0270-9139/87/0706-1346$02.00/0
HEPATOLOGY

Vol. 7, No. 6, pp. 1346–1358, 1987
Printed in U.S.A.

Special Articles

Multivariate Survival Analysis Using Cox's Regression Model

ERIK CHRISTENSEN

Medical Department A, Division of Hepatology, Rigshospitalet, Medical Department, Division of Hepatology, Hvidovre Hospital, and Medical Department F, Gentofte Hospital, University of Copenhagen, DK-2900 Copenhagen, Denmark

In recent years, the Cox regression model has been used increasingly for analysis of censored survival data. With this model, the pattern of association (covariation) of many predictor variables with survival is analyzed to identify the combination of variables which best predicts survival. The results can be presented as a "pocket chart," by which a prognostic index for a new subject can easily be obtained. By a simple graph, the prognostic index can be translated to estimates of the probability of surviving a given time or the median survival time predicted for the subject. In controlled clinical trials, the Cox model can be used to adjust for imbalance in variables influencing prognosis and to identify variables being associated with the treatment effect (therapeutic variables). This paper describes in rather simple and practical terms some of the concepts behind the model, how to perform the analyses and how to interpret and utilize the results.

SURVIVAL ANALYSIS

Like everything else, diseases develop and progress in time. Description of the course in time is an important aspect in the characterization of diseases, including their prognosis and the effects of therapies. However, a detailed description of the course of disease may be complex (1–4). Accordingly, the problem has been dealt with in simpler terms, namely by analyzing for each individual the time from a defined starting point, e.g., the time of diagnosis or randomization in a controlled clinical trial, to the occurrence of an *event* or *endpoint* of interest, traditionally death as in *survival analysis* (5). In principle, the first occurrence of other events such as complication, freedom of symptoms, recurrence of symptoms, diastolic blood pressure > 110 mm Hg, hemoglobin < 6 mmoles per liter, etc., may be defined as an endpoint for the subject and analyzed in a similar way (3, 5), although a precise registration of the time at which the event takes place may be difficult.

Since investigations have a limited duration, some subjects may not yet have had the event, but are still "alive" at the end of the investigation. Other subjects, while alive, may have dropped out for various reasons during the study without the event having occurred. Such "incomplete" survival times from the starting point to the latest observation, so called *censored* survival times, which hold the information that the event did *not* occur while the individual was being observed, are utilized along with the "complete" survival times in the Cox regression model (6) and other recent methods for analysis of survival data (7).

Table 1 presents a constructed set of survival data which will be used for illustration in this paper. The data set includes 30 subjects, of whom 18 have a complete observation time with an endpoint, and 12 have a censored observation time without an endpoint. The values of the variables presented (albumin, bilirubin and alcoholism) apply to the beginning of the follow-up period.

THE SURVIVAL CURVE

The established way of presenting survival data is to estimate the survival curve. If all of the survival times are complete, i.e., without censoring, the survival curve is estimated simply as the proportion of individuals in whom the event has not yet occurred at each point of time during the observation period. For survival data which includes censored survival times, the survival curve may be estimated by the method described by Kaplan and Meier (8) and illustrated by examples from Peto et al. (9). By including the censored survival times, that method gives a useful estimate of the probability of not having the event (i.e., to survive) as a function of time. Since this probability is a function of the probability of surviving, all time intervals from start to a given time t are denoted by the term *cumulative survival probability*, which is commonly designated S(t). The estimated cumulative survival probability curve $\hat{S}(t)$ for the total group of individuals presented in Table 1 is shown in Figure 1 (top panel).

HAZARD

The more recent methods for analysis of survival data, including Cox regression analysis (6–7, 10), are based on the instantaneous *hazard* (also called the *force of mortality*) designated $\lambda(t)$, which is the risk that the event will occur for a subject in a small time interval (Δt) at time t, given the subject did not have the event before that time. Since the hazard $\lambda(t)$ is the derivative of the *cumulative (integrated) hazard* designated $\Lambda(t)$ (7), it can be illustrated by the slope of the latter. Because the relation $\Lambda(t) = -\log_e S(t)$ (7), the cumulative hazard may easily be estimated by taking the negative natural logarithm of the corresponding cumulative survival probability estimates.

The estimated cumulative hazard curve $\Lambda(t)$ of the survival data in Table 1 is shown in the bottom panel of Figure 1. A steep rise in that curve corresponds to a high

Address reprint requests to: Erik Christensen, M.D., Medical Department F, Gentofte Hospital, University of Copenhagen, Niels Andersens vej 65, DK-2900 Copenhagen, Hellerup, Denmark.

TABLE 1. Constructed set of survival data

Subject no.	Survival time (days)	Death (1) or censoring (0)	Albumin (gm/liter)	Bilirubin (μmoles/liter)	Alcoholism [present (1)/absent (0)]
1	17	1	24	332	1
2	23	1	23	157	1
3	39	1	22	182	1
4	45	1	24	77	1
5	56	1	21	92	1
6	69	1	26	143	0
7	80	1	26	32	1
8	98	1	21	249	1
9	120	1	29	72	0
10	134	1	29	220	1
11	152	0	32	89	1
12	163	1	29	152	0
13	189	1	28	43	1
14	205	1	31	82	0
15	231	0	27	39	0
16	252	0	31	63	1
17	311	1	31	98	0
18	337	0	28	41	0
19	390	1	33	68	1
20	457	1	31	25	1
21	488	0	34	51	1
22	560	1	33	57	0
23	633	0	34	70	0
24	692	0	35	39	0
25	809	0	32	32	0
26	912	1	34	67	0
27	1,046	0	33	52	0
28	1,298	0	33	20	0
29	1,437	0	36	28	0
30	1,562	0	35	19	0

Note: the recordings for albumin, bilirubin and alcoholism apply to the beginning of the follow-up period.

