DOCUMENT RESUME

ED 293 899 TM 011 603

AUTHOR	Willett, John B.; Singer, Judith D.
TITLE	Doing Data Analysis with Proportional Hazards Models: Model Building, Interpretation and Diagnosis.
PUB DATE	Apr 88
NOTE	58p.; Paper presented at the Annual Meeting of the American Educational Research Association (New Orleans, LA, April 5-9, 1988).
PUB TYPE	Reports - Research/Technical (143) -- Speeches/Conference Papers (150)
EDRS PRICE	MF01/PC03 Plus Postage.
DESCRIPTORS	*Computer Simulation; Computer Software; *Data Analysis; *Databases; Elementary Secondary Education; Multivariate Analysis; *Predictor Variables; *Teacher Persistence; Teaching (Occupation)
IDENTIFIERS	*Proportional Hazards Models; *Univariate Analysis

ABSTRACT

Guidelines for articulation of a framework for practical application of proportional-hazards models (PHMs) to professional survival analysis are provided. Focus is on data analysis fitting the PHMs with the semi-parametric methods of partial likelihood; this strategy is available in the BMDP2L and SAS PROC PHGLM computer programs. Areas in which advice and clarification are particularly apt include the definition of terms, the model and its assumptions, model-building, and interpretation and reporting of estimated model parameters. The presentation is based on a data-based example from a study of 10-year teacher survival patterns for a cohort of teachers who entered the profession in 1972 in Michigan. Survival analysis seeks to predict the duration during which a subject remains in a situation, in this case, the teaching profession. Discrete and continuous time analysis are examined; and the use of univariate, bivariate, and multivariate analysis are outlined. Choosing predictors and building a hierarchy of models, including interactions and predictors, and summarizing and comparing fitted models are discussed. Nine graphs and two tables are provided. (TJH)

ED293899

DOING DATA ANALYSIS WITH PROPORTIONAL HAZARDS MODELS: MODEL BUILDING, INTERPRETATION AND DIAGNOSIS

John B. Willett Ph.D.

Judith D. Singer Ph.D.

Harvard University Graduate School of Education

Appian Way, Cambridge, MA 02138

Presented at the

American Educational Research Association Conference

New Orleans, April 8, 1988.

AUTHORS' FOOTNOTE

John B. Willett and Judith D. Singer are Assistant Professors of Quantitative Methodology at Harvard University Graduate School of Education. The order of the authors has been determined by randomization. The authors would like to thank their colleague Richard J. Murnane for instigating the teacher career-path collaboration, for the the use of his data, and for the stimulating questions that he asked. Many thanks are also due to Jim Kemple for his tireless efforts at the computer, and to Frank Harrell for his help and assistance via BITNET.

DOING DATA ANALYSIS WITH PROPORTIONAL HAZARDS MODELS: MODEL BUILDING, INTERPRETATION AND DIAGNOSIS

Research questions about *duration*, and its relationship with other factors, are of fundamental interest in educational research. For example, the researcher may ask whether certain groups (such as women, minorities, math/science graduates) stay in teaching longer, or whether graduate students in particular areas of specialization typically spend longer completing their graduate degrees. Thus, research interest focuses not only upon the length of time (the *waiting-time*) to the event in question (e.g., resignation from teaching, graduation from a degree program), but also upon those characteristics of the individuals and of their treatment and environment that are associated with inter-individual variation in waiting-time.

Unfortunately, building models of waiting-time as a function of selected predictors is not completely straightforward. Generally the sampled individuals are not followed for their entire lifetimes. Rather, the investigator will gather the waiting-time data during some pre-specified and finite data-collection period. Although information may be available on all of the sampled individuals for the entire length of the data-collection period, the value of dependent variable (waiting-time) may still not be known for all of these individuals because the event of interest may not yet have occurred for all of them. Some teachers will not have left teaching by the end of the data collection period, some students will not yet have graduated. These individuals, for whom the event of interest has not yet occurred by the end of the data collection period, possess truncated or *right-censored* waiting-times (see Miller, 1981, pp. 3-9).

Analyzing the waiting-times of *only those individuals for whom the event of interest has occurred* will *necessarily* result in an underestimate of the true length of service or time-to-degree (Allison, 1982, 1984; Karrison, 1987). For instance, in a naive assessment of how long it takes to complete a doctorate, the median lifetime of that sample of graduate studuate who have completed the degree will seriously under-estimate the true median length of time to degree because the estimation has ignored those students who have not yet finished (see, for instance, the results cited in Abedi & Benkins, 1987). Obviously, the *very presence* of ungraduated students in the institution indicates that the *true* time to completion is *much longer* than the naive approach would suggest. The fact that some percentage of any entering cohort has not graduated by some specific occasion contains much information, particularly information about the probability that earning a doctorate takes longer than the time that has already elapsed.

Methodologists have responded to the biasing effect of right-censoring by creating new, and statistically sophisticated, analytic methods known as *survival analysis* (Gross & Clark, 1975; Kalbfleisch & Prentice, 1980; Lee, 1980; Miller, 1981) or *event-history analysis* (Allison, 1984; Tuma, 1982; Tuma & Hannan, 1978). Rather than modeling the waiting-times directly, these strategies model mathematical transformations of the waiting-time that remain meaningful in the face of right-censoring: the survival function and the hazard function. Within this large class of techniques, one popular semi-parametric strategy that examines hazard as a function of selected predictors is the fitting of *proportional-hazards models* by the method of *partial-likelihood* ("Cox regression" -- see Cox, 1972; Kalbfleisch & Prentice, 1980). Although the statistical papers describing this technique are highly technical and difficult to understand, the strategy has recently become accessible to empirical researchers by virtue of appearing as a procedure in several statistical packages (e.g. as BMDP2L, SAS PROC PHLGM).

Recently, we were approached by a colleague with a sizable empirical dataset and a research question that fell readily into the framework of proportional-hazards modeling. He needed advice about the practical application of the technique. During our collaboration it became obvious that, although the statistical theory underpinning survival analysis has been well-described in the methodological literature, there is very little good *practical* advice for the empirical researcher. There is no source that clearly articulates how good data-analysis is performed with these models. And yet, the articulation of sound strategies for thorough and high quality data analysis is crucially important if the findings of educational research are to support the burden of strong inference.

This paper was written in order to begin to articulate such guidelines. In our consulting, we have identified four areas in which some advice and clarification may be required:

- **Definition of terms**: Among a variety of conflicting definitions of the survival and hazard functions, what minimal set of definitions permits the empirical researcher to implement survival analysis effectively? What are the potential sources of confusion?

- **The model and its assumptions**: What is the nature of the statistical model that is being fit during proportional-hazards modeling? What key assumptions must be satisfied if the results of proportional-hazards modeling are to be believed? Is there a useful data-analytic analogy to provide a conceptual framework for understanding the function of the proportional-hazards model?

- **Model-building**: What data-analytic strategies are useful when building proportional-hazards models? Among a confusing variety of strategies, what methods are the most practical for verifying that the assumptions have been satisfied? How should the researcher choose the predictors to be included as predictors in the model? Are all predictors equally acceptable? Can interaction terms be included in the models?

- **Interpretation**: How can the estimated model parameters be most meaningfully interpreted? How should we report the results of the proportional-hazards modeling?

This paper establishes a framework for doing good data-analysis with proportional-hazards models. Our presentation makes use of a data-based example from an investigation of decade-long teacher survival patterns for a cohort of teachers who entered the profession in 1972 in Michigan (Murnane & Olsen, 1988; Murnane, Singer & Willett, 1988).

I. DEFINITION OF TERMS

Survival analysis was originally intended for the analysis of clinical lifetime data. Such research examines how long cancer patients survive after diagnosis or treatment, and the event that terminates the observed waiting-times is death. Thus, (unfortunately for other disciplines, where the event of interest is often the happier onset of employment or the occurrence of graduation), the language of survival analysis is couched in dark and forboding terms. In the social sciences, survival analysis helps the researcher investigate whether certain types of teacher stay in the profession longer than others. Do men remain longer than women? Foreign language teachers longer than science teachers? Do teachers who are paid more, remain longer?

Broadly conceived, then, survival analysis seeks to predict the waiting-times by variables that describe the treatment, background and environment of the sampled individuals. To address such issues the simple regression of the waiting-times themselves on predictors of interest would be inappropriate because right-censored teachers would have missing values for the outcome. Thus, the regression model would be fit in a subsample of short-lived teachers with unavoidable bias to the findings. To avoid this bias, the dependent variable must be re-conceptualized so that it incorporates information from both the censored *and* the uncensored cases simultaneously. For this reason, survival analysis uses as its dependent variable a mathematical function of the waiting-times that remains meaningful in the face of right-censoring: the *hazard function*,

Unfortunately, not only is this new "dependent variable" onoe or twice removed from the actual dependent variable of interest (the waiting-time) but there is also some confusion as to its definition and interpretation. The literature is littered with alternative definitions and explanations of the hazard function, some of which are incorrect (compare, for instance, Allison, 1984, p. 23; Anderson et al., 1980, p. 228; Gregson, 1983, p. 48; Kalbfleisch & Prentice, 1980, pp. 5-8; Miller, 1981, p. 2: Tuma, 1982). This confusion arises chiefly when the observed waiting-time has not been specified clearly as a *discrete* or a *continuous* variable. There is an inclination on the part of some authors to insist on the former while defining the hazard function, but assume the latter while performing the analyses. This leads to considerable confusion because hazard is defined differently in discrete and continuous time. In this section, we resolve these conflicts by specifically distinguishing the discrete and continuous waiting-time metrics and clearly defining the hazard function in both domains.

