

A Test for Discriminating Between Additive and Multiplicative Relative Risks in Survival Analysis

By ELJA ARJAS†

University of Oulu, Finland

and DAVID VENZON

Fred Hutchinson Cancer Research Center, USA

[Received August 1986. Revised August 1987]

SUMMARY

The proportional hazards model of Cox (1972) specified a relative risk of the form $\exp(\beta'z)$ for a covariate vector z . We propose a test to detect true relative risks that are a less convex function of $\beta'z$ than the exponential, e.g. a linear function. The technique exploits the expected overestimation of conditional probabilities of failure at the ends of the range of values of $\beta'z$. Simulations indicate that the test is slightly conservative and illustrate how power is lost with certain covariate distributions and censoring patterns.

Keywords: Additive models; Cox regression; Goodness of fit; Multiplicative models; Relative risk

1. Introduction

As the proportional hazards regression model proposed by Cox (1972) has become widely used, generalisations of its assumptions are being more often considered. The hazard function for an individual with a p -vector of covariates z , originally

$$\lambda(t; z) = \lambda_0(t) \exp(\beta'_0 z), \quad (1)$$

for an unspecified baseline hazard $\lambda_0(t)$ and a p -vector of 'true' model parameters β_0 , can be modified in several ways. We will here retain the dependence on the covariates through the 'linear predictor' $\eta = \beta'z$, permitting estimation of β_0 by methods appropriate for generalised linear models (Nelder and Wedderburn, 1972). We wish, however, to allow for the possibility that the hazard depends on a different function of η . That is, assume

$$\lambda(t; z) = \lambda_0(t)r(\beta'_0 z) \quad (2)$$

for some twice-differentiable function $r(\eta)$ that is positive and increasing in η for β in some neighbourhood of β_0 and for all observed z . The construction of the partial likelihood (Cox, 1975) proceeds with this change in the same way as before. Prentice and Self (1983) established conditions for the consistency and asymptotic normality

† *Address for correspondence:* Department of Applied Mathematics and Statistics, University of Oulu, SF-90570 Oulu 57, Finland.

of $\hat{\beta}$, the maximum partial likelihood estimator of β_0 , in the general case of broadly defined, time-dependent covariate processes, with any such relative risk function r correctly specified.

A natural alternative form for r has the relative risk linear, rather than log-linear, in the covariates, i.e. $r(\eta) = 1 + \eta$ for $\eta > -1$. This specifies an additive, rather than multiplicative, model of interaction between two or more factors in epidemiological studies, for example (Rothman *et al.*, 1980).

One method for comparing the fit of these two models for a given set of data is to embed the linear and exponential functions in a parametric family. For example, relative risk functions that are mixtures of the two (Thomas, 1981) and Box-Cox transformations that include them both (Guerrero and Johnson, 1982; Breslow and Storer, 1985) have been proposed. In this way, likelihood ratio and score statistics can be computed for testing the specification of the relative risk function within the family. This is the approach that also leads to 'goodness of link' tests for generalised linear models (Pregibon, 1980).

Some difficulties in estimation and inference arise in non-exponential models, however. In a simulation study of the linear relative risk function, for example, Prentice and Mason (1986) found up to 21% of the replicates not convergent by a standard algorithm. The poor performance in small samples of methods based on asymptotic properties for hypothesis testing and confidence interval estimation with this model have also been addressed (Storer *et al.*, 1983; Moolgavkar and Venzon, 1987). In parametric families of relative risk functions, the interpretation of the parameter estimates that are found can be an additional problem (Lustbader *et al.*, 1984; Breslow, 1986).

When there is no independent information to recommend one function $r(\cdot)$ over another, provisional use of the exponential avoids the above difficulties. A second method for finding a suitable relative risk function, therefore, is to fit the Cox model (1) and then apply general tests of agreement of the data with the model's assumptions. These tests involve no extra parameters to specify families of alternative relative risk functions. For example, Kay (1977) describes some graphical techniques, and Schoenfeld (1980) derives a class of asymptotically χ^2 -distributed statistics of this type.

