates are presented in Tables 7 and 8 for the two age cohorts born before 1900 for the periods through 1969 and 1980, respectively. Here we use only the gamma baseline failure time, our preferred distribution based on the maximized value of the log likelihood. For both time periods, tobacco usage is important, but the results are stronger for the later-born cohort in both tables. The absolute magnitudes of the coefficients are greater, and in three of the four cases, the t values are larger for the longer period.

For the longer time period, the risk index has significant coefficient estimates for both cohorts and has about the same elasticity. In the shorter time period, it has a significant coefficient estimate for only the older cohort, with an elasticity similar to that in Table 8. The activity index has a significant coefficient estimate only for the younger cohort in both time periods; this estimate is much larger for the shorter time period, for which, once again, measurement error and parametric changes are less important.

Table 5. Dorn Sample, 1954–1969, Accelerated Time Models, All Cohorts

Variable	Weibull estimate	Lognormal estimate	Log logistic estimate	Gamma estimate
Intercept	5.20	5.15	5.14	5.15
	(1,750)	(1,666)	(1,718)	(1,572)
Smoking				
Proportion of years	−.032	−.029	−.033	−.029
used tobacco occasionally	(−5.73)	(−5.22)	(5.99)	(−5.07)
Proportion of years	−.080	−.081	−.083	−.081
used tobacco regularly	(−35.0)	(−35.2)	(−36.8)	(−35.0)
Operation				
Activity index	.003	.0044	.0037	.0047
	(2.78)	(3.85)	(3.33)	(4.00)
Risk index	−.002	−.0048	−.0031	−.0054
	(−4.34)	(−8.28)	(−5.46)	(−8.99)
Region				
South	−.007	−.0055	−.0073	−.0051
	(−3.43)	(−2.43)	(−3.31)	(−2.20)
Northeast	−.004	.00002	−.0042	.00113
	(−2.02)	(.03)	(−1.98)	(.51)
North Central	−.003	.0015	−.0023	.00250
	(−1.35)	(.67)	(−1.08)	(1.11)
Scale	.128	.186	.104	.192
	(228.7)	(247.3)	(226)	(150.1)
Shape	—	—	—	−.122
				(−6.58)
Log likelihood	−204,263	−201,959	−202,457	−201,931
χ^2	1,724	1,572	1,778	1,855

Note: *t* statistics are given in parentheses.

Sensitivity to Functional Form Choices

In this section we consider the sensitivity of our results to choices regarding functional forms. We examine various specifications for the accelerated failure-time model. The Cox and Weibull proportional-hazard models differ from the Weibull accelerated hazard by a scale factor. It is of interest also to consider robustness and goodness of fit in both the shorter and longer periods.

Consider first the results for the period 1954–1969 in Table 5 and those for 1954–1980 in Table 6. In both cases, given the sample length, the patterns across functional forms are quite similar. All four baseline hazards fit the data about equally well in terms of the maximized values of the log likelihood function, with the gamma model slightly the best fitting and the lognormal the worst. For the tobacco variables, the point estimates and *t* statistics are very similar in the four columns. The occupational and regional coefficient estimates have much greater variation, though the signs of the significant estimates are the same across the estimates.

A comparison of the Cox proportional-hazard (Table 3 and 4, col. 3) and the Weibull accelerated failure-time models (Tables 5 and 6, col. 1) indicates that, as expected, coefficient signs are reversed and the coefficients in the Cox model equal those in the Weibull model divided by minus the estimated scale parameter. The comparability of covariate estimates across different functional forms for the death hazard occurs despite the substantial differences