II.1 Hazard in discrete time

Rather than examine the waiting-times directly and be at the mercy of right-censoring, survival analysis focuses on the probability that a randomly-selected individual will "survive" beyond specific times. Then, although a given member of the population may ultimately be censored (i.e., may not ultimately "die" during observation), the fact that this individual is "alive" in some earlier period contributes information to the *probability of survival* beyond that earlier time.

The definition of a *survival-probability distribution function* is the first step in the ultimate definition of the *hazard function* that will serve as the foundation for the subsequent survival analyses. In discrete time, the value of the survival-probability distribution function (shortened to "survivor function" henceforth) at time t is the probability that a randomly-selected member of the population will "survive" beyond t:

$$S(t) = \text{Prob}\left[\text{survival beyond } t\right] \qquad [1]$$

(Kalbfleisch & Prentice, 1980, pp. 5-6 -- notice the conflict with Anderson et al., 1980, p.228 and Miller, 1981, p.2). In effect, the survivor function is nothing more than a "list" of probabilities -- one for each of the times of interest -- and therefore is best displayed graphically.

By definition, the first value of the survival-probability distribution function $S(0)$ is 1 because all the individuals are "alive" at the beginning of the period of study. As individuals "die" one by one, the function gradually "steps down" towards 0. However, it does not necessarily reach zero in the period under study because all of the members of the population may not have "died" by the end of the observation period -- these are the censored cases. Because the survivor function describes a sequence of probabilities (one for each discrete value of t), it can only take on values between 0 and 1. Because no more individuals can survive through the time period between t and t+1 than survived through the time period immediately preceding, it is a monotonically non-increasing, and usually decreasing, function of time. Because the waiting-times have been measured at discrete timepoints, S(t) is a step-function.

To have survived to at least t requires an individual to have also survived through all earlier time-periods. As a result, the value of the survival function at time t confounds cumulative information on survival for all of the preceding time-periods with specific information on survival in period between t and t+1. Because interest often centers upon the risk of "dying" in a particular time period (in a particular year, say) another function that is derived from the survivor function is usually examined in survival analyses: the *hazard function.*

With discrete waiting-time data, the value of the hazard function at time t is a probability. It is the *probability that a randomly-selected member of the population will "die" in the interval between t and t+1, given that s/he has survived until the beginning of that same interval:*

$$h(t) = \text{Prob}\left[\text{death between } t \text{ and } t{+}1 \mid \text{survival until } t\right] \qquad [2]$$

(Anderson et al., 1980, p. 228 -- note the misprint in Kalbfleisch & Prentice, 1980, p. 8, and the conflict with Miller, 1981, p.2). In other words, all of those members of the population who survive to time t enter the potential "risk set" for the period between t and t+1, and the hazard probability at time t is the proportion of this risk set who then die between t and t+1. Because h(t) represents the population probability of "dying" in the time-period between t and t+1 *conditional* on having survived to t, knowledge of the hazard function allows the investigator to determine whether there were certain time-periods in which the risk of "dying" was higher than usual, whether there are time periods that are intrinsically more "hazardous".

II.2 Hazard in continuous time

When survival time is measured continuously, the definition of the survival-probability distribution function in Equation [1] still holds and the obtained survivor function is a monotonically non-increasing and *continuous* function of time, rather than being a step-function. Therefore, S(t) is still interpretable as a sequence of probabilities.

Unfortunately, the definition of the hazard-function cannot be maintained in a transition from discrete to continuous time. When survival time is measured continuously, the definition of hazard must be modified. In continuous time, it makes little sense to talk of the probability of a death occurring at a given instant as this will automatically be zero, given the nature of continuous probability distributions. Moreover, it makes little sense to consider the probability of death in some arbitrary interval as this probability will differ depending upon interval length. Furthermore, the very act of choosing a specific interval length is essentially equivalent to arbitrarily discretizing the waiting-time measurements with consequent loss of information.

One solution to this problem is to compute the conditional probability of death within some arbitrary interval and, recognizing that this probability will increase with the length of the interval, divide the obtained probability by the length of the interval thus obtaining a sort of "conditional probability of death per unit time" or "rate of change of conditional probability". To ensure that the meaning of this probability rate is consistent from user to user, we must all agree on the length of the interval that will be used. Theoreticians, borrowing from differential calculus, have decided that interval should be *infinesimally small* -- that we should *take the limit of the probability rate as the length of the arbitrary interval disappears to zero.*

It is this mathematically-required modification that makes hazard a potentially confusing concept in continuous time. Consequently, in this paper, we will refer to continuous-time hazard as *hazard-rate* and the function will be given a new symbol, $\lambda(t)$, in order to distinguish it from the conditional probability in Equation [2]):

$$\lambda(t) = \underset{\Delta t \to 0}{\text{Limit}} \left[\frac{\text{Prob}\left[\text{death between t and t}+\Delta t \,\middle|\, \text{survival to t}\right]}{\Delta t} \right] \qquad [3]$$

(cf., Miller, 1981, pp. 2-3). Unfortunately, although this adjustment solves the mathematical problems inherent in the application of hazard to continuous time, it makes the function difficult to interpret because it can no longer be considered a probability.

Since the survivor function is relatively easy to interpret and maintains the same probabilistic interpretation in both discrete and continuous time, the job of the empirical researcher in interpreting hazard-rate is made easier by the demonstration of a simple relationship between the continuous-time hazard-rate and the survivor function. As Kalfleisch & Prentice (1980, p. 6) demonstrate, if the survivor function is transformed by taking the natural logarithm of S(t), then the value of the hazard-rate at time t is the (instantaneous) slope of the transformed curve at this time, multiplied by minus one. In other words, $\lambda(t)$ is the *negative slope of the logarithmically-transformed survivor function* and therefore, in a monotonically transformed world, the hazard-rate records fluctuations in the rate at which the survivor function decreases.

Although practical application of this new insight is not simple, the data analyst can learn to visualize S(t) when presented $\lambda(t)$, and vice versa: local maxima and minima in the hazard-rate function indicate more steep and less steep regions of the survivor function respectively, slope changes in the survivor function point out the peaks and valleys in the hazard-rate. Such comparisons emphasize that inspection of the hazard-rate function provides a very sensitive method -- a magnifying glass -- for the detection of changes in survival probability over time. This, of course, does not resolve the difficult problems involved in the reporting of the findings of a survival analysis to a relatively unsophisticated reader -- hazard-rate remains a difficult concept to interpret and discuss. Consequently, in Section IV, we suggest some relatively non-technical strategies that we have found useful when discussing our findings to others.

II. THE MODEL AND ITS ASSUMPTIONS

The generic label "survival analysis" refers, in general, to analytic strategies based upon the statistical modeling of the survivor and hazard-rate functions and subsumes a wide variety of different analytic methods, including:

- **Univariate descriptive strategies**: Product-limit and life-table estimation of the survivor and hazard-rate functions within a specific population. Eyeball inspection of estimated survivor and hazard-rate functions permits the investigator of teacher career paths to detect particular sections of the teaching career that are more "hazardous" than other periods (see, for instance, Singer & Willett, 1987).
- **Bivariate inferential strategies**: Non-parametric tests, including the Logrank and Wilcoxon tests, that permit the survivor and hazard-rate functions of two or more populations to be compared. These strategies can be used to answer questions such as: Do female teachers remain in teaching longer than their male counterparts? Do elementary-school teachers leave the profession more readily than other teachers?
- **Multivariate inferential strategies**: Statistical methods, that examine systematic differences in hazard-rate as a function of predictors selected to describe the individuals' treatment, background and environment. These strategies can be used to answer questions such as: What are the characteristics of those teachers with longer careers? Controlling for gender, do elementary-school teachers remain in teaching longer than secondary-school teachers? In the secondary school, controlling for age and gender, are teachers of physics and chemistry more inclined to leave teaching than other teachers?

The generic label of *life-testing* is often applied to the univariate descriptive and bivariate inferential strategies listed above (see Lee, 1980). A variety of life-testing methods are available in the major statistical packages as subprograms such as BMDP1L, SAS PROC LIFETEST, and SPSS^{X} SURVIVAL. In this paper, we focus principally on doing good data-analysis with one particularly robust and readily-available multivariate strategy that has seen wide application in clinical and sociological research, and that is becoming more familiar in educational research: the fitting of *proportional-hazards models by the semi-parametric methods of partial-likelihood* (Cox, 1972). This strategy is available in BMDP2L and SAS PROC PHGLM.

Proportional-hazards modeling bears a considerable resemblance to the more familiar multiple regression analysis. In both cases, an appropriate dependent variable is modeled as a function of user-selected predictors so that questions about the association between the dependent variable and the predictors can be answered. As with multiple regression analysis, proportional-hazards modeling permits more substantively and statistically powerful analyses to be performed than are available through the corresponding component bivariate analyses. However, just as preliminary univariate descriptive and bivariate correlational analyses can inform the building of multiple regression models, so can the univariate and bivariate methods of life-testing inform the multivariate modeling of hazard-rate. These latter methods will therefore become part of our exploratory data-analytic toolkit.

In this section of the paper, in the context of our data-based example, we define the mathematical form of the proportional-hazards model and discuss the key assumptions on which its analytic validity rests. We also provide a simple and more-familiar analogy that can help the empirical researcher build an intuitive feel for the way that the proportional-hazards model operates.

II.1 The proportional-hazards model

In our analyses of beginning teachers in Michigan through the 1970's, we are interested in the extent to which teachers with different characteristics have careers of different duration[1]. We focus particularly on two classes of predictor: (1) **Demographic characteristics** represented by *the age at which the teacher first entered teaching* (a continuous variable measured in years) and *the teacher's gender* (a dichotomous variable that takes on a value of 0 for males and a value of 1 for females). (2) **Background characteristics** represented by the *teacher's subject-matter specialty* (a dichotomous variable that takes on a value of 0 for english teachers and a value of 1 for elementary-school teachers).