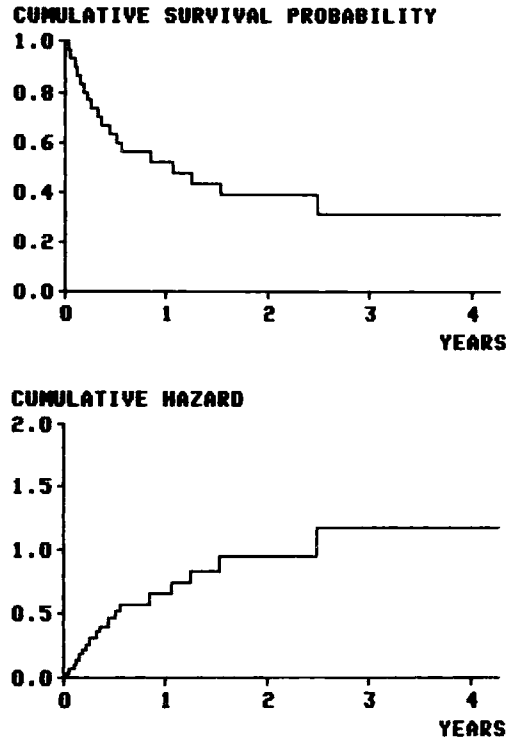

FIG. 1. Estimated cumulative survival probability $\hat{S}(t)$ [Kaplan-Meier plot (8)] (*top*) and estimated cumulative hazard $\hat{\Lambda}(t)$ (*bottom*) for the survival data presented in Table 1. One can estimate the one from the other using the relations: $\hat{\Lambda}(t) = -\log_e \hat{S}(t)$ and $\hat{S}(t) = e^{-\hat{\Lambda}(t)}$.

hazard, a slight rise to a low hazard. It appears from the curve that the hazard is high initially and less thereafter.

A cumulative survival curve and the cumulative hazard curve derived from it are summarizing descriptions concerning the studied total group of individuals. However, there may be a wide variation in the survival time (and hazard) between individual subjects. Although the curves illustrate the variation among the subjects, they do not allow identification of who had a long survival (low hazard) and who had a short survival (high hazard).

COVARIATES

To make such an identification possible or to allow prediction of survival time in individual subjects, it is necessary to identify and utilize variables *covarying* with survival. For example, it may be that serum albumin at the starting point covaries with the subsequent survival time; i.e., in subjects with a low albumin, the survival time may be short (hazard high), and in subjects with a high albumin, the survival time may be long (hazard low). If the covariation (or correlation) between the level of albumin and the survival time is large, the level of albumin may to some degree "*explain*" the variation in survival time or hazard between the subjects (11). In that case, the level of serum albumin in a new subject may to some degree be used to *predict* his/her survival time or hazard. In a controlled clinical trial, the treatment given may be an important covariate which may "explain" a difference in survival between the treatment groups.

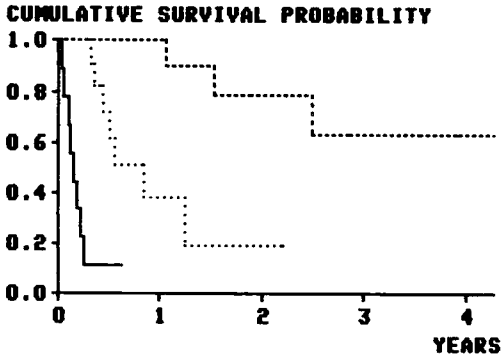

FIG. 2. Cumulative survival probability for the data presented in Table 1 in subgroups defined according to the level of serum albumin: —— = albumin ≤ 27 gm per liter; · · · · = 27 gm per liter < albumin ≤ 32 gm per liter; and - - - - = albumin > 32 gm per liter. The difference in survival between the groups is marked, indicating that the level of albumin to some degree can predict survival.

COMPARISON OF SUBGROUPS (STRATA)

Thus, the simplest way to identify prognostic covariates is to divide the subjects in subgroups (strata) according to different levels of a given variable. For example, if we divide the subjects presented in Table 1 according to the level of albumin (e.g., albumin < 28, 28 to 32 and >32 gm per liter), it can be shown in Figure 2 that the survival curves for these three groups are markedly different. In a similar way, one can show for the subjects presented in Table 1 that the survival curves are different in subgroups defined according to the level of bilirubin or the presence or absence of alcoholism.

Normally, a single variable, even if it shows a strong covariation with survival, will not completely "explain" survival. Usually, it is to be expected that more variables in combination may "explain" survival to a higher degree.

It is possible to stratify according to more than one variable at a time (3, 9). However, with an increasing number of strata, the number of subjects in each stratum will rapidly decrease to such an extent that the corresponding survival curves will have too little "confidence" [the curves will have too wide confidence limits (9)] to be of any value. Hence, in practice, stratified analyses can only be performed with one or few variables at a time. This puts a serious limitation on stratification. However, the method may be used for a crude screening to identify variables which should be analyzed further in a Cox regression model.

COX REGRESSION MODEL

The regression model proposed by Cox (6) is a multiple regression model for analysis of censored survival data. Provided that the more strict assumptions (described later) of this model may be considered fulfilled, it may be used to study and utilize the pattern of covariation of many variables with the hazard. The Cox regression model has this form:

$$\lambda(t, z) = \lambda_0(t) \exp(b_1z_1 + \cdots + b_iz_i + \cdots + b_pz_p).$$

Thus $\lambda(t, z)$, the hazard at time t after a defined starting point [diagnosis, randomization etc. (being time zero)] for an individual with variables $z = (z_1 \cdots z_i \cdots z_p)$ is being "dependent on" or "explained" or "predicted" by $\lambda_0(t)$, the so-called underlying hazard at time t, and the predictor variables z_1 to z_p (recorded at time zero), each variable z_i being multiplied by a corresponding regression coefficient b_i. Here, exp stands for exponential function, e.g., $\exp(bz) = e^{bz}$. The underlying hazard $\lambda_0(t)$ may be considered a "reference" hazard from which the hazard $\lambda(t, z)$ at time t of given subject may be obtained by multiplication with a factor, namely the exponential function of the subject's variables "weighted" by the regression coefficients. Formally, the underlying hazard $\lambda_0(t)$ is the hazard at time t of an individual whose z_i's are all zero. Usually, $\lambda_0(t)$ is of little interest in itself, since it may depend on the scoring of the variables (7).

Thus, the Cox model assumes that the hazards of any two patients are proportional over time, i.e., the ratio between the hazards is the same at any time t. This does not preclude that the hazard may change over time. Often, the hazard will be relatively high soon after the time of diagnosis and thereafter it may decrease as in Figure 1. However, the Cox model assumes that changes in the hazard of any patient over time will always be proportional to changes in the hazard of any other patient and to changes in the underlying hazard over time.