Table 6. Dorn Sample, 1954–1980, Accelerated Time Models, All Cohorts

Variable	Weibull estimate	Lognormal estimate	Log logistic estimate	Gamma estimate
Intercept	5.15	5.10	5.10	5.13
	(2,899)	(2,473)	(2,576)	(2,599)
Smoking				
Proportion of years used tobacco moderately	−.021	−.021	−.023	−.022
	(−6.44)	(−5.55)	(−6.33)	(−6.22)
Proportion of years used tobacco heavily	−.064	−.071	−.071	−.068
	(−46.5)	(−46.0)	(−48.1)	(−47.0)
Occupation				
Activity index	.0011	.0027	.0020	.0016
	(1.63)	(3.39)	(2.70)	(2.25)
Risk index	−.003	−.0056	−.0040	−.0038
	(−8.65)	(−14.2)	(−10.3)	(−10.3)
Region				
South	−.008	−.0055	−.007	−.0074
	(−5.98)	(−3.53)	(−4.68)	(−5.18)
Northeast	−.003	.0015	−.0022	−.0021
	(−2.43)	(1.0)	(−1.55)	(−1.5)
North Central	−.001	.0035	.0001	.00016
	(−.846)	(2.29)	(.073)	(.12)
Scale	.109	.146	.082	.126
	(321.6)	(357.3)	(361)	(229.3)
Shape	—	—	—	.545
				(29.2)
Log likelihood	−338,697	−339,291	−338,209	−337,936
χ^2	2,532	2,227	2,601	2,789

Note: *t* statistics are given in parentheses.

in the highly nonlinear functional forms for the accelerated time-to-failure and proportional-hazard models. The age profiles for the logarithm of the hazard (relative to which our estimated covariate effects are calibrated) are, however, fairly flat because the hazard has a strong upward trend with age. The nonlinear covariate estimates are evaluated at the mean of the sample, but there may be little nonlinearity at this point in the sample space; thus our finding of robustness across different functional forms may be an artifact of the pattern of our sample death rates.

Unobserved Heterogeneity in Frailty

The previous results assume no individual differences in unobserved frailty. Manton et al. (1986) argued that one can expect heterogeneity in a mortality hazard. Heterogeneity can be modeled as following either a particular parameterization as in Manton et al. or a nonparametric distribution as in Heckman and Singer (1984). Our third contribution is to explore whether control for such heterogeneity makes a difference.

To investigate the importance of heterogeneity, we used a Weibull proportional-hazard model. In Table 9 we present estimates for the earliest cohort. The first column presents the Weibull with no allowance for heterogeneity. This column has elasticities similar in sign and magnitude to those obtained with the Cox model in Table 4, column 1, though the impact of the occupational risk index is much larger and statistically significant in Table 9.

Table 7. Dorn Sample, 1954–1969, Accelerated Time Model, Gamma Baseline Failure Time, Two Cohorts

Variable	Cohort estimate (t)	
	Born before 1891	Born 1891–1899
Intercept	5.03	4.97
	(1,090)	(1,435)
Smoking		
Proportion of years used tobacco occasionally	−.012	−.015
	(−1.65)	(−3.06)
Proportion of years used tobacco regularly	−.029	−.043
	(−9.23)	(−21.3)
Occupation		
Activity index	.0016	.0042
	(1.05)	(3.99)
Risk index	−.0016	−.0006
	(−1.97)	(−1.11)
Region		
South	−.00069	−.0058
	(−.236)	(−2.74)
Northeast	.00052	−.00447
	(.18)	(−2.19)
North Central	−.0054	−.00296
	(−1.86)	(−1.45)
Scale	.11	.174
	(85.4)	(171.2)
Shape	−1.79	−2.36
	(−37.7)	(−74.9)
Log likelihood	−37,930	−137,900
χ^2	82.5	334.0

The Heckman–Singer nonparametric model using two points of support,[17] given in column 2, fits the data a bit better. The coefficients on the variables that were significant in column 1 remain significant, with those on regular tobacco use and on the ln of duration increasing significantly.