If teaching career duration differs systematically by entryage, gender and subject-matter specialty, then the hazard-rate function describing career persistence in the population will depend not only upon time (the early years of teaching have been

---

[1] Of course, it is entirely possible that some teachers will teach for a while, leave and then return to the profession for a second or third "spell". In the analyses presented here, we are concerned only with the duration of the first spell of teaching. Murnane, Singer & Willett (1988) present a more complete analysis of the multispell data.

shown to be more "hazardous" -- see Mark & Anderson, 1985; Singer & Willett, 1987) but also upon entryage, gender, and subject-matter specialty. In order to represent such dependencies, Cox (1972) has introduced a statistical model that specifies the hazard-rate function in terms of:

- A vector of predictors $\mathbf{X}_p$ whose values differ among individuals and which describe selected characteristics of each individual's treatment, background and environment (the subscript referring to the p^{th} individual). In our analyses, $\mathbf{X}_p$ includes entryage A_p, gender G_p and subject-matter specialty S_p (and possibly interactions among them).

- A vector of regression parameters $\boldsymbol{\beta} = (\beta_a, \beta_g, \beta_s, \ldots)'$ that are common to all individuals in the population and that describe the direction and magnitude of the relationship between the predictors and the hazard-rate function. These estimation of these parameters is of principal interest in the analysis.

Using the Cox (1972) formulation, the following mathematical function is proposed as an appropriate model for representing hazard-rate as a function of the main effects of the three predictors:

$$\lambda_p(t) = \lambda_0(t)\, e^{\boldsymbol{\beta}'\mathbf{X}} = \lambda_0(t)\, e^{\left[\, \beta_a A_p + \beta_g G_p + \beta_s S_p \,\right]} \qquad [4]$$

where $\lambda_0(t)$ is an *unknown* function of time[2].

Providing that the predictors are time-invariant, the statistical model in Equation [4] separates out the universal time-dependence of the hazard-rate function from its dependence on the entryage, gender and subject-matter specialty of the members of the population. The overall hazard-rate for the p^{th} member of the population is a *product of two distinct parts*: a term that depends *solely upon time* and is the *same for all individuals*, and a term that is *unrelated to time* but depends upon a *linear combination of the values of the predictors for a specific individual.*

II.2 The assumptions of the proportional-hazards model

As a consequence of the specific mathematical form adopted for the statistical model in Equation [4], certain crucial assumptions must be met if the model is to be validly applied in practice: the "proportional-hazards assumption" and the "linear additivity assumption".

The proportional-hazards assumption

Notice that, in the model of Equation [4], $\lambda_0(t)$ does not

2 Both time-invariant and time-varying predictors can be included in the Cox formulation. Therefore, proportional-hazards modeling can easily incorporate teacher salary and teacher qualifications (both of which are intrinsically time-variable). However, for conceptual clarity, we restrict the current presentation to time-invariant predictors.

possess a subscript denoting the identity of any particular individual and therefore it is the same for everyone in the population. Furthermore, among those particular members of the population for whom all the predictors have the value zero (if any such exist), the hazard-rate function is *equal* to $\lambda_0(t)$. For all other individuals (for whom the predictors take on values different from zero), the exponentiated term in Equation [4] acts to shift the hazard-rate function multiplicatively away from $\lambda_0(t)$.

Conceptually then, $\lambda_0(t)$ is acting as a *baseline hazard-rate function*, all other individuals having hazard-rate functions that are *magnified* or *diminished* versions of the baseline. The process of magnification, or diminution, of an (unspecified) baseline function to generate all other required hazard-rate functions, implies that the hazard-rate functions of all individuals are *proportional*. This inbuilt proportionality must be checked during data-analysis if the analytic validity of the model is to be maintained. We will discuss strategies for checking the assumption, based on the methods of life-testing, in Section III.

The proportional-hazards constraint is both the curse and the reward of proportional-hazards modeling. It is a curse because it is an additional assumption that must be satisfied if the model is to be applied in practice. However, if the assumption is validly made, then it permits the statistical model in Equation [4] to be estimated in *two stages* by the method of *partial-likelihood* (Cox, 1972). First, by focusing only on the second (exponentiated) term in Equation [4], the parameter vector β can be estimated *without the necessity of specifying the functional form of the baseline* *hazard-rate*. Cox (1972, 1975) has indicated that obtained *partial-likelihood* estimates of β are endowed with large-sample distributional properties typical of their *maximum-likelihood* counterparts (see also Efron, 1977; Oakes, 1977). Second, subsequent to the estimation of β, estimates of the baseline hazard-rate function can be obtained (Breslow, 1974).

Thus, in Cox regression, the shape of the baseline hazard-rate function need never be specified "up-front" but can remain happily indeterminate until estimates are required. Therefore, unlike other types of survival analysis (which require the specification of a particular probability distribution for the waiting-times), Cox regression remains *partially-parametric* (Tuma, 1982). This has proven to be a considerable boon for empirical research in education, in which little is yet known about the nature of the underlying survival processes.

The linear additivity assumption

A useful and informative conceptual perspective on the operation of the proportional-hazards model can be obtained by taking natural logarithms throughout Equation [4]:

$$\log_e\left[\lambda_p(t)\right] = \log_e\left[\lambda_0(t)\right] + \left[\ \beta_a A_p + \beta_g G_p + \beta_s S_p\ \right] \qquad [5]$$

Notice that, in a logarithmically-transformed world, the action of the predictors is to shift the entire baseline hazard-rate function systematically and additively by an amount equal to the corresponding parameter β per unit increase in the predictor.

Thus, in the model, the influence of the predictors on the logarithm of the hazard-rate function is both *linear in the parameters* and *additive*.

This assumption has considerable implication for good data-analysis. First, the data-analyst must ensure that the linearity assumption has been satisfied. This presents no problem when the predictors are dichotomous, as are gender and subject-matter specialty in our analyses. However, in the case of a continuous predictor such as entryage, the data-analyst must check that differences in the log-hazard-rate due to differences in entryage are equivalent at all levels of entryage in order that the variable can be included as a continuous predictor in the model. Second, the data-analyst must check that the contribution of each of the predictors to the log-hazard-rate is additive. In practice, this involves specifying interactions among the predictors and testing that their inclusion has a non-zero effect on model fit. We will discuss both of these issues further, in Section III.

II.3 A useful analogy with ANCOVA

Conceptually, the transformed model in Equation [5] bears a remarkable resemblance to an analysis of covariance model. It is as though the influence of the predictors (the "treatment" variables) on the logarithm of the hazard-rate (the "dependent" variable) is being investigated, with time being treated as the "covariate". In this context, what is the proportional-hazards assumption in Cox regression corresponds to the assumption of homogeneity of within-group regression lines in the analysis of covariance. In our consultations with empirical researchers, we have found this mental image (of the log baseline hazard-rate shifting vertically up or down under the influence of the individual-specific predictors) to be a very useful heuristic[3].

---

3 The model in Equation [5] indicates that the hazard-rate function is linearly-related to the individual-specific predictors in a logarithmically-transformed world. As data-analysts we should not find this log-linear representation particularly unusual. The hazard-rate is strictly non-negative and, given a "dependent variable" with this property in any other data-analytic situation, we might extend its possible range to negative infinity prior to analysis by applying a logaraithmic transformation.

III. MODEL-BUILDING

In order to examine the relationship between the selected predictors (entryage, gender and subject-matter specialty), and the censored teaching career durations, a series of proportional-hazards models must be fitted and compared. And, for every model that is to be interpreted, the investigator must demonstrate that the underlying proportional-hazards and linear additivity assumptions are met and that the model does, in fact, fit. In this section, we demonstrate how preliminary survivor function estimation and life-table testing provide simple methods for: (1) checking the validity of the linear additivity and proportional-hazards assumptions, (b) eliminating less promising predictors from consideration in the model-building process, and (c) suggesting particular predictor combinations that might be included as interaction terms in the proportional-hazards models.

III.1 Checking the assumptions

In multiple regression analyses of uncensored data, the underlying assumptions are assessed easily by the careful examination of regression residuals. In the application of Cox regression to the analysis of censored waiting-time data, the situation is considerably more opaque. Not only are the underlying assumptions themselves more complex but there is also no non-systematic ("error") term in the proportional-hazards model to represent variation unexplained by the predictors and consequently there is no *direct* analogy for the regression residual[4]. In the literature, a large and confusing variety of formal and informal tests have been proposed for examining the validity of the underlying assumptions, many of them difficult to apply and hard to interpret (Allison, 1984; Harrell & Lee, 1986; Miller, 1981; Kalbfleisch & Prentice; 1981)[5].

Until new tests are evolved, or until the glitches are worked out of the old ones, the empirical researcher requires a few workable assumption-evaluation strategies -- particularly those based on plots -- that can readily be applied in practice. Such strategies must necessarily be able to be applied *before* the proportional-hazards models are fit. A natural approach is to use the univariate descriptive and bivariate inferential methods of

---

4 Miller (1981, pp. 168-172) and Kalbfleisch & Prentice (1980, pp. 96-98) discuss so-called generalized residuals that can be used to check the adequacy of the proportional-hazards model (see also, Cox & Snell, 1968). The conceptual and theoretical justification for generalized residuals is not easy to grasp and the residuals are difficult to obtain with commercial software packages.