The amount by which each predictor variable z_i contributes to the prediction of the hazard $\lambda(t, z)$ of an individual depends on the magnitude of the corresponding term b_iz_i. If the term is numerically big, then the contribution is big; if the term is numerically small (close to zero), then the contribution is small.

Consider a Cox model including only one variable and having an underlying hazard $\lambda_0(t)$ of A years $^{-1}$ at a given time t. If the variable z_1 has the value (score) 2 in one subject and 1 in another subject, b_1 being 0.5, then the model assumes that the ratio between the hazards of the two patients is

$$(A \times e^{0.5 \times 2})/(A \times e^{0.5 \times 1}) = (e^{0.5 \times 2})/(e^{0.5 \times 1}) = e^{0.5} = 1.65.$$

This ratio is assumed to be constant over time; it is independent of the actual value A of $\lambda_0(t)$ which may change with the time t.

If b_i had been −0.5, then the ratio would have been

$$(A \times e^{-0.5 \times 2})/(A \times e^{-0.5 \times 1}) = e^{-0.5} = 1/e^{0.5} = 1/1.65 = 0.61.$$

Thus, higher values (scores) of z_i indicate higher hazard (shorter survival) if b_i is positive and vice versa if b_i is negative. If $b_i = 0$, then z_i has no influence on the hazard.

FITTING A COX MODEL TO A SET OF DATA

The estimation of the b coefficients and the underlying hazard in the Cox regression model is complex. The statistical and computational details are described in the literature (6, 7, 12). However, to perform Cox analyses using available standard computer programs (see later data), knowledge of these details is not necessary.

The procedure is illustrated by performing a Cox regression analysis on the data presented in Table 1. In the analysis, serum albumin will be scored by its value in grams per liter and alcoholism as 1 if present and 0 if

TABLE 2. Seven Cox regression analyses of the data set presented in Table 1[a]

Model no.	χ^2 model	d.f.	p model	R^2	Variable(s) included	Regression coefficient b	SE(b)	N.D. [b/SE(b)]	p coefficient
1	30.99	1	<0.0001	0.28	Albumin	−0.42	0.089	−4.71	<0.0001
2	21.24	1	<0.0001	0.18	Log_{10} bilirubin	4.44	1.06	4.17	<0.0001
3	8.79	1	0.003	0.06	Alcoholism	1.55	0.55	2.82	0.005
4	*35.89*	*2*	*<0.0001*	*0.30*	*Albumin*	*−0.35*	*0.10*	*−3.43*	*0.0006*
					Log_{10} bilirubin	*2.36*	*1.11*	*2.12*	*0.03*
5	32.50	2	<0.0001	0.27	Albumin	−0.39	0.094	−4.16	<0.0001
					Alcoholism	0.79	0.64	1.23	0.22
6	25.13	2	<0.0001	0.20	Log_{10} bilirubin	3.88	1.06	3.66	0.0002
					Alcoholism	1.14	0.59	1.93	0.056
7	37.04	3	<0.0001	0.30	Albumin	−0.32	0.11	−3.07	0.002
					Log_{10} bilirubin	2.25	1.11	2.03	0.04
					Alcoholism	0.71	0.66	1.08	0.28

[a] Models 1 to 3 include one predictor variable, Models 4 to 6 include two predictor variables and Model 7 includes three predictor variables.

absent. Serum bilirubin will be scored by $\log_{10}$ of the values in μmoles per liter as in a previously published study (13). For example, if serum bilirubin is 92 μmoles per liter, it will be scored as 1.98227 . . . . Later, it will be shown if this scoring is adequate.

As in simple multiple regression analysis (14), variables may be selected according to certain procedures (forward selection or backward elimination). To illustrate how this works (the details will be explained in the following), Table 2 presents the results of seven Cox analyses comprising all possible combinations of the three variables in Table 1, i.e., three including only one variable (Models 1 to 3), three including two variables (Models 4 to 6) and one including all three variables (Model 7).

OVERALL SIGNIFICANCE OF THE MODEL (LIKELIHOOD RATIO TEST)

Estimation and significance testing of a given Cox model involves the concept of likelihood, meaning the probability of the observed data being "explained" by a certain model.

The overall significance of each model shown in Table 2 is based on the ratio between the likelihood of a model in which the variables show no covariation with the survival time, the b coefficients all being zero, L(0), and the likelihood of the model with the b coefficient(s) obtained by the analysis, L(b) (7), the b coefficient(s) being estimated in such a way that L(b) is as great as possible. Thus, the estimated parameters (the underlying hazard and the coefficients) of a Cox model are so-called "maximum likelihood estimates" (7). The greater the L(b) or the less the *likelihood ratio* L(0)/L(b), the better the model actually "explains" or fits the observed data (7). The significance of each model can be tested statistically using the relation

$$\chi^2 \text{ model} = -2 \times \log_e [L(0)/L(b)] = -2 \times [\log_e L(0) - \log_e L(b)] = 2 \times [\log_e L(b) - \log_e L(0)]$$

with degrees of freedom (d.f.) being equal to the number of coefficients estimated in the model (7). While $\log_e L(0)$ is the same for all of the models, i.e., −52.319, L(b) and hence the χ^2 model depend on the variable(s) included. For example, for Model 1, $\log_e$ L(b) is −36.825, and the χ^2 model = 2[−36.825 − (−52.319)] = 30.988 ≃ 30.99 as shown in Table 2. This high χ^2 value with 1 d.f. is highly significant. Considering the three models in Table 1, including only one variable (Models 1 to 3), it appears that the highest χ^2 model is provided by Model 1, which therefore has the greatest significance of those three models.

INFORMATION CRITERION FOR A COX MODEL

In Table 2 is also shown R^2, which has been proposed as a counterpart to the coefficient of determination of simple multiple regression analysis (15, 16). R^2 for a Cox model may be estimated as (χ^2 model − 2 × p)/[−2 × $\log_e$L(0)], where p is the number of variables in the model (15). The subtraction of 2 × p is made to adjust for the number of parameters being estimated (15, 16). As in simple multiple regression analysis, R^2 lies between 0 and 1. If R^2 is 0, the model is of no value. The closer R^2 is to being 1, the more perfect the hazards of the individuals can be "explained" by the model. Because of the adjustment, R^2 can never be exactly 1. In Table 2, the model having the highest R^2 is Model 4. Thus, Model 4 "explains" best the hazards of the individuals. However, the value of R^2 for that model is only 0.30, indicating that the fit is far from perfect.