The estimates assuming that heterogeneity follows a normal distribution are given in column 3.[18] This estimator fits the data slightly worse than the Heckman–Singer model and slightly better than the model with no allowance for heterogeneity. The t statistics tend to be lower, with that on occasional tobacco use falling to 1.75. The coefficients are very similar to those in the first two columns, though a little closer to the ones in the first column.

Table 10 contains the 1891–1899 cohort Weibull proportional-hazard results, which highlight differences in very large samples between specifications with and without heterogeneity. The model without heterogeneity yields parameter estimates somewhat different from those obtained with the Cox model (Table 4). In Table 10 the activity and risk indices have much bigger impacts (in absolute value), but the regional and tobacco-use variables have smaller effects (some significantly so). The allowance for heterogeneity[19] improves the fit slightly, but it does not alter the overall impression concerning the magnitude of the impact of the coefficients. For example, the regular tobacco-use coefficient estimate goes from .45 to .48 and the activity index's coefficient estimate goes from −.084 to −.098, but the differences are relatively minor.

Table 8. Dorn Sample, 1954–1980, Accelerated Time Model, Gamma Baseline Failure Time, Two Cohorts

Variable	Cohort estimate (t)	
	Born before 1891	Born 1891–1899
Intercept	5.09	5.09
	(1,271)	(2,064)
Smoking		
Proportion of years	−.015	−.023
used tobacco occasionally	(−1.64)	(−5.76)
Proportion of years	−.041	−.064
used tobacco regularly	(−14.3)	(−39.5)
Occupation		
Activity index	.0011	.0019
	(.77)	(2.23)
Risk index	−.0017	−.0015
	(−2.24)	(−3.39)
Region		
South	−.045	−.010
	(−1.64)	(−5.72)
Northeast	−.00066	−.0078
	(−.245)	(−4.76)
North Central	−.0056	−.0042
	(−2.06)	(−2.55)
Scale	.10	.137
	(150.4)	(262.8)
Shape	−.29	−.141
	(−9.44)	(−8.30)
Log likelihood	−52,402	−245,401
χ^2	119	527

Comparison of Tables 9 and 10 shows that the activity index has an insignificant coefficient estimate for the older cohort but a significant one for the younger group. The risk index has a somewhat larger coefficient in the older cohort, though not significantly so. The smoking effects are somewhat larger for the older cohort (significantly so for regular use) when heterogeneity is allowed.

We also use the procedure proposed by Lancaster (1985) to test for unobserved heterogeneity. The test is based on generalized residuals (Cox & Snell 1968) and is a special case of the information matrix test for misspecification proposed by White (1982) and expanded in the context of duration models by Chesher (1984). We consider here only the frailty differences among individuals that affect the constant term, not the entire set of parameters such as in the tests of White and Chesher. The statistical test is outlined in the Appendix.

Computational constraints limited the use of this test to the first cohort of the Dorn sample, in which we have approximately 13,000 individuals. The results are summarized in Table 11 for representatives of three classes of models.

Although results from the three classes of models suggest the presence of heterogeneity, the evidence indicates that the modeling of unobserved heterogeneity directly in a proportional-hazard setting may not be as important as allowing covariates to affect the hazard in the highly nonlinear way that the gamma accelerated failure-time model allows. Comparisons of alternative methods to reduce the impact of distributional assumptions about the

Table 9. Weibull Proportional Hazard With and Without Heterogeneity, Dorn Sample, 1954–1980, Cohort Born Before 1891

Variable	Estimate (t)		
	No heterogeneity	Nonparametric heterogeneity[a]	Normal heterogeneity
Smoking			
Proportion of years used tobacco occasionally	.180 (2.74)	.196 (2.51)	.150 (1.75)
Proportion of years used tobacco regularly	.458 (15.9)	.542 (15.7)	.424 (3.21)
Occupation			
Activity index	.00435 (.08)	−.00777 (.12)	−.0380 (.61)
Risk index	.0975 (2.03)	.123 (2.12)	.134 (1.95)
Region			
South	.0668 (2.69)	.0775 (2.54)	.0501 (1.54)
Northeast	.0189 (.76)	.0207 (.69)	.008 (.29)
North Central	.0695 (2.79)	.0759 (2.50)	.0593 (1.77)
ln duration	9.87 (112.2)	11.2 (117.9)	9.93 (97.5)
Log likelihood	−52,019	−50,427	−50,662
χ^2	129	147	139