5 SAS PROC PHGLM provides a simple test of the adequacy of the proportional-hazards assumption -- the "z:ph" test (see also Harrell & Lee, 1986). The provided z-statistic is intended to permit the testing of the null hypothesis that the proportional-hazards assumption is met, for each predictor separately. In our experiences, however, the z:ph test is so very highly sensitive to the presence of ties among the waiting-times -- the obtained z-statistic being inflated by factors of 10 to 100 as the proportion of ties increases -- that the test has almost no practical utility.

life-testing referenced earlier to estimate the survivor function within strata defined by the selected predictors and use these estimates as the basis for diagnostic testing.

Integrating and taking logarithms in Equation [4], provides the following reformulation of the proportional-hazards model:

$$\log_e(-\log_e(S_p(t))) = \log_e(-\log_e(S_0(t))) + \boldsymbol{\beta}'\mathbf{X} \quad . \qquad [6]$$

Thus, under the Cox regression model, the *log(-log) survivor function* is an additive linear function of the predictors. In other words, provided a dichotomous predictor such as gender satisfies the proportional-hazards assumptio plots of the log(-log) survivor function against time for men and women will have the same shape but be simply displaced vertically by a constant amount. Provided a continuous predictor such as entryage satisfies both the proportional-hazards and the linear-additivity assumptions, plots of the log(-log) survivor function against time will have the same shape but be displaced vertically by a constant amount for each unit increase in the predictor.

Because the survivor function is easily estimated (independent of proportional-hazards model-fitting) by the methods of life-testing and because inspection of the log(-log) survivor function readily reveals the failure of either the proportional-hazards or the linear additivity assumption, estimation of this function provides the most accessible diagnostic strategy that can be applied during the exploratory data-analytic phase of model building. In other words, in advance of model-fitting, the univariate and bivariate methods of life-testing (Lee, 1980) can be used to estimate survivor functions within each distinct level or stratum of each selected predictor so that the estimated log(-log) survival probabilities can be plotted against time and examined. Most commercial software packages (BMDP1L, SAS PROC LIFETEST, $SPSS_x$ SURVREG) provide life-table and product-limit (nonparametric maximum-likelihood) estimates of the survival-function so that the required log(-log) survival plots can easily be obtained.

Checking the proportional-hazards assumption

In practice, instead of plotting estimated values of log(-log(S(t))) against time, it is more useful to examine the *negative* log(-log) survivor plot (that is, a plot of $-\log(-\log(\hat{S}(t)))$ versus time). This latter plot is a monotonic transformation of the original estimated survivor function (i.e., both decrease steadily as time passes), whereas the former plot is a "flipped over" transformation of the original and therefore less intuitively appealing.

Thus, for our sample of beginning Michigan teachers, Figure [1] presents -log(-log) survivor plots estimated separately for men and women by the product-limit method (Kaplan & Meier, 1958). Figure [2] presents similar plots, by subject-area specialty[6]. Notice that, in both figures, the pairs of plots appear to have an approximately constant vertical separation over time indicating that the proportional-hazards assumption is met for both predictors.

---

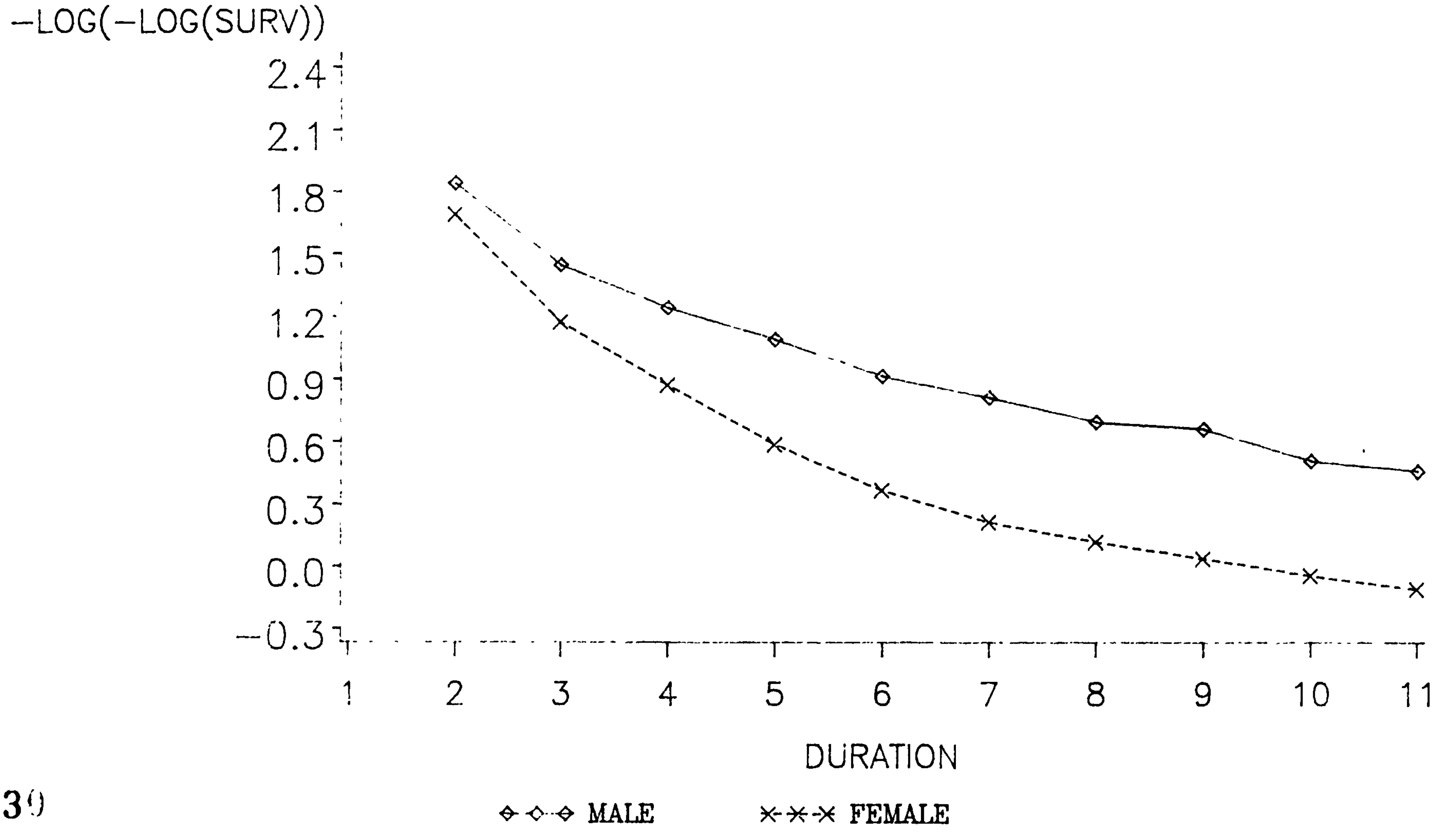
[6] All life-test estimation reported in this paper was performed with SAS (Version 5) PROC LIFETEST and all plots were created with SASGRAPH PROC GPLOT.

Insert Figures [1] and [2] here

When performing such comparisons purely by inspection, it is difficult to know how deviant the constant separation of the estimated -log(-log) survivor functions must be before the proportional-hazards assumption is compromised. The work of experts in the field (Kalbfleisch & Prentice, 1980, pp. 89-95, Figures 4.3 through 4.10) and our own practical experience suggests that the methodology is quite robust against violations of this assumption. Nevertheless, just as we would be very cautious in the interpretation of an analysis of covariance in which there existed a possibility that the within-group regression slopes were not homogeneous, so we should be similarly circumspect in our interpretations of Cox regression models fitted when the proportional-hazards assumption may not be met.

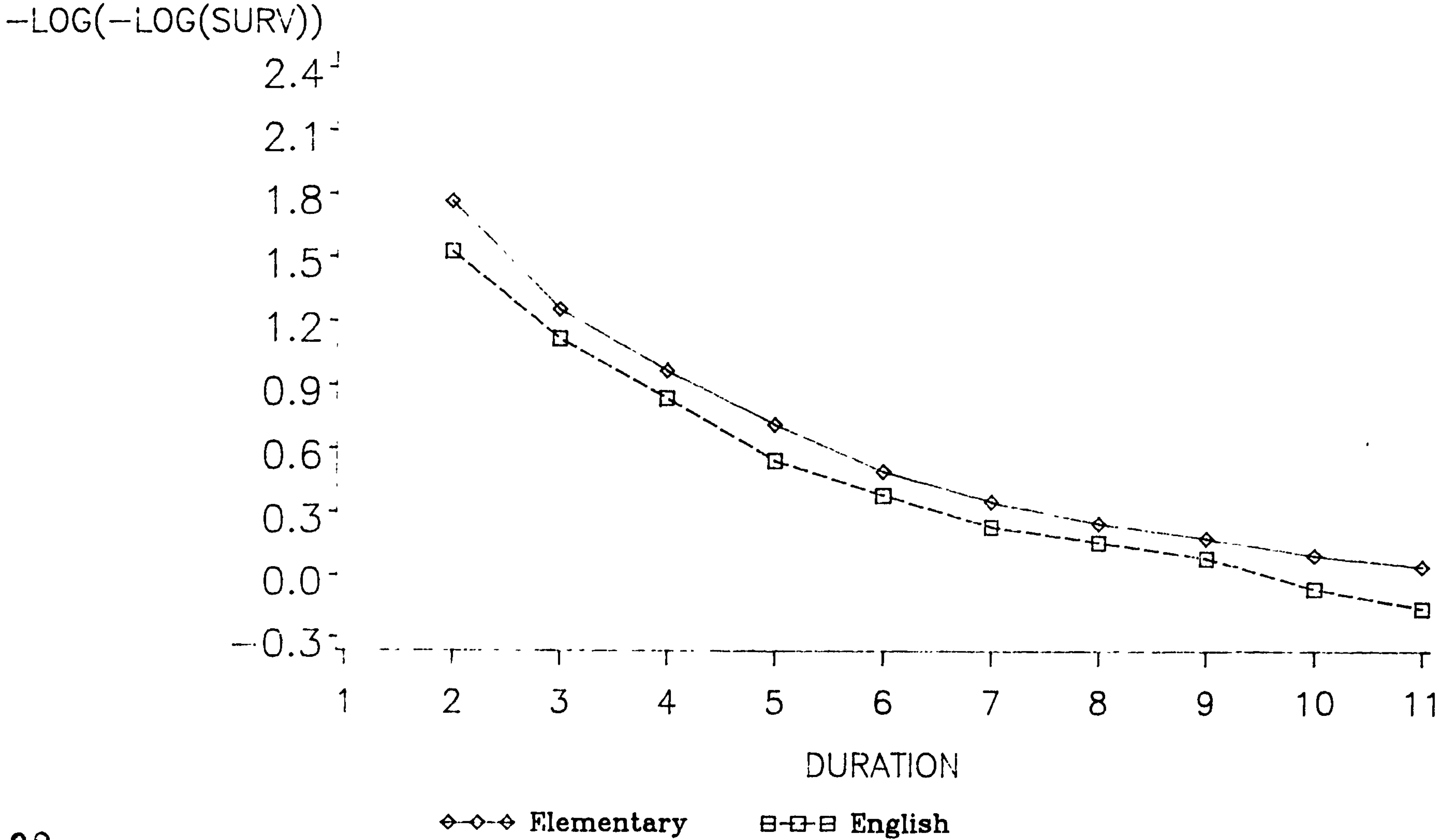
Unfortunately, the "intra-ocular" testing of the proportional-hazards assumption fails to take into account the standard errors associated with the survival probability estimates. When standard errors are large, -log(-log) survival functions that appear to violate the proportional-hazards assumption may in reality be indistuishable from one another and the question of their "constant separation over time", therefore, becomes mute. In practice, especially when relatively few subjects are involved in some particular level or stratum of the predictor being examined, adjacent estimated -log(-log) survivor-functions may overlap and