SIGNIFICANCE OF EACH INCLUDED VARIABLE

For each variable in each analysis, Table 2 presents the regression coefficient b and the standard error of b [SE(b)], which indicates the "confidence" of the estimated b value and may be used to estimate confidence limits of b. As shown in the table, the significance of each coefficient can be estimated by comparing the normal deviate, N.D. = b/SE(b), with the standardized normal distribution (7). Identical results are obtained by comparing the square of the normal deviate, i.e., $N.D.^2$

with the χ^2 distribution with 1 d.f. {the so-called Wald test [sometimes presented in computer printouts (16)]}.

The relative importance of the variables is given by the numerical value of N.D. The greater the numerical value of N.D., the more significant it is in the model. Considering Model 7 in Table 2, the importance of the variables decrease in this order: albumin; $\log_{10}$ bilirubin and alcoholism (the latter being insignificant).

RELATIVE RISK (HAZARD) PREDICTED BY EACH REGRESSION COEFFICIENT

From the regression coefficient b of a variable, it is possible to estimate relative risks (ratios between hazards) attributable to various levels of that variable, all other variables being unchanged. Considering the dichotomous variable alcoholism, Model 3 predicts the relative risk of alcoholism to nonalcoholism being $e^{1.55 \times 1}/e^{1.55 \times 0} = 4.7/1 = 4.7$.

In Model 4, a 1 gm per liter lower concentration of serum albumin (e.g., 29 gm per liter relative to 30 gm per liter) will be associated with a relative increase in hazard of $e^{-0.35 \times 29}/e^{-0.35 \times 30} = e^{-0.35 \times (-1)} = 1.42$ times. The latter expression shows that the relative risk being associated with a 1 gm per liter lower albumin is independent of the absolute level of albumin. A 3 gm per liter lower albumin is associated with a $1.42 \times 1.42 \times 1.42 = 1.42^3 = 4.26$ times higher risk, and a 2 gm per liter higher albumin is associated with a halved risk $[1/(1.42 \times 1.42) = 1/1.42^2 = 1.42^{-2} = 0.50]$, everything else being unchanged.

Looking at bilirubin in Model 4 in the same way, a doubling in values (e.g., 100 compared to 50 μmoles per liter) corresponding to an increment in $\log_{10}$ bilirubin of 0.3 (2.0 compared to 1.7) is associated with a risk being $e^{2.36 \times 2.0}/e^{2.36 \times 1.7} = e^{2.36 \times 0.3} = 2.0$ times higher or doubled, everything else being unchanged. It is possible to obtain relative risks attributable to more variables in combination by multiplying the relative risks attributable to each variable in the model. The same result may be obtained more simply as the ratio between the exponential functions of the prognostic indices (see later) corresponding to the two sets of values of the variables considered.

The limitation of the relative risk in itself is that it is just relative and does not give an absolute estimate of the survival time or probability of surviving a given span of time for a given subject.

SELECTION OF VARIABLES

With the *forward selection* method, the model is built up step-wise by including at each step the variable giving the largest reduction in the likelihood ratio or equivalently the largest increase in the χ^2 model. Thus, in the first step, albumin (Model 1 in Table 2) would be included because this variable gives the highest significant χ^2 model of all possible models with one variable (Models 1 to 3). In the next step, $\log_{10}$ bilirubin would be added (Model 4) because this variable increases the χ^2 model significantly ($35.89 - 30.99 = 4.90$ with 1 d.f., $p < 0.05$) in contrast to alcoholism (Model 5), which only gives an insignificant increase in the χ^2 model ($32.50 - 30.99 = 1.51$ with 1 d.f., $p > 0.2$). (Here, d.f. is the difference between the number of estimated coefficients in the models being compared.) Inclusion of alcoholism in a model comprising albumin and $\log_{10}$ bilirubin (Model 7) does not lead to a significant increase in the χ^2 model ($37.05 - 35.89 = 1.15$ with 1 d.f., $p > 0.2$). Therefore, Model 4 would be the final model if the forward inclusion technique was used.

Utilizing the *backward elimination* method, one starts with a model which includes all variables, and then insignificant variables are removed step-wise from the model by excluding the most insignificant variable at each step until each remaining variable contributes significantly to the model. Thus, one would start with Model 7 and then remove alcoholism because this variable is insignificant. This would lead to Model 4, which would be the final model because both variables (albumin and $\log_{10}$ bilirubin) are statistically significant.

For the data in Table 1, forward selection and backward elimination of variables lead to the same final model. In more complex analyses, including many variables, the two methods of selection of variables may lead to slightly different final models. Normally, the selection of variables should not be made solely according to automatic rules. The selection process should be guided by the investigator taking into account, among other things, the *a priori* prognostic value of each variable considered.

INFLUENCE OF COVARIATION BETWEEN PREDICTOR VARIABLES

In general, the pattern of covariation between the predictor variables will, to some extent, determine which will be significant in the final Cox regression model. It is not always possible from the results of univariate analyses (including only one variable, e.g., Models 1 to 3 in Table 2) to predict which variables will be significant in a Cox regression model, including more variables (17). If two variables, each of which has shown a significant covariation with survival time by univariate analysis, are strongly intercorrelated, and therefore holding nearly the same information, only one of them may be significant if both are included in the model. On the other hand, a variable which has shown no significant covariation with survival if included as the only variable may be significant if included together with other variables. The reason for this is that multivariate—in contrast to univariate—statistical analyses can adjust for the influence (covariation) of other variables with the variable in question.

Furthermore, the magnitude of the regression coefficient and the degree of significance of each included variable depend on which other variables are also included in the model as shown in Table 2. For example, by comparing Models 3 and 5, it appears that alcoholism is significant if it is the only variable in the model (Model 3); but if albumin is also included (Model 5), the influence of alcoholism is no longer significant. Furthermore, the value of the b coefficients changes from the models, including one variable to the model having both variables [b for albumin changes from −0.42 (Model 1) to −0.39 (Model 5) and b for alcoholism changes from 1.55 (Model 3) to 0.79 (Model 5)]. The reason for these differences is the covariation between the predictor variables (in this

case between albumin and alcoholism). This covariation is adjusted for in the Cox model, so that only the independent association of each variable with the hazard is presented in the estimated model.