[a] We use two points of support.

form of heterogeneity have generally been made in the context of proportional-hazard models with few covariates and limited sample sizes. These last results suggest that renewed focus might be placed on generalizing the specification of the conditional hazard as well as on limiting the distributional impacts of unobserved heterogeneity.

Conclusion

Our study uses up to 100,000 individuals, more general functional forms for the hazard distribution than in most previous work, and seven covariates. We present results that permit comparison of estimators of hazard functions as we vary the length of the sample period, functional forms for the baseline hazard, and the treatment of heterogeneity.

Our first concern is what happens with longer samples and the associated fewer censored observations and greater precision and the concomitant dependence on the assumed stability of prospective variables over a longer period and of the underlying parameters. A longer time period with fewer censored observations tends to increase the precision of the estimated coefficients and sometimes significantly changes their magnitude. Over the longer time period, the estimated effects of tobacco smoking and occupational activity, both reported in the mid-1950s, on life expectancy tend to be reduced, but those of occupational risk are increased in some cases. These changes may have occurred because of chance, the reduction in censored observations, the more precise fitting of hazards to longer time periods, or the

Table 10. Weibull Proportional Hazard With and Without Heterogeneity, Dorn Sample, 1954–1980, Cohort Born 1891–1899

Variable	Estimate (t)	
	No heterogeneity	Nonparametric heterogeneity
Smoking		
Proportion of years used tobacco occasionally	.164 (4.40)	.158 (3.74)
Proportion of years used tobacco regularly	.448 (28.7)	.475 (26.7)
Occupation		
Activity index	−.084 (2.73)	−.098 (−2.95)
Risk index	.073 (2.73)	.094 (3.15)
Region		
South	.040 (2.67)	.035 (1.92)
Northeast	.028 (1.93)	.013 (.80)
North Central	.009 (.58)	.0027 (.17)
ln duration	7.01 (202)	8.07 (212)
Log likelihood	−247,002	−241,613
χ^2	501	519

growing irrelevance of data on tobacco use and occupation collected in the mid-1950s to explain developments in the 1970s. The last possibility, of course, would be detrimental to the use of prospective studies undertaken without periodic remeasurement of right-side variables, which may change substantially over time, some examples of which are given in the Introduction. Because in some cases the longer period estimates are larger in absolute magnitude, however, the last possibility is probably not the only relevant one.

Our second focus concerns alternative specifications of the baseline hazard functions. We explore four alternatives. Formal testing of the statistical dominance of a particular nonnested specification among the array of models that we consider is not possible. We find, though, that variations in assumed hazard functions do not make much difference in parameter estimates, given the length of sample period.[20]

Table 11. Specification Error Tests

Model	Score statistic
Cox's proportional hazard (partial likelihood)	−15.9
Nonparametric maximum likelihood estimate with Weibull hazard (two points of support)	−11.1
Gamma accelerated hazard	−2.61

Our third concern relates to the importance of unobserved heterogeneity. The introduction of unobserved heterogeneity and nonparametric methods yields coefficients that are generally similar to those obtained with parametric techniques and no heterogeneity. Thus the hazard estimates are robust to control for such heterogeneity. Furthermore, score tests of heterogeneity indicate that more general accelerated time-to-failure models with no heterogeneity are not dominated by proportional-hazard models that control for heterogeneity.

The substantive findings, finally, are in accordance with previous results. Cigarette smoking, less occupational activity, and greater occupational risk are associated with shorter life spans.