In this paper we examine a statistic similar to some others in that the number of failures in a subset of the individuals in a study is compared to the sum of their conditional probabilities of failure at observed failure times. The choice of the subset for each technique is determined by the type of deviation from model assumptions that is emphasised. In Schoenfeld (1980), for example, to detect arbitrary deviations, any independently chosen partition of covariate space can define several subsets. In Thomas (1983), for discontinuity in a dose-response curve, the location of the discontinuity separates the individuals into subsets above and below it. In work related to that described here, Arjas (1987) used the values of a particular covariate or stratification variable for such a definition. Our choice of subset emphasises the difference between the exponential and the linear relative risk functions.

This difference is depicted in Fig. 1. We fitted the Cox model (1) to simulated data that arose from linear relative risks with two continuous covariates. The two sets of points show the fitted conditional probabilities for the individuals at risk at two observed failure times (at the beginning and near the middle of the study), plotted against their probabilities under the true model. Note that maximum partial likelihood estimation, even under the incorrect model, produces probabilities that approximate

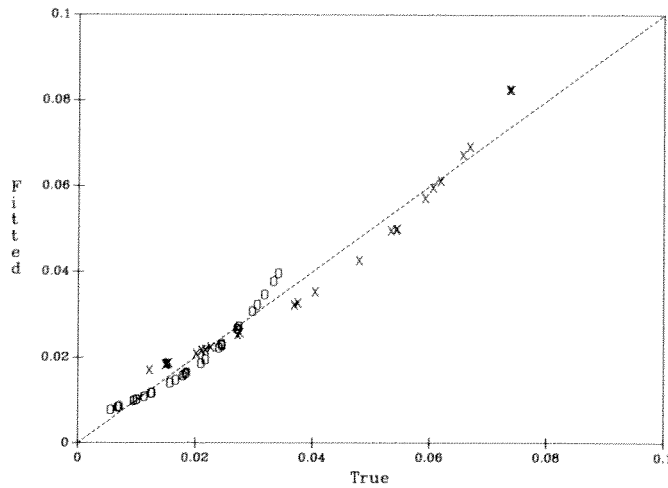


Fig. 1. Fitted exponential *vs* true linear conditional probabilities for all individuals at risk in two risk sets from simulated data

to some degree the true values. However, the fitted probability appears clearly as a convex function of the true probability, over any interval in η , due directly to the convexity of the exponential relative risk function. This results in overestimation of the probability for individuals with 'extreme' values of η and underestimation for individuals with intermediate values. Since the true risks are unknown in practice, our diagnostic statistic compares failure probabilities in the fitted multiplicative risk model with a corresponding failure count. The definitions appear in Section 2. Simulation results presented in Section 3 indicate the agreement of the distribution of the statistic with an approximate null distribution and its power to detect linear relative risks. The application to a published data set is reported in Section 4.

The convexity in Fig. 1 would be less apparent if the true relative risks were, in the terms of Prentice *et al.* (1983), superadditive but submultiplicative over the entire range of η . Although the linear relative risk function is the primary alternative against which a test is suggested, it should have some power for any underlying pattern of relative risks which is less convex in η than the exponential.

2. Notation and Definition of the Statistic

Consider a set of n individuals, indexed by j , $1 \leq j \leq n$. With each individual, we associate an observation (T_j, δ_j) , where T_j is the failure time or the censoring time, whichever occurs first, and

$$\delta_j = \begin{cases} 1 & \text{if } j \text{ fails at } T_j \\ 0 & \text{if } j \text{ is censored at } T_j. \end{cases}$$

We hypothesise that the failure intensity of j at time t can be expressed in the form

$$\lambda_j(t) = Y_j(t) \cdot \lambda_0(t) e^{\beta' z_j},$$

where $Y_j(t) = 1_{\{T_j \geq t\}}$ is the indicator of being at risk at time t . In this simple setting tied failure times are ruled out.