−LOG(−LOG(SURVIVAL)) BY DURATION
GENDER EFFECTS



T e are 2309 white teachers who entered in 1972
taught either elementry school or high school English

−LOG(−LOG(SURVIVAL)) BY DURATION
SUBJECT EFFECTS



…e are 2309 white teachers who entered in 1972
… taught either elementary school or high school English

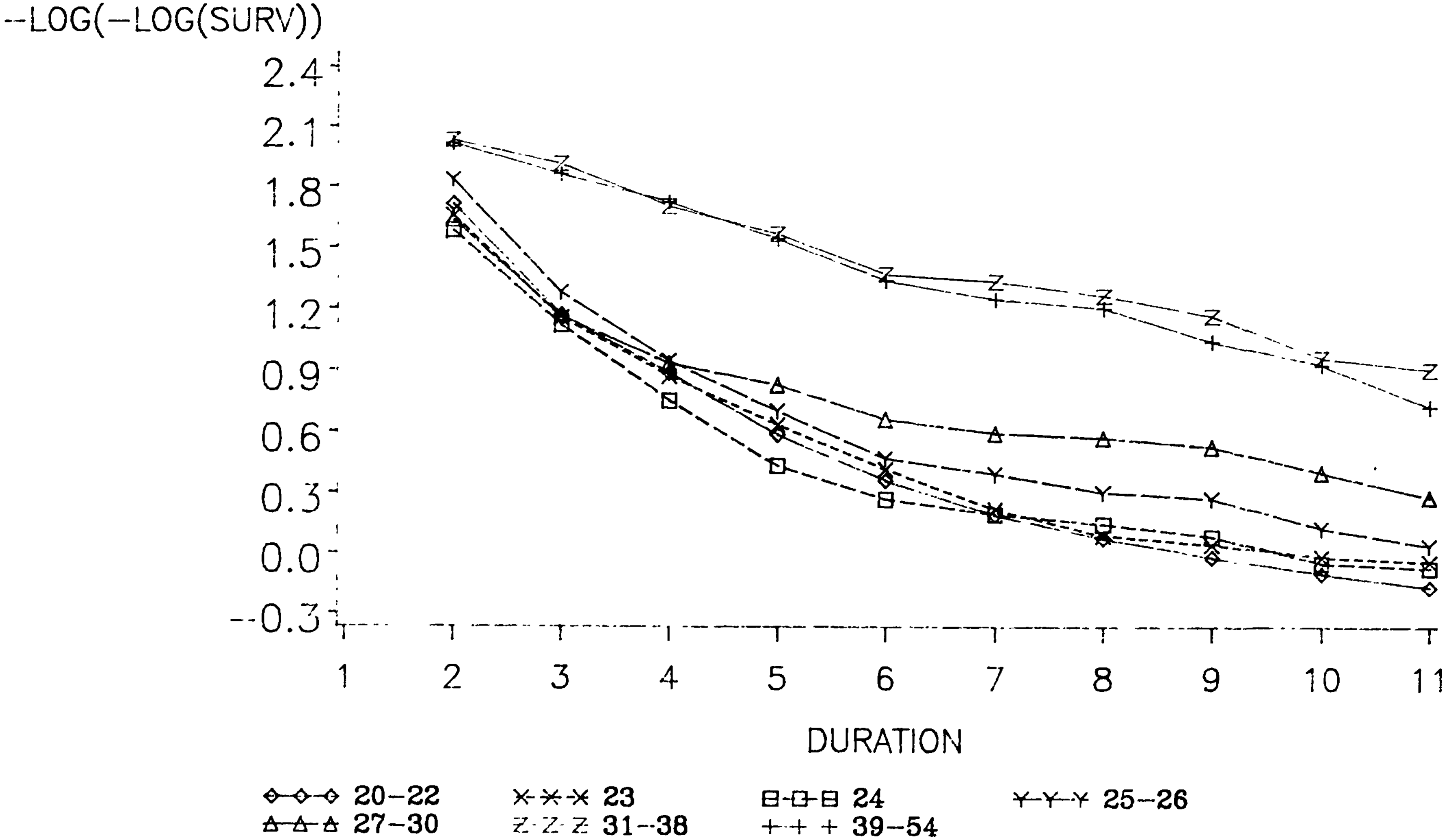
criss-cross quite frequently -- a condition that, if it represents the population situation accurately, of necessity violates the proportional-hazards assumption. However, by judicious application of the the Logrank or Wilcoxon tests of survivor function equivalence, the investigator may be unable to reject that the offending functions are, in fact, identical (in which case, the question of "constant separation over time" need not be addressed). Potentially, this provides the investigator with a diagnostic strategy of pairwise survival-function comparison that can be applied effectively in practice although, as with any other form of multiple comparisons, the family-wise α-level must be adjusted appropriately.

Checking the linear additivity assumption

In our analyses of the Michigan beginning teacher dataset, if entryage is to be included as a continous predictor in a Cox-regression model, then we must demonstrate that its action satisfies the assumption of linear additivity. In other words, we must check that equal increments of entryage are related to equal vertical shifts of the -log(-log) survival function. The simplest way to test such assumptions is to stratify according to the values of entryage and simultaneously display the estimated within-stratum -log(-log) survivor functions. Figure [3] presents such a plot.

Insert Figure [3] about here

−LOG(−LOG(SURVIVAL)) BY DURATION
ENTRYAGE EFFECTS



These are 2309 white teachers who entered in 1972
taught either elementary school or high school English

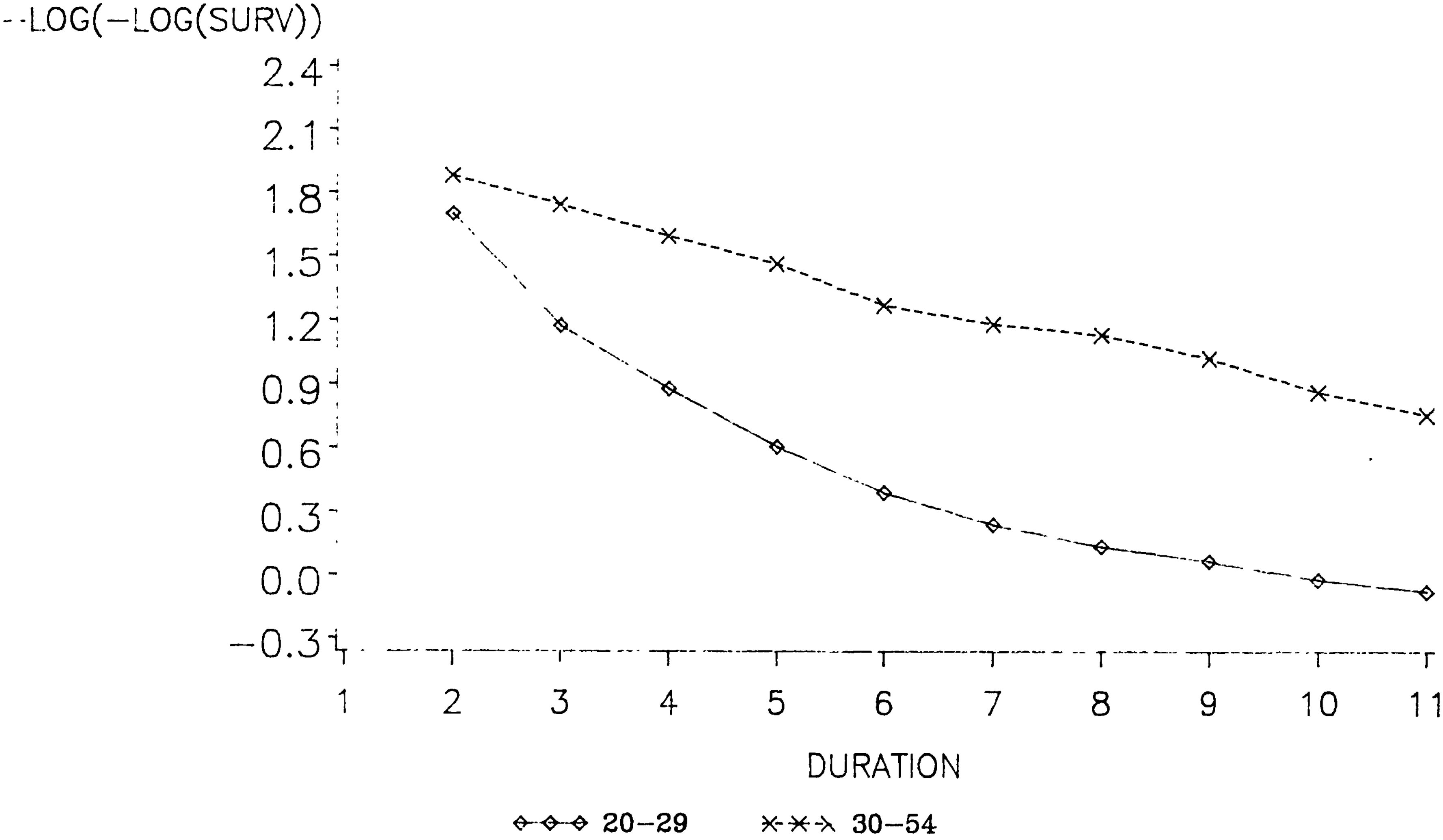
Inspection of Figure [3] not only indicates the failure of the linear additivity assumption for entryage, but also points up an interesting effect. Distinguishing only on entryage, there appear to be *two sorts of teachers younger* teachers (of entryage 30 and less), and *older* teachers (of entryage 31 and more). Because of this interesting natural dichotomization, in all subsequent analyses entryage has been recoded as a dummy variable with value 0 for younger teachers and value 1 for older teachers. Judgement of the adequacy of the proportional-hazards assumption for groups distinguished by the newly-dichotomized entryage is informed by the estimated -log(-log) survival functions in Figure [4].