Notes

[1] For example, see Rosen and Taubman (1984), Kitagawa and Hauser (1973), Duleep (1986), and references therein.

[2] Alternative statistical methods have recently been developed to estimate a baseline survivor distribution (what percentage of the sample is alive at a given age), which shifts with covariates. For example, see Cox (1972), Kalbfleisch and Prentice (1980), Manton, Stallard, and Vaupel (1986), and Vaupel, Manton, and Stallard (1979).

[3] We also presented such functions for the Retirement History Survey (RHS) sample for the 1969–1977 period.

[4] Rogot is our source for much of the following description.

[5] The sample has only a small number of nonwhites.

[6] We do not give separate estimates for those born after 1899 because of the much greater right-censoring problem for this group.

[7] That is, expected life (and thus mortality) may be determined simultaneously with tobacco smoking and occupational characteristics. If such simultaneity is not controlled, the estimated coefficients of tobacco use and occupational characteristics may be biased because they may reflect underlying factors (e.g., tastes) that affect all three outcomes. If so, they will not be unbiased estimates of the effect, say, of changing tobacco use on mortality, though they will still indicate the predictive association in the sample between tobacco use and occupational characteristics on the one hand and mortality on the other. See any standard econometrics text (e.g., Maddala 1977).

[8] In future work we hope to develop a set of instruments to link to the Dorn data.

[9] This index was developed from the physical capacity classification in U.S. Employment Service (1961).

[10] The only family of distributions closed under both multiplication of failure time and multiplication of the hazard function by an arbitrary nonzero constant is the Weibull family (Kalbfleisch & Prentice 1980), which includes the exponential and Weibull log-linear regression models.

[11] We also used the exponential, which has a constant hazard over time. Since the actual hazard rates rise with age over the part of the life cycle considered, we do not show these results.

[12] This is the largest number that we can handle on the smallest computer system we are using, an IBM 3083 with 16 megabytes of memory.

[13] The elasticity of x with respect to y is the percentage change in x for a given percentage change in y. The estimates of .454 in the second row and first column indicate, for example, that the hazard for the cohort born before 1891 increases 4.5% if there is a 10% increase in the proportion of years in which tobacco was used regularly.

[14] This physical activity result is also found in Moore and Haywood (1988), using the National Longitudinal Survey of older men.

[15] These calculations reflect the full range of values possible for these variables. See the discussion of data above.

[16] Most results hold for the other baseline hazard distributions, but not all. The other regional effects are significantly nonzero for the lognormal case. Increased physical activity lengthens life expectancy significantly in the other cases.

[17] We have one estimate using three points of support for this cohort. Most estimates are very similar to those in Table 9, but the coefficient on regular tobacco use rises to .58. The third support increases CPU time on the NEC SX-2 supercomputer used in these analyses by about 50%. For a sample of about 13,000 observations, the running time is about 4½ hours.

[18] Although distributions other than the normal can be specified in the CTM program (Yi et al. 1986), the extreme computational demands implied by the relatively large number of covariates and sample observations meant that only a limited number of parametric heterogeneity distributions could be considered. These included the lognormal and gamma. Due to convergence problems and CPU constraints, however, the only parametric heterogeneity distribution on which we report estimates is the normal. Although the normal is not a flexible distribution, its ubiquity in applied work and its justification therein provide us with a rationale for its selection for this study.

[19] We consider in Table 10 only nonparametric heterogeneity, not normal heterogeneity.

[20] It is possible that in future work further developments in knowledge about biological properties, as suggested in Manton and Stallard (1979, 1981), may help determine the appropriate functional form.