Let P^β be a probability which is in agreement with the above assumptions and write $P = P^{\beta_0}$. We denote by $K = \sum_{j=1}^n \delta_j$ the number of observed failure times, by $T_{(1)} < T_{(2)} < \dots < T_{(K)}$ the ordered failure times, and by $R(T_{(i)}) = \{1 \leq j \leq n : T_j \geq T_{(i)}\}$ the risk set at $T_{(i)}$.

The partial likelihood expression L^β of Cox (1972, 1975) is a product of terms of the form

$$p_j^\beta(i) \stackrel{\text{def}}{=} \frac{Y_j(T_{(i)}) \cdot e^{\beta'z_j}}{\sum_{k \in R(T_{(i)})} e^{\beta'z_k}}, \quad 1 \leq i \leq K, \quad 1 \leq j \leq n,$$

where the product is taken over those values of i and j for which individual j fails at $T_{(i)}$. It is well known that $p_j^\beta(i)$ can be viewed as the conditional P^β -probability that the j th individual fails at $T_{(i)}$, given the risk set, the corresponding covariate values, and the fact that exactly one individual fails at $T_{(i)}$.

As mentioned in the introduction, our diagnostic method is based on selecting a subset of individuals whose values of $p_j^\beta(i)$ emphasise the convexity *vs* linearity of the relative risk function. These are the individuals with linear predictors $\eta_j = \beta'z_j$ near the ends of the range of η over all the individuals. Here we look at one precise way of defining such sets: fixing some value of β , write

$$G^\beta(x) = n^{-1} \sum_{j=1}^n 1_{\{\beta'z_j \leq x\}}$$

for the empirical distribution of the sample $\{\beta'z_j : 1 \leq j \leq n\}$ of linear predictors corresponding to all individuals. Then denote

$$I = I^\beta = \{1 \leq j \leq n : G^\beta(\beta'z_j) \leq \frac{1}{4} \text{ or } G^\beta(\beta'z_j) \geq \frac{3}{4}\}. \quad (3)$$

The proportions $\frac{1}{4}$ and $\frac{3}{4}$ are specified as reasonable values for a broad range of failure and censoring parameters. Thus it is not a cause for concern if those proportions are only approximately attained, as in our example with real data in Section 4. Corresponding to this I , define the counting variables $N_I(i)$, $1 \leq i \leq K$, by $N_I(i) = 1$ if the failing individual at $T_{(i)}$ belongs to I and $N_I(i) = 0$ otherwise. The sums $p_I^\beta(i) = \sum_{j \in I} p_j^\beta(i)$, $1 \leq i \leq K$, are then conditional expected values of the corresponding $N_I(i)$, in the same sense as the $p_j^\beta(i)$ above. We now make a direct comparison between the counting variables $N_I(i)$ and their expected values $p_I^\beta(i)$. Considering $\beta = \beta_0$, we define the cumulative differences

$$M_k = \sum_{i \leq k} [N_I(i) - p_I^{\beta_0}(i)], \quad 1 \leq k \leq K.$$

Apart from a normalising factor, M_K is going to be our test statistic. Negative values of M_K are an indication that the model assigns too high relative risks, and therefore failure probabilities, to individuals with extreme values of the linear predictor.