Insert Figure [4] about here

III.2 Choosing predictors and building a hierarchy of models

Corresponding to each of Figures [1], [2] and [4], there are tests of survival-function equivalence such as the nonparametric Wilcoxon and Logrank tests (see Lee, 1980, Ch. 5). The significance levels of these tests are usually in good agreement, although the Wilcoxon statistic does place more weight on early differences between the survival functions whereas the Logrank statistic attaches equal importance to all differences regardless of location along the time axis. However, because there is a particularly "close relationship between the logrank [test] and the

−LOG(−LOG(SURVIVAL)) BY DURATION
ENTRYAGE EFFECTS



These are 2309 white teachers who entered in 1972
taught either elementary school or high school English

proportional-hazards model" (Kalbfleisch & Prentice, 1980, p.215), we favor this latter test as an exploratory tool.

If the investigator carries out the Logrank test when plotting the -log(-log) survival functions above, then the test statistics can be used to inform the model-building process. In particular, survival functions that can be inferred to differ in the population on the basis of the Logrank test are likely to be widely separated as estimated -log(-log) survival functions. Consequently, they are also likely to have been generated by grouping on a predictor that has substantial influence in the Cox regression model. For this reason, just as simple bivariate correlational analyses can inform the building of multiple regression models, so can systematic life-testing be one foundation for the construction of proportional-hazards models.

In the Michigan data, although the vertical separation of the subject-area-specific -log(-log) survivor functions in Figure [2] is much smaller than the corresponding vertical separations in Figures [1] and [4], the null hypotheses of survival function equivalence can be rejected for all three predictors: entryage (logrank chi-square statistic=67.9, df=1, p<.000), gender (logrank chi-square statistic=59.8, df=1, p<.000), and subject-area specialty (logrank chi-square statistic=6.8, df=1, p<.009). Thus, all three variables are likely to be decent predictors in subsequent proportional-hazards models, with the possibility that subject-area specialty will have the weakest effect.

Table [1] presents the results of fitting a variety of proportional-hazards models by the method of partial likelihood[7]. For the moment let us only consider Models #1 through #3, each of these models contains a single predictor. Notice that the fitted relationships are as we might have expected on the basis of the survival-function comparison above. By inspection of both the parameter estimates (in relation to their standard errors) and the chi-square improvement-in-fit statistics, we see that entryage and gender both demonstrate a strong non-zero relationship with hazard-rate ($p<.001$) and, while the influence of subject-area specialty is much weaker, it remains significantly different from zero ($p<.05$).

Insert Table [1] about here

III.3 Including interactions as predictors in the model

As in multiple regression analysis, the influence of the predictors on the dependent variable may not be solely additive. The predictors may *interact* in their prediction of the hazard-rate. For this reason, products of the predictors may be legitimately included as interaction terms in the model. And, just as exploratory -log(-log) survivor plots can inform the selection of predictors for inclusion as main effects in proportional-hazards

[7] All of the Cox regression analyses reported in this paper were carried out by SAS (Version 5) PROC PHGLM.

Table [1]. Investigating career duration of teachers entering the profession in Michigan in 1972: estimated proportional-hazards model parameters, standard errors and improvements-to-fit, models #1 through #6.

Model ID#	Parameter Estimate for each covariate				Model Chisquare Statistic† (df)
	A (se)	G (se)	A*G (se)	S (se)	
[1]	-.794*** (.105)				72.48*** (1)
[2]		.520*** (.072)			58.36*** (1)
[3]				-.164* (.067)	5.86* (1)
[4]	-.854*** (.105)	.570*** (.072)			143.08*** (2)
[5]	- 023 (.246)	.624*** (.075)	-.955*** (.271)		153.41*** (3)
[6]	-.013 (.246)	.638*** (.075)	-.957*** (.271)	-.185*** (.069)	160.79*** (4)

*p<.05 **p<.01 ***p<.001

†The model chi-square statistic is suitable for testing H_0: there is no improvement in fit on the simultaneous addition of all the predictors in the model to a model that contains no predictors. It is obtained by subtracting approximate chi-square (-2 log likelihood) goodness-of-fit statistics for the fitted model and for a model containing no predictors.

models (Figures [1], [2] and [4]), so can similar plots inform the identification of potential interactions among the predictors.

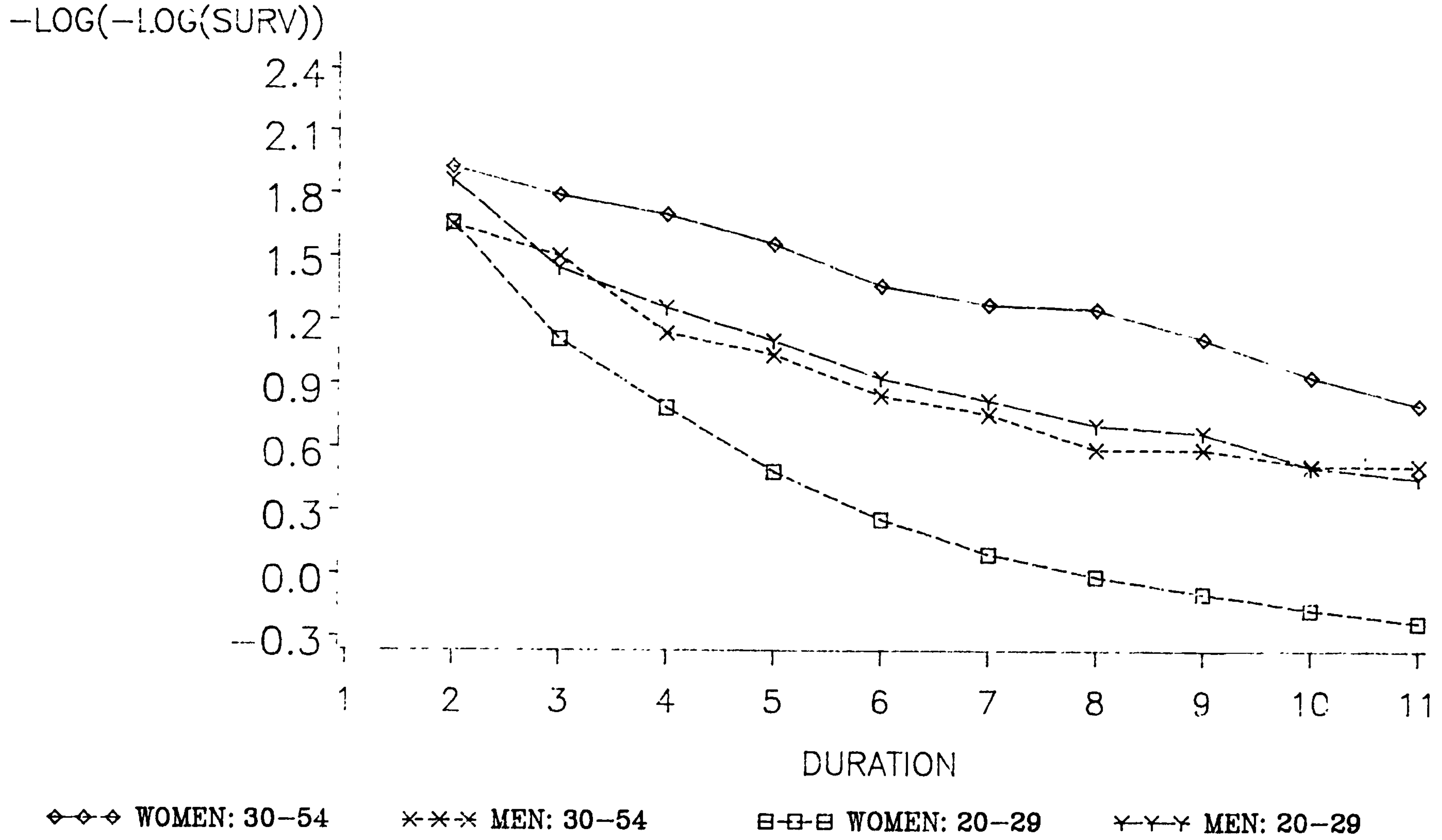
For our sample of beginning Michigan teachers, Figure [5] presents -log(-log) survivor functions estimated separately for young men and women, and for older men and women. Notice that, ignoring for the moment the possibility that the proportional-hazards assumption may not hold, the four plots offer considerable evidence of an interaction between entryage and gender. In particular, whereas the estimated -log(-log) survival plots for young men and older men coincide, the equivalent plots for women are widely separated and fall on opposite sides of the plots for men. This impressive gender variation in separation between entryage survival functions suggests that the effect of entryage on teacher career duration is very important for women but not at all important for men. It suggests that *entryage and gender interact very strongly in the prediction of the hazard-rate function*, and that a product term that captures this synergy must be included in any proportional-hazards model which includes entryage and gender as predictors[8].