Appendix: Specification Error Tests

To outline the test, first define the generalized error of the integrated hazard as

$$\varepsilon(t) = \int_0^t \lambda(\tau \mid x)\, d\tau. \tag{A.1}$$

Maximum likelihood estimates of the generalized residuals are based on the estimated parameters and give

$$\hat{\varepsilon}(t) = \int_0^t \hat{\lambda}(\tau \mid x)\, d\tau. \tag{A.2}$$

Consider the hazard function that is shifted by heterogeneity of the form $v = \exp\{\theta\}$. The conditional survivor function becomes

$$S(t \mid x, v) = \exp\left\{ -\int_0^t \lambda(\tau \mid x, v)\, d\tau \right\} \tag{A.3}$$

or

$$S(t \mid x, v) = \exp\{-v\varepsilon(t)\}, \tag{A.4}$$

where by assumption θ is assumed to be distributed independently of x. The unconditional survivor function is

$$S(t) = \int g(v) \exp\{-v\varepsilon(t)\}\, dv = E[\exp\{-v\varepsilon(t)\}], \tag{A.5}$$

where g is the density function of v. If we scale the constant term in the hazard function so that $\text{var}(v) = \sigma^2$ and $E(v) = 1$ and approximate equation (A.5) by a second-order Taylor series about $v = 1$, then

$$S(t) \approx S^*(t)[1 + (\sigma^2/2)\varepsilon(t)^2], \tag{A.6}$$

where $S^*(t)$ is the survivor function if there is no neglected heterogeneity [i.e., $S^*(t) = \exp\{-\varepsilon(t)\}$]. Differentiating equation (A.6) gives the approximation to the unconditional duration density:

$$h(t) \approx h^*(t)\{1 + \sigma^2[\varepsilon(t)^2 - 2\varepsilon(t)]/2\}, \tag{A.7}$$

where $h^*(t)$ is the duration density if there is no neglected heterogeneity. Thus $h(t)$ and $h^*(t)$ differ by the variability in frailty differences among individuals. Taking natural logs of equation (A.7) gives

$$\ln h(t) = \ln h^*(t) + \ln\{1 + \sigma^2[\varepsilon(t) - 2\varepsilon(t)]/2\}. \tag{A.8}$$

Let the parameter vector characterizing the hazard be (β, σ^2) and let $\hat{\beta}$ be the maximum likelihood estimate of β when $\sigma^2 = 0$. Then the derivatives of $\ln h(t)$ at the point $(\hat{\beta}, 0)$ are zero except for the term

$$\left. \frac{\partial \ln h(t)}{\partial \sigma^2} \right|_{(\hat{\beta},0)} = [\hat{\varepsilon}(t)^2 - 2\hat{\varepsilon}(t)]/2. \tag{A.9}$$

Under the null hypothesis of no neglected heterogeneity, the expression in equation (A.9) has an expectation of zero. The score test of the null hypothesis that $\sigma^2 = 0$ is then based on the quantity

$$Q^* = \left(\sum_{i=1}^{N} \hat{\varepsilon}_i^2 - 2 \sum_{i=1}^{N} \hat{\varepsilon}_i \right) \Big/ N, \tag{A.10}$$

where N is the number of observations in the sample, and on the estimated asymptotic variance of equation (A.9). The asymptotic variance of equation (A.9) is obtained by evaluating the information matrix for the approximation of $h(t)$ in equation (A.7) on the basis of the parameters (β, σ^2). Reordering parameters, the information matrix at the point $(0, \beta)$ is

$$I(0, \beta) = \begin{vmatrix} 2 & l^T \\ l & i \end{vmatrix}, \tag{A.11}$$

where $l = (1, 1, \ldots, 1)$ and i is the information matrix based on the density $h^*(t)$. The variance of the score statistic is proportional to the (1, 1)th element in $I(0, \beta)^{-1}$. The test statistic is

$$T = (\sqrt{N}/2)[(\Sigma \hat{\varepsilon}_i^2 - 2\Sigma \hat{\varepsilon}_i)/N]/(2 - l'_i{}^{-1} l)^{1/2}, \tag{A.12}$$

which under appropriate conditions converges in distribution to a standardized normal.