If, however, the covariation between two or more predictor variables is very strong (collinearity), the estimated regression coefficients may be affected so much that the results may no longer be interpreted in a simple meaningful way. The solution will often be to include in the model only one variable from the highly correlated set (14).

CHECKING OF MODEL ASSUMPTIONS

This is an important but difficult aspect of Cox regression analyses, where a close cooperation with a qualified statistician is well-justified. For a more detailed description, the reader is referred to the book of Kalbfleisch and Prentice (pp. 87–98) (7) and Refs. (20) and (21). Here, the main principles will be illustrated.

The most important issue is insuring that the assumption of proportional hazards is not violated. Consider two patients (A and B), who for a given predictor variable have the values z_A and z_B, respectively, the difference $z_A - z_B$ being equal to d. If the regression coefficient of the variable in question is b, the Cox model predicts that the ratio between the hazards of the two patients $\lambda_A(t)/\lambda_B(t)$ should be $e^{b \times d}$, keeping the other variables unchanged. This ratio should be constant, irrespective of time t and values z_A and z_B of the predictor variable being high or low. Under the assumption of proportional hazards, the corresponding proportion between the cumulative (integrated) hazards would be the same, i.e.,

$$\Lambda_A(t)/\Lambda_B(t) = e^{b \times d} \text{ or equivalently } \log_e\Lambda_A(t) - \log_e\Lambda_B(t) = b \times d.$$

Thus plots of the logarithm of the cumulative hazards corresponding to values differing by d (all other variables being unchanged) should be parallel and approximately b × d apart vertically (independent of time t) (7, 17–19).

In practice, checking the assumption of proportional hazards is done for each predictor variable at a time by a stratified analysis where the predictor variable in question is only defining a small number of strata but otherwise not included. The defined strata should be equally spaced, i.e., the spacing d between the means of the scored values from one stratum to the next should be approximately the same. In the analysis, regression coefficients (common for all strata) for the remaining variables are estimated. From such a model, one can plot the logarithm of the cumulative hazard against time for each of the strata, keeping the values of the other variables constant. If the curves are approximately parallel and equidistant vertically with a distance of about b × d, then the assumption of proportional hazards may be considered to be met for that variable.

To check the final model (Model 4 in Table 2), such analyses have been performed where one of the variables (albumin or $\log_{10}$ bilirubin) have been included only as a stratified variable in equally spaced strata and the other maintained unchanged in the model. The resulting plots of the logarithm of the cumulative hazard are shown in Figures 3 and 4. The fit seems to be quite good for albumin and somewhat poorer for $\log_{10}$ bilirubin. However, in view of the limited number of subjects included in the analysis, a very good fit may not be expected. Thus, the hypothesis of constant vertical differences between the curves cannot be rejected with this small set of data. To illustrate what would have happened if bilirubin had been scored as μmoles per liter (without logarithmic transformation), Figure 5 shows that with that kind of scoring the curves do not seem equidistant. Therefore, the logarithmic scoring of bilirubin seems the preferable alternative in this case. Other transformations may be necessary in other cases.

The assumption of proportional hazards may also be tested by the goodness of fit test by Andersen (20) or other similar tests (21).

Sometimes, it may not be possible to obtain proportionality. One reason may be that an important predictor

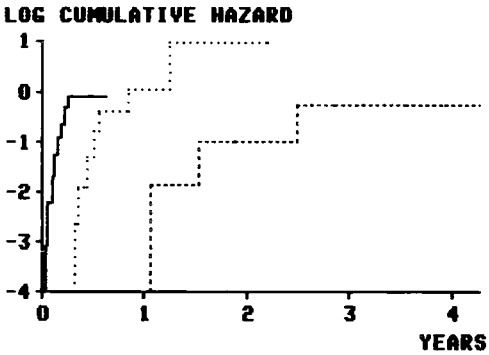

FIG. 3. Plots of the logarithm of the cumulative hazard [$\log_e\hat{\Lambda}(t)$] for a Cox model including albumin in three equally spaced strata: —— = albumin ≤ 27 gm per liter; · · · · = 27 gm per liter < albumin ≤ 32 gm per liter; - - - - = albumin > 32 gm per liter and $\log_{10}$ bilirubin from the data in Table 1. The plots are made for the mean of $\log_{10}$ bilirubin (1.83). Since the means of albumin values for each stratum are about 5 gm per liter apart, then d = 5. Since b for albumin is −0.35 (Model 4 in Table 2), then b × d = −0.35 × 5 = −1.8. In view of the limited number of subjects, the assumption of proportionality, i.e., vertical equidistance of about 1.8 units between the curves, does not seem to be grossly violated.

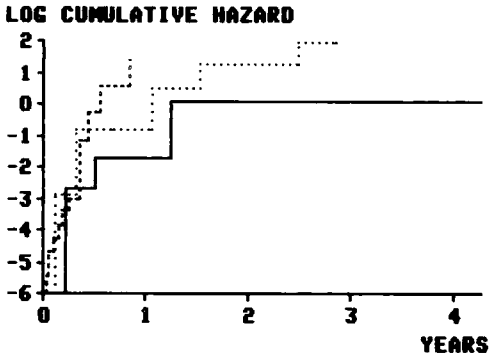

FIG. 4. Plots of the logarithm of the cumulative hazard [$\log_e\hat{\Lambda}(t)$] for a Cox model including $\log_{10}$ bilirubin in three equally spaced strata: —— = $\log_{10}$ bilirubin ≤ 1.7; · · · · = 1.7 < $\log_{10}$ bilirubin ≤ 1.9; and - - - - = $\log_{10}$ bilirubin > 1.9 and albumin from the data in Table 1. The plots are made for the mean of albumin (29.5 gm per liter). Since the means of $\log_{10}$ bilirubin values for each stratum are about 0.34 apart, then d = 0.34. Since be for $\log_{10}$ bilirubin is 2.36 (Model 4 in Table 2), then b × d = 2.36 × 0.34 = 0.8. In view of the limited number of subjects, the assumption of proportionality, i.e., vertical equidistance of about 0.8 between the curves, does not seem to be grossly violated.