Insert Figure [5] about here

---

8 Murnane, Singer and Willett (1988), in a more extensive analysis of these same questions, conjecture that there are forces (such as the decision to embark upon full-time child-rearing) which draw young women out of the profession early causing their survival functions to fall below those of the other groups and making the profession appear much more hazardous for them.

−LOG(−LOG(SURVIVAL)) BY DURAT1
AGE AND GENDER EFFECTS



These are 2309 white teachers who entered in 1972
taught either elementary school or high school English

In Figure [5], because the estimated -log(-log) survivor functions for the four entryage/gender subgroups do not maintain constant vertical separation over time, there is also evidence that the proportional-hazards assumption may no longer be met. This potential failure of a crucial assumption serves to emphasize that there are *two types of interaction* that are of interest in proportional-hazards modeling: (a) *predictor-predictor interactions*, and (b) *predictor-time interactions*. The former interactions can be represented by products of the appropriate variables and are legitimate candidates for inclusion as predictors in any proportional-hazards model. The latter interactions undermine the very fabric of the analysis by challenging the proportional-hazards assumption itself and must be demonstrated to be zero.

The earlier analogy with ANCOVA provides insight into, and a method of evaluating, the potential failure of the proportional-hazards assumption. The predictor-predictor interactions noted above correspond, in analysis of covariance, to perfectly acceptable interactions among the ANCOVA treatment variables whereas the predictor-time interactions correspond to interactions between the treatment variables and the covariate in analysis of covariance -- interactions that would violate the assumption of homogeneity of within-group regression slopes. Just as failure of the homogeneity of within-group regression slopes can be evaluated by testing the addition to the ANCOVA model of interactions between the group dummies and the continuous covariate, so can the addition to the Cox model of interactions between the predictors and *time* inform the evaluation of the proportional-hazards assumption.

The modeling of predictor-time interactions requires the inclusion of time-varying predictors in the proportional-hazards model (Allison, 1984; Kalbfleisch & Prentice, 1980). While this is not difficult, it demands computer software that is not widely available and is a topic that we have chosen not to discuss in this paper. Typically, an investigator lacking such software and faced with the dilemma of Figure [5] would either develop proportional-hazard models for subject-area specialty within each of the four entryage/gender subpopulations separately, or would decide that the violation was not serious and could safely be ignored. Based on exemplary plots presented in the literature (Kalbfleisch & Prentice, 1980, pp. 91-95), we have decided to adopt the latter approach.

Model #4 of Table [1] presents parameter estimates and improvement-in-fit statistics for a proportional-hazards model that includes only the main effects of entryage and gender. Model #5 in the same table includes the interaction between entryage and gender, in addition to their main effects. Comparison of the two models by subtraction of the model chi-square statistics and the corresponding degrees of freedom indicates that the inclusion of the interaction term leads to a considerable improvement in fit (change in model chi-square statistic = 10.33, change in df=1, $p<.005$). Thus, the non-additive influence of entryage and gender suggested by Figure [5] is strongly confirmed[9].

---

[9] Notice that, compared to earlier models, the magnitude of the estimated age effect in Model #5 is very small. This is not unexpected because, when the interaction of age and gender is

included in the model, the main effect of age indicates the differences in survival between younger and older men and, as we have seen by our earlier inspection of Figure [5], this difference is close to zero.

IV. THE ISSUE OF INTERPRETATION

Having carried out a lengthy and detailed exploratory analysis to support the building of a hierarchy of proportional-hazards models, the investigator ultimately faces the problem of interpreting the findings. This can be an onerous burden because the obscure nature of the "dependent variable" (the hazard-rate function) and the logarithmic construction of proportional-hazards models themselves makes the interpretation somewhat opaque.

Furthermore, in applied research, the substantive questions being answered often have large-scale policy implications and must be reported to methodologically-naive clients. In our case, a large section of the audience for our findings consists of school principals, school superintendents and state and federal administrators. This group typically displays minimal statistical training, if at all. Thus, careful consideration needs to be given to the format of the research-reporting and largely non-technical approaches based on intuitively-oriented plots and summary statistics must be applied. In this section, we give examples of displays and summary statistics that can be used to enhance the substantive interpretation of the findings of a survival analyis.

IV.1 Summarizing and comparing the fitted models

Table [1] represents what we consider to be the minimally acceptable format for reporting the fits of the proportional-hazards modeling process. The table presents the parameter estimates, their standard errors and the model improvement-to-fit statistics.

In addition, the table systematically presents results for a *hierarchy of fitted models*. To structure this hierarchy, as Mosteller & Tukey (1977) recommend in the case of multiple regression analysis, we have established an order of priority among the predictors on the basis of our research question. Other than baseline analyses to determine the "zero-order" influence of the three predictors individually, our principal intention is to control for the effects of the demographic variables (entryage and gender) and then examine the influence of the background characteristics (subject-area specialty) on career duration. Therefore, rather than reporting a single "best" model, we have reported several fitted models in order to: (a) demonstrate the priority among the various classes of predictor, (b) illustrate that we have done a credible job of considering alternative explanations of the phenomenon under study, and (c) provide the reader with a basis for comparison when interpreting the later, more complicated models.

Thus. Models #1 through #3 display the zero-order influence of each predictor separately on career duration. Models #4 and #5 summarize the joint influence of the demographic characteristics on hazard-rate and their comparison allows the importance of the entryage/gender interaction to be evaluated (see Section III.3). Models #5 and #6 permit the influence of subject-area specialty to be examined, having controlled for demographic variation. In this latter comparison, by inspecting both the change in the model chi-square statistic and the parameter estimate in relation to its standard error, we note that subject-area specialty appears to play a role in the prediction of teaching career duration even when the influence of entryage and gender have been controlled.

IV.2 Interpreting the fitted models

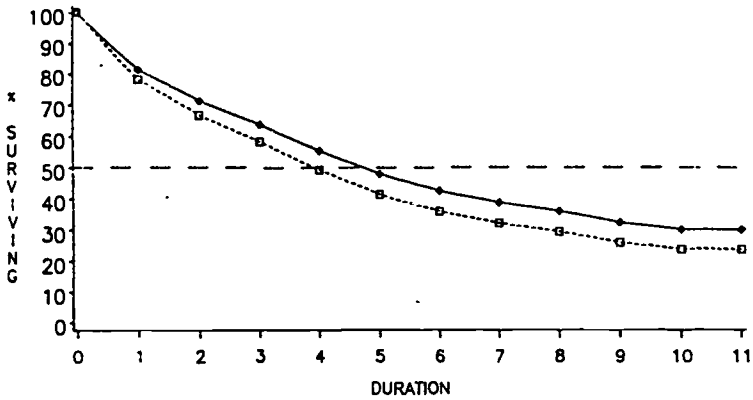
Algebraically, the magnitude and direction of the effects of the predictors on the hazard-rate function are not difficult to interpret, particularly if the logarithmic reformulation of the proportional-hazards model in Equation [5] is utilized. In this re-expression, predictors can be interpreted in terms of their additive influence on the logarithm of the hazard-rate.

Thus, as the value of predictors with *positive* "slope" parameters increases, the log-hazard-rate also increases and there is a corresponding decline in the probability of survival. For example, in the case of the zero-order influence of gender (Model #2), the parameter estimate (.520) indicates that, as gender changes from G=0 (men) to G=1 (women) the log-hazard-rate is shifted "vertically upwards" by .520. Thus, at all times, the fitted hazard-rate for women will be "higher" (i.e., more positive) than the fitted hazard-rate for men. Consequently, we can regard the teaching career as generally "more hazardous" for women; they are less likely to survive, more likely to leave. On the other hand, the negative parameter estimate for subject-area specialty in Model #6 indicates that, after controlling for the demographic variables, the teaching profession is slightly less hazardous for elementary school teachers (S=1) than for english teachers (S=0) and therefore the former will remain longer than the latter.