To determine the normalisation, note first that the conditional P -variance of $N_I(i)$ is obviously $p_I^{\beta_0}(i)[1 - p_I^{\beta_0}(i)]$. Although the increments of $(M_k)_{1 \leq k \leq K}$ are not independent, they are uncorrelated, and the 'correct' variance expression for the asymptotic normality of M_K , as $K \rightarrow \infty$, is given by the sum

$$V_K = \sum_{i \leq K} p_I^{\beta_0}(i)[1 - p_I^{\beta_0}(i)].$$

A formal justification for this is that $(M_k)_{1 \leq k \leq K}$ is a P -martingale (stopped at K) with respect to a natural embedded history, while $(V_k)_{1 \leq k \leq K}$ with $V_k = \sum_{i \leq k} p_I^{\beta_0(i)} [1 - p_I^{\beta_0(i)}]$ is the corresponding variance process. (For an excellent introduction into the martingale ideas associated with Cox's model, see Gill (1984) and for general background, see for example Gill (1980).) By assuming a set of mild regularity conditions and then applying Rebolledo's central limit theorem (see Andersen and Gill (1982)), one concludes easily that $M_K/\sqrt{(V_K)}$ approaches in distribution the standardised normal law as $K \rightarrow \infty$.

In practice, β_0 is unknown and we must use the partial ML-estimate $\hat{\beta}$ instead. We are thus led to consider the statistic

$$D = \frac{\sum_{i=1}^K [N_I(i) - \hat{p}_I(i)]}{\sqrt{\left\{ \sum_{i=1}^K \hat{p}_I(i) [1 - \hat{p}_I(i)] \right\}}}, \quad (4)$$

where the definition of I is based on $\beta = \hat{\beta}$, and $\hat{p}_I(i)$ denotes $p_I^{\hat{\beta}(i)}$. Unfortunately this change also destroys our direct martingale argument for proving the asymptotic normality of the test statistic. We have been unable to overcome this mathematical difficulty and must, therefore, rely on simulation results. The next section describes the simulations in detail.

Our starting point, then, is that if the multiplicative model (1) is correct, the test statistic D is approximately normal. If, on the other hand, the relative risk function is additive and (1) is fitted, we can expect that the estimated probabilities $\hat{p}_I(i)$ are often too large to balance against the counting variables $N_I(i)$. Consequently, we use critical regions of the form $\{D < d\}$.

3. Simulation Results

We conducted a series of simulations to test the normality of the null distribution of the statistic in equation (4) and to assess its power. Sample sizes $n = 500$ and $n = 100$ were run with no censoring; sample size $n = 200$ was run with the largest 50% of the failure times censored. Four covariate distributions, two univariate and two bivariate, were used:

- (a) normal – a pseudorandom sample of size n from the truncated normal with mean 0.5 and standard deviation 0.166, truncated at 0 and 1.
- (b) uniform – $z_j = j/(n-1)$, $0 \leq j \leq n-1$.
- (c) normal \times binary – z_1 normal as in (a); $z_2 \in \{0, 1\}$, each value taken $n/2$ times, uncorrelated with z_1 .
- (d) binary \times binary – $(z_1, z_2) \in \{(0, 0), (0, 1), (1, 0), (1, 1)\}$, each value taken $n/4$ times.

When the estimated range of relative risks over the individuals in a study is small, e.g. when $\hat{\beta}$ is close to 0, the difference between the exponential and linear functions, the power of a test to distinguish them, and the need to do so are all correspondingly small. Thus we chose true parameter values that make the covariates strong risk factors for failure. This should make it possible to detect the convexity evident in Fig. 1 when the additive model is the correct one.

Since the baseline hazard $\lambda_0(t)$ is irrelevant to the estimation procedure, it was fixed at 1, and exponentially distributed failure times were generated. The null distribution

of the statistic was tested by fitting the Cox model to failure times using the exponential as the true relative risk function. The proportion of the replicates where D fell into the lower 0.1 tail of the standard normal is given in Table 1. We chose the 10% level test for indicating model inadequacy rather than a more stringent one which would heavily favour the retention of the multiplicative relative risk model. All the results are based on 1000 replicates, except for one case ($n = 100$, binary \times binary, $\beta'_0 = (4, 1)$) where five of the replicates had monotone likelihoods.