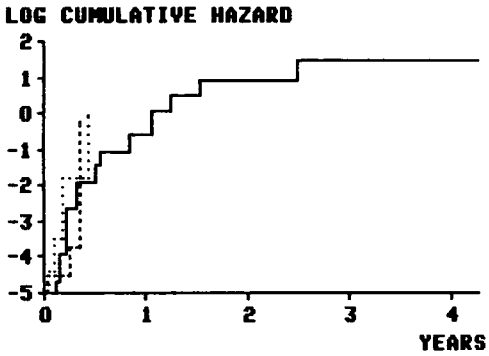

FIG. 5. Plots of the logarithm of the cumulative hazard [$\log_e \hat{\Lambda}(t)$] for a Cox model including bilirubin in three equally spaced strata: —— = bilirubin ≤ 100 μmoles per liter; · · · · = 100 μmoles per liter < bilirubin ≤ 200 μmoles per liter; and – – – – = bilirubin > 200 μmoles per liter and albumin from the data in Table 1. The plots are made for the mean of albumin (29.5 gm per liter). Since b for bilirubin without logarithmic transformation would be 0.0063 and since the means of the bilirubin values for each stratum are about 100 μmoles per liter apart, then d = 100 and b × d = 0.0063 × 100 = 0.6. In contrast to Figure 4, the curves do not seem equidistant vertically with a distance of about 0.6, i.e., the assumption of proportionality does not seem to be fulfilled.

variable is missing from the analysis. The solution would then be to include that variable if possible. Another reason for lack of proportionality may be that some predictor variables interact, i.e., the association of one predictor variable with the hazard depends on the value of another predictor variable. Interaction between predictor variables may be identified by different methods. The most widely used approach is to include multiplicative terms in the model (14). For example, if the association of albumin with hazard depended on whether the individuals were alcoholic or not, this could be tested in a model including the following three variables: albumin; alcoholism, and the multiplicative term albumin × alcoholism (the product of the scorings used). Because the correlation between the interaction variable and its component predictor variables will often be quite high, problems of collinearity leading to "distortion" of the regression coefficients may arise. Furthermore, it would not be feasible to investigate all possible interactions between the predictor variables. One should limit the interaction analyses to those variables which *a priori* may be suspected to interact. If it is not possible to fit a given variable to the proportional hazards assumption, the analysis may be performed stratified, the strata being defined by the variable in question (7, 16).

UTILIZATION OF A COX REGRESSION MODEL TO ESTIMATE PROGNOSIS IN NEW SUBJECTS

PROGNOSTIC INDEX (PI), ITS ESTIMATION AND INTERPRETATION

The term in the Cox regression model by which the variables of the subject affect his/her hazard is $b_1z_1 + \cdots + b_iz_i + \cdots + b_pz_p$. Denoting this expression PI (13, 17, 22, 23), the Cox model equation is simplified to $\lambda(t, z) = \lambda_0(t)\, e^{PI}$.

Estimation of PI for a given subject is the first step in estimating the prognosis of that subject. Considering Model 4 in Table 2 as the final model, the PI for a subject with a serum albumin of 30 gm per liter and a serum bilirubin of 50 μmoles per liter, the PI is estimated as follows:

$$PI = -0.35 \times 30 + 2.36 \times \log_{10} 50 = -0.35 \times 30 + 2.36 \times 1.7 = -6.5.$$

It appears that higher levels of albumin will lead to lower levels of the first term and thus lower levels of PI because the regression coefficient for albumin is negative. In contrast, higher levels of bilirubin will lead to higher values of the second term and hence higher values of PI because the regression coefficient for bilirubin is positive.

Higher values of PI mean higher hazard or shorter survival, lower values mean lower hazard or longer survival. Differences in PI for two patients can be used to estimate their relative risk. A PI being 0.7 higher corresponds to a doubled risk ($e^{0.7}$ per liter = 2). A PI being 0.7 lower corresponds to a halved risk ($e^{-0.7} = 0.5$); except for this, PI cannot be interpreted in a meaningful way by itself, since it depends on the scoring of the variables.

The estimation of PI can be simplified markedly by using a pocket chart (22) as presented in Table 3 in which the regression terms for various values of the variables have already been calculated for Model 4 in Table 2. For example, for a bilirubin of 60 μmoles per

TABLE 3. Pocket chart for easy estimation of PI corresponding to model 4 in Table 2

Variable		Points to add (A)	Points to subtract (B)
Bilirubin (μmoles/liter)	10	24	
	14	27	
	18	30	
	25	33	
	33	36	
	45	39	
	60	42	
	80	45	
	110	48	
	150	51	
	200	54	
	270	57	
	360	60	
Albumin (gm/liter)	20		70
	22		77
	24		84
	26		91
	28		98
	30		105
	32		112
	34		119
	36		126
	38		133
	40		140
Points to be added (A) =			
Points to be subtracted (S) =		←	
A – S			
PI = (A – S)/10 =			

Note: for each variable, only one number should be used. If a patient has values between those in the table, interpolation should be used.

TABLE 4. Ranked individual estimated values of PI based on model 4 in Table 2

Subject no.	Estimated PI
8	−1.7
3	−2.4
1	−2.5
5	−2.7
2	−2.9
4	−3.9
6	−4.0
10	−4.6
12	−5.0
7	−5.5
15	−5.7
9	−5.8
13	−5.9
18	−6.0
17	−6.2
14	−6.3
11	−6.6
16	−6.6
19	−7.2
22	−7.4
27	−7.5
23	−7.5
20	−7.6
26	−7.6
25	−7.6
21	−7.9
28	−8.5
24	−8.5
29	−9.2
30	−9.2

liter, the regression term would be $\log_{10} 60 \times 2.36 = 1.778 \times 2.36 = 4.2$. In Table 3, this has been multiplied by 10 to obtain the integer 42. Later, the result is divided by 10 and thus a precision of PI of one decimal is obtained. By using a pocket chart of this kind, it is possible to estimate PI by very simple algebra. For example, for the aforementioned subject, PI may be estimated simply as $(40 - 105)/10 = -6.5$ or the same as before.

Table 4 presents the PI according to Model 4 for each of the 30 subjects in Table 1. The PI values ranging from −1.7 to −9.2 are ranked according to decreasing PI, i.e., increasing predicted survival. The ranking is of course not the same as in Table 1, although the subjects with lower PI (Table 4) tend to have longer observed survival times (Table 1) and vice-versa.