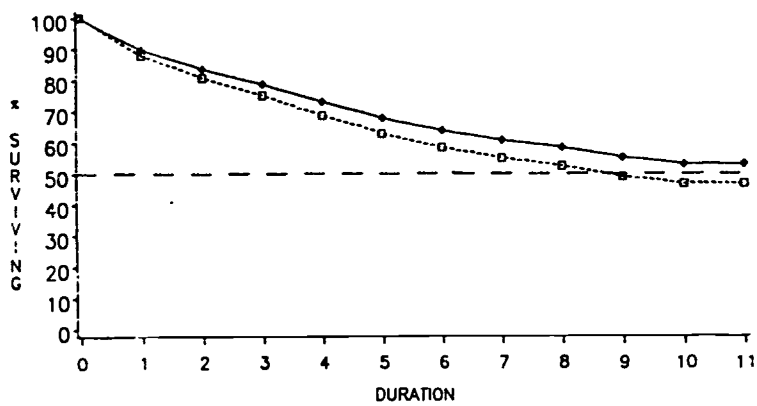
Even though such interpretations are not particularly difficult to make, they are not readily accessible to the naive reader. To make our research findings more palatable, we have found *fitted survival functions* to be of the greatest assistance as this avoids discussion of the hazard-rate entirely. By systematically varying the values of the predictors, it is a relatively simple matter to generate and plot fitted survival functions for hypothetical subgroups in the population. Thus, Figure [6] presents a series of fitted survival plots in which the career persistence of the four demographic groups is documented, by subject-area specialty. We have found such plots to be acceptable to most non-technical audiences, as it seems intuitively obvious to the methodologically-unsophisticated that the size of the entering cohort will decline over time and that there will be differences by entryage, gender and subject-area specialty.

Insert Figure [6] here

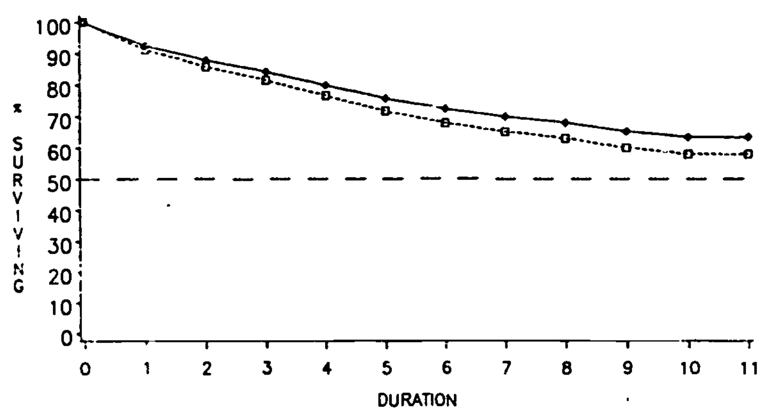
Notice, in Figure [6], that a horizontal dashed line has been drawn across the plots where the fitted survival probability is



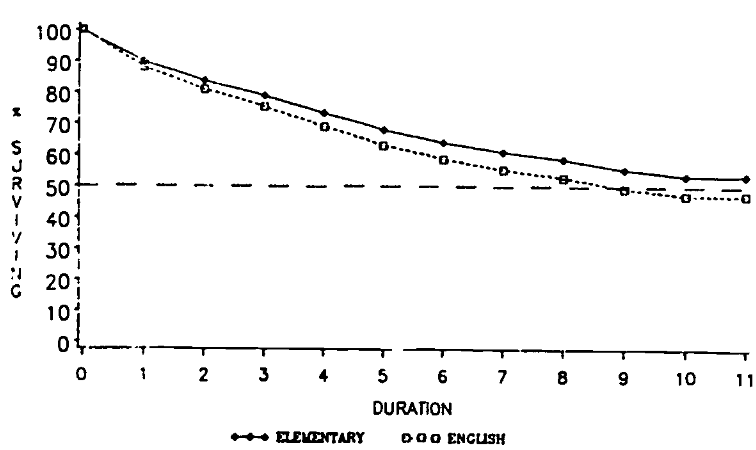
Predicted Survival Functions
Subject Effects for Young Women



Predicted Survival Functions
Subject Effects for Young Men



Predicted Survival Functions
Subject Effects for Older Women



Predicted Survival Functions
Subject Effects for Older Men

equal to one half. By reading off the time on the abscissa corresponding to this probability, a useful summary of the fitted survival distribution is estimated -- the *predicted median lifetime* -- the length of time taken by 50% of the particular subgroup to "die". Thus, on the basis of the fitted survival plots in Figure [6], the predicted median lifetimes of each of the distinct groups in our analyses are presented in Table [2]. This latter table is relatively simple for the non-technical to understand and can form the basis of an interesting report or oral presentation.

Insert Table [2] here

Table [2]. Investigating career duratic . of teachers entering the profession in Michigan in 1972: estimated median lifetimes for elementary-school and english teachers based on Model #6 in Table [1], by entryage and gender.

Subject-area Specialty	Entryage<=30 (A=0)		Entryage>30 (A=1)	
	Men (G=0)	Women (G=1)	Men (G=0)	Women (G=1)
English (S=0)	8.8	3.9	8.8	>11.0
Elementary (S=1)	>11.0	4.7	>11.0	>11.0

REFERENCES

Abedi, J., & Benkin, E. The effects of students' academic, financial and demographic variabl.s on time to the doctorate. Research in Higher Education, 1987, 27, 3-14.

Allison, P.D. Discrete-time methods for the analysis of event histories. In S. Leinhardt (Ed.), Sociological Methodology 1982. San Francisco, CA: Jossey-Bass, 1982.

Allison, P.D. Event History Analysis: Regression for Longitudinal Event Data. Sage University Paper series on Quantitative Applications in the Social Sciences, #46. Beverly Hills, CA: Sage Publications, 1984.

Anderson, S., Auquier, A., Hauck, W.W., Oakes, D., Vandaele, W., & Weisberg, H.I. Statistical Methods for Comparative Studies. New York, NY: John Wiley & Sons, 1980.

Breslow, N. Covariance analysis of censored survival data. Biometrics, 1974, 30, 89-99.

Cox, D.R. Regression models and life tables. Journal of the Royal Statistical Society Series B, 1972, 34, 187-202.

Cox, D.R. Partial likelihoods. Biometrika, 1975, 62, 599-607.

Cox, D.R., & Snell, E.J. A general definition of residuals. Journal of the Royal Statistical Society, Series B, 1968, 30, 248-275.

Efron, B. The efficiency of Cox's likelihood function for censored data. Journal of the American Statistical Association, 1977, 72, 557-565.

Gregson, R.A.M. Time Series in Psychology. Hillsdale, NJ: Lawrence Erlbaum Associates, 1983.

Gross, A.J., & Clark, V.A. Survival Distributions: Reliability Applications in the Biomedical Sciences. New York, NY: John Wiley & Sons, 1975.

Harrell, F., & Lee, K.L. Verifying assumptions of the Cox proportional-hazards model. Proceedings of the Eleventh Annual SAS User's Group International Conference. Cary, NC: SAS Institute Inc., 1986.

Kalbfleisch, J.D., & Prentice, R.L. The Statistical Analysis of Failure Time Data. New York, NY: John Wiley & Sons, 1980.

Kaplan, E.L., & Meier, P. Nonparametric estimation from incomplete observations. Journal of the American Statistical Association, 1958, 53, 457-481.

Karrison, T. Restricted mean life with adjustment for covariates. Journal of the American Statistical Association, 1987, 82, 1169-1176.

Lee, E.T. Statistical Methods for Survival Data Analysis. Belmont, CA: Lifetime Learning Publications, 1980.

Mark, J.H. and Anderson, B.D. Teacher survival rates in St. Louis 1969-1982. American Educational Research Journal, 1985, 22, 413-421.

Miller, R.G. Jr. Survival Analysis. New York, NY: John Wiley & Sons, 1981.

Mosteller, F., & Tukey, J.W. Data Analysis and Regression. Reading, MA: Addison-Wesley Publishing Company, 1977.

Murnane, R.J., & Olsen, How long teachers stay in teaching depends on salaries and opportuniti . Paper presented at the Econometrics Society Meeting, Chicago, December, 1987.

Murnane, R.J., Singer, J.D., & Willett, J.B. Public school teachers' career paths: implications for teacher supply. Submitted to Educational Researcher, March, 1988.

Oakes, D. The asymptotic information in censored survival data. Biometrika, 1977, 64, 441-448.

Singer, J.D., & Willett, J.B. Uncovering involuntary layoffs in teacher survival data: The year of leaving dangerously. Paper presented at the annual conference of the American Educational Research Association, Washington, April, 1987. Submitted to Economics of Education Review, February, 1988.

Tuma, N.B. Nonparametric and partially parametric approaches to event history analysis. In S. Leinhardt (Ed.), Sociological Methodology 1982. San Francisco, CA: Jossey-Bass Publishing Company, 1982.

Tuma, N.B., & Hannan, M.T. Approaches to the censoring problem in analysis of event histories. In K.F. Schuessler (Ed.), Sociological Methodology 1979. San Francisco, CA: Jossey-Bass Publishing Company, 1978.

FIGURE CAPTIONS

Figure [1]: Testing the proportional-hazards assumption for gender. Product-limit (nonparametric maximum likelihood) estimates of -log(-log) survivor probability plotted against time, for men and women.

Figure [2]: Testing the proportional-hazards assumption for subject-area specialty. Product-limit (nonparametric maximum likelihood) estimates of -log(-log) survivor probability plotted against time, for elementary-school and english teachers.

Figure [3]: Testing the linear additivity assumption for entryage. Product-limit (nonparametric maximum likelihood) estimates of -log(-log) survivor probability plotted against time, for a variety of entering age cohorts.

Figure [4]: Testing the proportional-hazards assumption for entryage. Product-limit (nonparametric maximum likelihood) estimates of -log(-log) survivor probability plotted against time, for teachers for whom entryage<=30 and entryage>30.

Figure [5]: Detecting the interaction of entryage and gender. Product-limit (nonparametric maximum likelihood) estimates of -log(-log) survivor probability plotted against time for men and women, by dichotomized entryage.

Figure [6]: Fitted survival-functions. Survivor functions estimated from Model #6 in Table [1] for each of the entryage/gender subgroups, by subject-area specialty.