The 10th percentile of the distribution of $M_K/\sqrt{(V_K)}$ appears from Table 1(a) to be close to that of the normal when the true relative risk function is the exponential. The test based on D tends to be conservative, however. The extremely low rejection probability in the binary \times binary case with $n = 200$ and 50% censoring is due to the standard deviations being approximately 0.8 and 0.5 rather than 1 in that case. This is discussed in the last section.

We also generated failure times with the linear as the true relative risk function to estimate the power of the test. The results for some values of β_0 that produced moderate power are reported in Table 2. The range of relative risks in the univariate cases with the linear model (e.g. 1–33 with $\beta_0 = 32$) is somewhat smaller than with the exponential (1–55 with $\beta_0 = 4$). In the bivariate cases, the range with the linear is much smaller. Note that in the normal \times binary case, with the second coefficient

TABLE 1
Proportions of replicates with indicated statistic less than
–1.282 for one-sided test at 10% level

a) $M_K/\sqrt{(V_K)}$					
n	$\beta_0 =$	Uniform		Normal	
		2	4	2	4
500		.098	.099	.100	.099
100		.111	.112	.111	.112
200*		.097	.083	.094	.093
n	$\beta'_0 =$	Normal \times Binary		Binary \times Binary	
		(2, 1)	(4, 1)	(2, 1)	(4, 1)
500		.108	.106	.094	.103
100		.108	.099	.091	.094
200*		.110	.112	.096	.087

b) D , from equation (4)					
n	$\beta_0 =$	Uniform		Normal	
		2	4	2	4
500		.098	.095	.098	.097
100		.112	.112	.110	.115
200*		.095	.066	.092	.082
n	$\beta'_0 =$	Normal \times Binary		Binary \times Binary	
		(2, 1)	(4, 1)	(2, 1)	(4, 1)
500		.109	.095	.092	.102
100		.113	.104	.106	.100
200*		.096	.083	.040	.000

* 50% censored.

TABLE 2
Proportions of replicates rejected by one-sided 10%-level test using D with true linear relative risk function

		Uniform			
<i>n</i>	$\beta_0 =$	4	8	16	32
500		.60	.87	.98	.99
100		.27	.43	.58	.68
200*		.25	.35	.45	.54
		Normal			
<i>n</i>	$\beta_0 =$	4	8	16	32
500		.22	.30	.37	.42
100		.15	.18	.20	.21
200*		.14	.16	.17	.18
		Normal \times Binary			
<i>n</i>	$\beta_0 =$	(4, 8)	(8, 8)	(16, 8)	(32, 8)
500		.56	.60	.51	.42
100		.25	.26	.23	.19
200*		.18	.21	.22	.18
		Binary \times Binary			
<i>n</i>	$\beta_0 =$	(2, 1)	(4, 1)	(8, 1)	(16, 1)
500		.83	.93	.97	.98
100		.38	.48	.54	.56
200*		.33	.36	.30	.18

* 50% censored.

held fixed, the power ultimately decreases as the coefficient of the normally distributed covariate increases. This results from the trade-off between the increasing range of the bimodal distribution of relative risks and its decreasing variance relative to the range.

With the binary \times binary covariate distribution, this nonparametric test can be compared to the likelihood ratio test of the null hypothesis $\beta_3 = 0$ for the coefficient β_3 of the interaction covariate $z_3 = z_1 \cdot z_2$. The power for the latter, seen in Table 3, is not a great deal higher, except in the censored case where the test derived here is extremely conservative. Even in this case, the test has some power for small- to moderate-sized risks.

The covariate distributions were chosen for similarity to data observed in practice. Higher powers may be found with distributions designed specifically for testing the exponential against the linear model. At the suggestion of a referee, we conducted

TABLE 3
Proportion of binary \times binary replicates rejected by likelihood ratio test at 10%-level with true linear relative risk function

<i>n</i>	$\beta_0 =$	(2, 1)	(4, 1)	(8, 1)	(16, 1)
500		.83	.93	.97	.98
100		.40	.50	.56	.59
200*		.38	.45	.48	.44

* 50% censored.