Thus, the PI for a subject defines his/her place within the prognostic "spectrum" defined by the model. In the following, it will be shown how the information in PI may be utilized further in the estimation of a survival curve, the probability of surviving a given time and the median survival time for the subject.

SURVIVAL CURVE ESTIMATE FOR THE INDIVIDUAL SUBJECT

From a given Cox model, it is possible to estimate the cumulative survival probability or so-called survivorship function S(t, z) corresponding to any combination of the variables $z = z_1 \cdots z_p$. Discussing hazard previously, the equation $\Lambda(t, z) = -\log_e S(t,z)$ was presented. Solving for S(t, z) this becomes $S(t, z) = e^{-\Lambda(t, z)}$. Following from the assumption of proportionality of the Cox model, $\Lambda(t, z) = \Lambda_0(t)\, e^{PI}$. Since $\Lambda_0(t)$ is estimated in a Cox analysis for the whole range of time t from zero to the longest observation time, it is thus possible using these equations to estimate $\Lambda(t, z)$ and then S(t, z), which is the estimated survival curve corresponding to a given value of PI or combination of variables z (13, 17, 23–24).

Using Model 4 in Table 2, the estimated survival curve for each of the two subjects with given values of albumin and bilirubin is shown in Figure 6.

Figure 7 (top panel) shows survival curves corresponding to different values of PI. The middle panel of Figure 7 shows the cumulative hazard corresponding to the same values of PI. As a consequence of the assumption of proportionality of the Cox model, both the slope (the hazard) and the level increase with the same factor $e^1 = 2.718$, from one curve to the next above. In the bottom panel of Figure 7 showing the logarithm of the cumulative hazards, the curves are equidistant vertically with a distance of 1 ($\log_e 2.718 = 1$) between the curves. The shape of these curves is exactly the same; they are all based on the survival structure of the whole group of subjects (as are the curves in the upper and middle parts of Figure 7). The curves only differ in their level being determined by the value of PI. Accordingly, the logarithm of the cumulative underlying hazard $\log_e \hat{\Lambda}_0(t)$, which is equal to $\log_e \hat{\Lambda}(t)$ for a PI of zero, can be illustrated by a curve having exactly the same shape as these curves, but lying 4 units higher than the curve for a PI of −4.

Methods for deriving point-wise confidence limits for estimated survival curves are given in Ref. (13).

PROBABILITY OF SURVIVING A GIVEN TIME

For given values of $\hat{\Lambda}_0(t)$ corresponding to a given span of time, e.g. 1 year, it is possible to estimate the value of S(t, z) corresponding to various values of PI (13, 17, 23). The estimation is performed using the same equations as in the previous section, the only difference being that here $\hat{\Lambda}_0(t)$ is kept constant for each curve, whereas PI is allowed to vary. Figure 8 has been constructed in this way. The reverse sigmoid shape of the curves is a con-

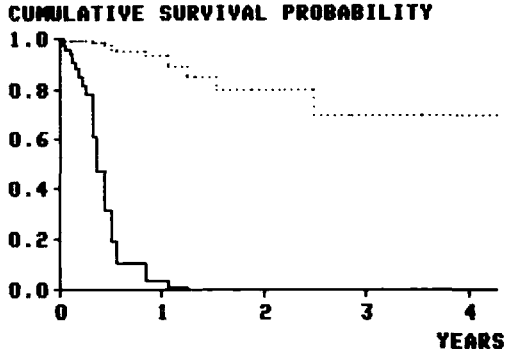

FIG. 6. Estimated survival curves based on Model 4 in Table 2 for a subject having an albumin of 29 gm per liter and a bilirubin of 220 μmoles per liter ($\log_{10}$ bilirubin = 2.342) (——) and a subject having an albumin of 33 gm per liter and a bilirubin of 20 μmoles per liter ($\log_{10}$ bilirubin = 1.3) (· · · ·).

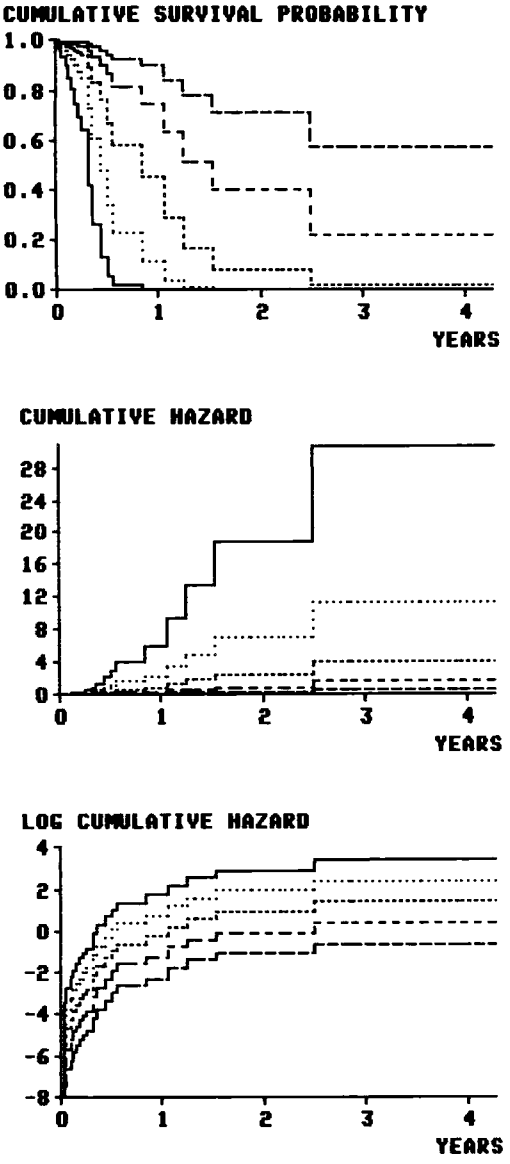

FIG. 7. Estimated survival curves (*top*), estimated cumulative hazards (*middle*) and logarithm of estimated cumulative hazards (*bottom*) based on Model 4 in Table 2 for different values of the PI: —— PI = −4; · · · · PI = −5; - - - - PI = −6, – – – – PI = −7; and – – PI = −8.