Acknowledgments

An earlier version of this article was presented at the 1987 European Econometric Society Meetings, Copenhagen, and at the 1987 Annual Meeting of the Population Association of America, Chicago. The work was supported by National Institute on Aging Grant 1-RO1-AG-05384. Computer services were made available through the Houston Area Research Center's NEC/SX-2 supercomputer, one of the world's fastest uniprocessors. We thank Abdo Yazbeck and Keun Huh for their valuable computational assistance, James Heckman and James Walker for making available to us an early version of their CTM software (Yi et al. 1986), and two anonymous referees for useful comments.

[*Received July 1989. Revised November 1989.*]

References

Anderson, K., & Burkhauser, R. (1985). "The Retirement Health Nexus: A New Measure of an Old Puzzle." *Journal of Human Resources*, 20, 315–333.

Beebe, G., & Simon, A. (1969). "Ascertainment of Mortality in the U.S. Veteran Population. *American Journal of Epidemiology*, 89, 636–643.

Behrman, J. R., Hrubec, Z., Taubman, P., & Wales, T. J. (1980). *Socioeconomic Success: A Study of the Effects of Genetic Endowments, Family Environment and Schooling*. Amsterdam: North-Holland.

Behrman, J. R., Sickles, R., & Taubman, P. (1988). "Age Specific Death Rates." In *Issues in Contemporary Retirement*, eds. E. Lazear & R. Ricardo-Campbell. Stanford, CA: Hoover Institution Press, pp. 162–190.

Behrman, J. R., & Wolfe, B. L. (1987). "How Does Mother's Schooling Affect Family Health, Nutrition, Medical Care Usage, and Household Sanitation. *Journal of Econometrics*, 36, 185–204.

Chesher, A. D. (1984). "Testing for Neglected Heterogeneity." *Econometrica*, 52, 865–872.

Cohen, B. (1953). "Methodology of Record Follow-up Studies on Veterans." *American Journal of Public Health*, 43, 1292–1298.

Cox, D. (1972). "Regression Models and Life Tables." *Journal of the Royal Statistical Society*, Ser. B, 33, 187–202.

———. (1975). "Partial Likelihood." *Biometrika*, 62, 269–276.

Cox, D. R., & Snell, E. J. (1968). "A General Definition of Residuals." *Journal of the Royal Statistical Society* 131, 265–279.

DeBakey, M., & Beebe, G. (1952). "Medical Follow-up Studies on Veterans." *Journal of the American Medical Association*, 182, 1103–1109.

Dorn, H. (1958). "The Mortality of Smokers and Non-smokers." In *Proceedings of the Social Statistics Section, American Statistical Association*. Alexandria, VA: American Statistical Association, pp. 34–71.

Duleep, H. (1986). "The Socioeconomic Determinants of Mortality: The Role of Income." *Journal of Human Resources*, 21, 238–251.

Grossman, M. (1972). *The Demand for Health: A Theoretical and Empirical Investigation*. New York: Columbia University Press.

Heckman, J., & Singer, B. (1984). "A Method for Minimizing the Impact of Distributional Assumptions in Econometric Models for Duration Data." *Econometrica*, 52, 271–320.

Kahn, H. (1966). "The Dorn Study of Smoking and Mortality Among U.S. Veterans: Report of 8½ Years of Observation." In *Epidemiological Approaches to the Study of Cancer and Other Diseases*, ed. W. Haenszel, Monograph 19. Bethesda, MD: U.S. Public Health Service, National Cancer Institute, pp. 1–125.

Kalbfleisch, J., & Prentice, R. (1980). *Statistical Analysis of Failure Time Data*. New York: Wiley.

Kaplan, E., Seeman, T. E., Cohen, R. D., Knudsen, L. P., & Guralnik, J. (1987). "Mortality Among the Elderly in the Alameda County Study: Behavioral and Demographic Risk Factors." *Journal of Public Health*, 77, 307–312.