TABLE 4

Proportions of replicates rejected by one-sided 10%-level test using D with true linear relative risk function and discrete covariate distributions having mass p_1 , p_2 , p_3 at $z = 0, 0.5, 1$, respectively

(p_1, p_2, p_3)	$n = 100$			$n = 200^*$			
	$\beta_0 =$	4	8	16	4	8	16
(.20, .60, .20)		.52	.83	.98	.46	.75	.93
(.25, .50, .25)		.56	.86	.98	.50	.78	.95
(.33, .34, .33)		.53	.85	.98	.48	.80	.96
(.50, .30, .20)		.52	.84	.98	.55	.88	.99
(.60, .20, .20)		.44	.77	.96	.51	.84	.99
(.60, .30, .10)		.41	.72	.94	.52	.87	.99

* 50% censored.

some simulations with discrete covariate distributions having proportions p_1 of their mass at $z = 0$, p_2 at $z = 0.5$ and the remainder p_3 at $z = 1$. Selected results appear in Table 4. With $n = 500$, the estimated power was over 0.95 with $\beta_0 \geq 4$ for each set of proportions shown. The power is not highly sensitive to small imbalances in the proportions, but the trends indicate that more low-risk individuals are needed with the censoring used here.

4. Application to a Real Data Set

As an example of the use of this method, we analysed the subset of the Radiation Therapy Oncology Group data listed in Appendix I of Kalbfleisch and Prentice (1980). Survival or censoring times for 195 individuals with cancers of the oropharynx are given. In our model, we included three covariates that were important in the full data set: general condition (z_1), T staging of the tumour (z_2) and N staging of the regional lymph nodes (z_3). Two individuals were omitted because of apparently missing values for z_1 . This left 193 individuals in the study, of whom 53 had censored survival times.

Some simplification of these covariates improved the fit. The T and N classifications were reduced to indicator variables for the most severe conditions (large invasive tumour and multiple positive nodes, respectively), and the one individual with the poorest general condition (class 4) was combined with those in class 3 to avoid having a highly influential observation. Of the 12 distinct covariate vectors possible, 11 were observed. Such discreteness makes the grouping of the highest and lowest 25% of the relative risk values, as specified in equation (3), a goal that is often unattainable. In our attempt to meet this goal with these data, the set $I^{\hat{\beta}}$ indexed the 35 highest and 57 lowest risks. Since the optimal fractions evidently depend on the censoring, it is difficult to see what effect this might have here.

The results of maximising the partial likelihood of these data under the Cox model (1) were the coefficient estimate $\hat{\beta}^t = (0.956, 0.456, 0.328)$, at which point the log-likelihood equals -629.062 , and the test statistic $D = -2.28$. The value of D should alert us to the need for a less convex relative risk function than the exponential. Thus we were directed to fit the linear relative risk function, yielding $\hat{\beta}^t = (3.150, 1.094, 0.790)$, with a maximum log-likelihood of -627.137 . At this stage, we could make a

more formal inference by embedding these two in any parametric family and performing a likelihood ratio test, with the null hypothesis that the true relative risk function is the exponential against the alternative that it is the linear. Such a test gives an approximate p -value of 0.05. Note that the values 0, 1 and 2 must be assigned to z_1 , rather than 1, 2 and 3, to obtain finite estimates with the linear relative risk function.

The implication of this for the estimated relative risks is shown in Fig. 2. We plotted against each other the relative risks under the linear and the exponential models for the observed covariate values, at the respective $\hat{\beta}$ s. To standardise the arbitrary scale of each set of relative risks, each value is divided by the geometric mean of the relative risks of the 193 individuals. The set of points at which the exponential relative risks exceed the linear then corresponds roughly to the grouping of the indices in I^β . In particular, the largest relative risks may be greatly overestimated in the exponential model.

5. Discussion

Our goal has been to provide a method for testing the multiplicativity versus additivity of the relative risk function. We have aimed at simplicity rather than complete generality or maximal power. Once the multiplicative model has been fitted, no further estimation of alternative models is needed to calculate the test statistic D .

We emphasise that finding a suitable relative risk function $r(\cdot)$ is not the only variation in the general proportional hazards model (2) needed to apply it to a given set of data. Even when the covariate effects are adequately summarised by a linear predictor, it may still, for example, be necessary to stratify over a set of different baseline hazards. Our technique can easily be extended to this form of the partial likelihood. If, however, the correct fitting would require the use of time-dependent covariates, the definition of the set I would also have to be made to depend on time. While such an extended definition is technically feasible, we will not pursue it in this paper.

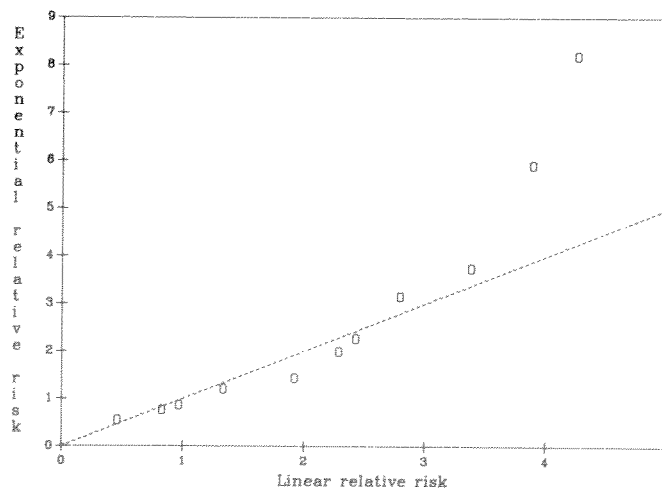


Fig. 2. Fitted exponential vs fitted linear relative risks for observed covariate values in Radiation Therapy Oncology Group data

In Table 2, the highest overall power was achieved with $n = 500$ and binary \times binary covariates. Censoring decreased the power, substantially in some cases. For example, for the binary \times binary covariates with $\beta_0 = (16, 1)$, there is very little convexity to detect as long as the true relative risks take values 1, 2, 17 and 18. The 50% censoring compounds this difficulty, however, by ending almost every simulation while a few individuals with high relative risk remain at risk. The contribution to the $\hat{p}_I(i)$ from individuals with low relative risk is significant in only a very few risk sets, so that $\hat{p}_I(i)$ is consistently close to 0.5, even if the (wrong) multiplicative model is used. This leads to low power in the test, and may also account for the reduced variance of D when the multiplicative model is correct. It appears that the censoring falling much more heavily on low-risk individuals causes this problem, and that a different censoring mechanism, such as independent censoring, would make some improvement. Thus the general rule drawn from Table 4 applies here as well: for possible overestimation to be detected by this method, there should be sufficient proportions of high- and low-risk individuals at risk throughout the study.

It is plausible that modifying the definition of the set I would improve the power of the test in general. The maximum would be achieved if the relative risks that were overestimated at the upper and lower ends of the range could be precisely identified. This would necessarily depend on $\hat{\beta}$ and be more complicated. Some experimentation with different definitions and the results in Table 4 lead us to believe that the increase in power would not be large in any case.

As noted in the introduction, our statistic is one of several proposed for tests of model specification. Some studies of the small-sample properties of such tests have been reported, e.g. Barlow (1986). In Venzon (1987), when the true relative risks were linear, the power of the test using one version of the statistic of Schoenfeld (1980) was lower than the score test in a parametric family. A more comprehensive comparative study of some of these tests for common covariate distributions and parameter values would now be of use.

Acknowledgement

This research was supported in part by grant GM-24472 from the National Institute of General Medical Sciences and by USPHS grant CA-39949 from the National Institutes of Health of the USA. The authors wish to thank the referees for useful comments.