Kitagawa, E., & Hauser, P. (1973). *Differential Mortality in the United States of America: A Study in Socioeconomic Epidemiology*. Cambridge, MA: Harvard University Press.

Lancaster, T. (1985). "Generalized Residuals and Heterogenous Duration Models." *Journal of Econometrics*, 28, 155–169.

Madans, J. (1987). "10 Years After NHANES 1: Report of Initial Followup, 1982–84." Office of Analysis and Epidemiology Program, National Center for Health Statistics (mimeograph).

Madans, J. H., Cox, C. S., Kleinman, J. C., Makuc, D., Feldman, J. J., Finucane, F. F., Barbano, H. E., & Cornoni-Huntley, J. (1986). "10 Years After NHANES 1: Mortality Experience at Initial Followup, 1982–1984." *Public Health Reports*, 101, 465–473.

Maddala, G. S. (1977). *Econometrics*. New York: McGraw-Hill.

Manton, K. G., Stallard, E., & Vaupel, J. W. (1986). "Alternative Models for the Heterogeneity of Mortality Risks Among the Aged." *Journal of the American Statistical Association*, 81, 635–644.

Manton, K. G., & Stallard, E. (1979). "Maximum Likelihood Estimation of a Stochastic Compartment Model of Cancer Latency: Lung Cancer Mortality Among White Females in the U.S." *Computers and Biomedical Research*, 12, 313–325.

———. (1981). "Heterogeneity and Its Effect on Mortality Measurement." In *Methodologies for the Collection and Analysis of Mortality Data*, eds. J. Vallin & J. H. Pollard. Liege, Belgium: pp. 265–299.

Moore, D., & Haywood, M. (1988). "Death of a Salesman: Occupational Difference in Mortality Among a Cohort of Older Men." Battelle Memorial Institute, Columbus, Ohio (mimeograph).

Rogot, E. (1974). *Smoking and General Mortality Among U.S. Veterans 1954–1969*. Washington, DC: Dept. of Health, Education and Welfare. [DHEW Publication (NIH) 74-544].

Rogot, E., & Murray, J. (1980). "Smoking and Causes of Death Among U.S. Veterans: 16 Years of Observation." *Public Health Reports*, 95, 213–222.

Rosen, S., & Taubman, P. (1984). "Changes in the Impact of Education and Income on Mortality in the U.S." *Statistical Uses of Administrative Records: Recent Research and Present Prospects*. Washington, DC: U.S. Dept. of Health, Education and Welfare, pp. 549–554.

Sickles, R., & Taubman, P. (1986). "An Analysis of the Health and Retirement Status of the Elderly. *Econometrica*, 54, 1339–1356.

Taubman, P., & Rosen, S. (1982). "Healthiness, Education and Marital Status. In *Economic Aspects of Health*, ed. V. Fuchs. Chicago: University of Chicago Press, pp. 121–140.

U.S. Employment Service, Bureau of Employment Security. (1961). *Estimates of Worker Trait Requirements for 4,000 Jobs as Defined in the Directory of Occupational Titles*. Washington, DC: U.S. Government Printing Office.

Vaupel, J., Manton, K., & Stallard, E. (1979). "The Impact of Heterogeneity in Individual Frailty on the Dynamics of Mortality." *Demography*, 16, 439–454.

White, H. (1982). "Maximum Likelihood Estimation of Misspecified Models." *Econometrica*, 50, 1–26.

Wolfe, B. L., & Behrman, J. R. (1987). "Women's Schooling and Children's Health: Are the Effects Robust With Adult Sibling Control for the Women's Childhood Background?" *Journal of Health Economics*, 6, 239–254.

Yi, K. M., Honorer, B., & Walker, J. (1986). *CTM: A Program for the Estimation and Testing of Continuous Multi-Spell Models—Users Manual, Program Version 5.0*. Chicago: National Opinion Research Corporation.