References

- Andersen, P. K. and Gill, R. D. (1982) Cox's regression model for counting processes: a large sample study. *Ann. Statist.*, **10**, 1100-1120.
- Arjas, E. (1987) A graphical method for assessing goodness-of-fit in Cox's proportional hazards model. *J. Amer. Statist. Ass.* (in press).
- Barlow, W. E. (1986) General relative risk models for stratified epidemiologic studies. Ph.D. dissertation, University of Washington.
- Breslow, N. E. (1986) Use of the power transform to discriminate between additive and multiplicative models in epidemiologic research. In *Modern Statistical Methods in Chronic Disease Epidemiology*, (S. H. Moolgavkar and R. L. Prentice, eds). New York: Wiley.
- Breslow, N. E. and Storer, B. E. (1985) General relative risk functions for case-control studies. *Amer. J. Epid.*, **122**, 149-162.
- Cox, D. R. (1972) Regression models and life tables (with discussion). *J. R. Statist. Soc. B*, **34**, 187-220.
- (1975) Partial likelihood. *Biometrika*, **62**, 269-276.
- Gill, R. D. (1980) *Censoring and Stochastic Integrals*. Amsterdam: Mathematical Centre.
- (1984) Understanding Cox's regression model: a martingale approach. *J. Amer. Statist. Ass.*, **79**, 441-447.

- Guerrero, V. M. and Johnson, R. A. (1982) Use of the Box-Cox transformation with binary response models. *Biometrika*, **69**, 309-314.
- Kalbfleisch, J. D. and Prentice, R. L. (1980) *The Statistical Analysis of Failure Time Data*. New York: Wiley.
- Kay, R. (1977) Proportional hazard regression models and the analysis of censored survival data. *Appl. Statist.*, **26**, 227-237.
- Lustbader, E. D., Moolgavkar, S. H. and Venzon, D. J. (1984) Tests of the null hypothesis in case-control studies. *Biometrics*, **40**, 1017-1024.
- Moolgavkar, S. H. and Venzon, D. J. (1987) Confidence regions in curved exponential families: application to matched case-control and survival studies with general relative risk function. *Ann. Statist.*, **15**, 346-359.
- Nelder, J. A. and Wedderburn, R. W. M. (1972) Generalized linear models. *J. R. Statist. Soc. A*, **135**, 370-384.
- Pregibon, D. (1980) Goodness of link tests for generalized linear models. *Appl. Statist.*, **29**, 15-24.
- Prentice, R. L. and Mason, M. W. (1986) On the application of linear relative risk regression models. *Biometrics*, **42**, 109-120.
- Prentice, R. L. and Self, S. G. (1983) Asymptotic distribution theory for Cox-type regression models with general relative risk form. *Ann. Statist.*, **11**, 804-813.
- Prentice, R. L., Yoshimoto, Y. and Mason, M. W. (1983) Relationship of cigarette smoking and radiation exposure to cancer mortality in Hiroshima and Nagasaki. *JNCI*, **70**, 611-622.
- Rothman, K. J., Greenland, S. and Walker, A. M. (1980) Concepts of interaction. *Amer. J. Epid.*, **112**, 467-470.
- Schoenfeld, D. (1980) Chi-squared goodness-of-fit tests for the proportional hazards regression model. *Biometrika*, **67**, 145-153.
- Storer, B. E., Wacholder, S. and Breslow, N. E. (1983) Maximum likelihood fitting of general risk models to stratified data. *Appl. Statist.*, **32**, 172-181.
- Thomas, D. C. (1981) General relative-risk models for survival time and matched case-control analysis. *Biometrics*, **37**, 673-686.
- (1983) Nonparametric estimation and tests of fit for dose-response relations. *Biometrics*, **39**, 263-268.
- Venzon, D. J. (1987) Parameter estimation in the proportional hazards model with general relative risk functions. Ph.D. dissertation, Program in Applied and Computational Mathematics, Princeton University.

2

3

4

5