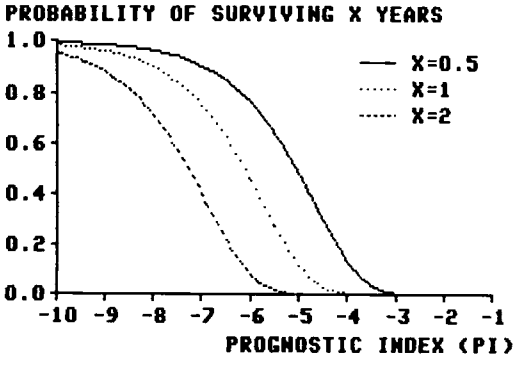

FIG. 8. Estimated probability of surviving 0.5, 1 and 2 years by PI based on Model 4 in Table 2.

sequence of the exponential functions involved. From a given value of PI on the abscissa, one can read the estimated probability of surviving 0.5, 1 and 2 years on the ordinate. For example, for a subject having a PI of −6, the probability of surviving 0.5, 1 and 2 years would be 0.76, 0.45 and 0.08, respectively.

ESTIMATED MEDIAN SURVIVAL TIME

Another useful way of interpreting PI is to estimate the median survival time for the subject. This is the span of time the subject will survive with 50% probability. For a subject having the variables z giving a certain PI, the median survival time is estimated as the time t for which the estimated survival curve $\hat{S}(t, z)$ reaches 0.5 (13, 17, 23). A graph showing the estimated median survival time as a function of PI for Model 4 in Table 2 is shown in Figure 9. Each point of the curve is obtained in the following way: for a value of PI, a survival curve is estimated as previously described. The time t where the curve reaches a cumulative survival probability of 0.5 is the estimated median survival time corresponding to that PI. This is repeated for the whole range of possible PI values to obtain the whole curve. Since these estimations have not yet been included in the available standard computer programs, they have to be done by the investigator.

From a given value of PI on the abscissa in Figure 9, one reads the corresponding value of the estimated median survival time on the ordinate. For example, for a subject having a PI of −6, the estimated median survival time would be 0.9 years. Since for PI values less than −7.8 the estimated survival curves do not reach 0.5, one can only say that the estimated median survival time will be longer than 4.3 years for such low values of PI.

There is a close correspondence between Figures 7 (upper part), 8 and 9. They illustrate the predictive information of the same Cox model in complementary ways.

THE COX MODEL IN CONTROLLED CLINICAL TRIALS

ADJUSTMENT FOR IMBALANCE

In controlled clinical trials, randomization is performed with the purpose of eliminating bias in treatment assignment. This will tend to lead to comparable treat-

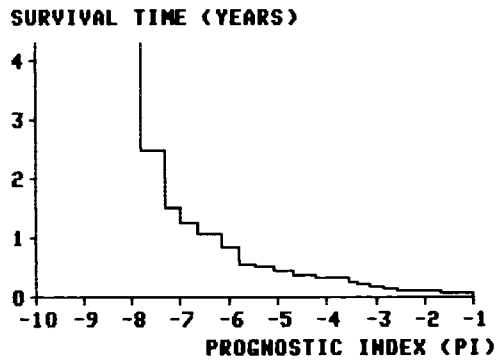

FIG. 9. Estimated median survival time by PI based on Model 4 in Table 2.

ment groups. However, random allocation does not guarantee complete balance. Imbalance in variables being known or suspected to covary with survival (prognostic variables) may occur randomly. If this happens, the "spontaneous" survival may be different between the treatment groups. Therefore, to insure a fair comparison between the studied treatments, imbalance in prognostic variables should be detected and adjusted for (25–29; Christensen, E. et al., *Gastroenterology* 1986; 90:508–509, Correspondence). This can be done by performing a Cox regression analysis in which the prognostic variables are included together with the treatment variable (13). Even slight imbalance may have a marked influence on the result if the prognostic variable is very important (13, 30). In such cases, there may be a substantial difference in the estimated therapeutic effect as obtained with and without adjustment for imbalance in prognostic variables (13, 30).

To illustrate these points, consider the treatment allocation data in Table 5 to supplement the data in Table 1. The Kaplan-Meier survival curves for the two treatment groups shown in Figure 10 are not significantly different, although the survival in the prednisone group tends to be longer than in the placebo group. Performing a Cox regression analysis, including the treatment variable together with albumin, $\log_{10}$ bilirubin and alcoholism, the model presented in Table 6 is obtained. In this model, the treatment as well as alcoholism are also

TABLE 5. Constructed treatment allocation data to supplement the data in Table 1

Subject no.	Treatment [prednisone (0)/placebo (1)]
1	1
2	1
3	0
4	1
5	0
6	1
7	1
8	0
9	1
10	0
11	1
12	1
13	0
14	1
15	1
16	0
17	1
18	1
19	0
20	1
21	0
22	0
23	1
24	0
25	1
26	0
27	0
28	0
29	0
30	0

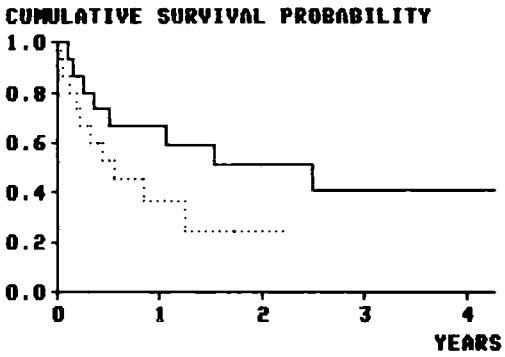

FIG. 10. Unadjusted survival curves for prednisone (——) and placebo (· · · ·)-treated subjects based on the treatment allocation data in Table 5 in combination with the survival data in Table 1. The difference in survival between the two groups is not significant.

significant. Figure 11 shows the estimated survival curves for the two treatments adjusted for the influence of albumin, $\log_{10}$ bilirubin and alcoholism. The difference in survival is more marked and statistically significant. The main reason for this effect is that alcoholism is slightly more frequent in the prednisone (53%) than in the placebo group (40%), this difference being adjusted for by the Cox model.

Although a powerful tool, statistical adjustment for the influence of prognostic variables cannot replace randomization. In general, only a smaller part of the variation in the hazard between the individuals can be "explained" by the variation in the predictor variables included in a Cox regression model. A number of unknown not yet identified variables may be associated with the hazard. Still, the only way in which such variables may be modeled as being distributed equally between the treatment groups is randomization (31). Since, as mentioned previously, randomization does not guarantee complete balance, it is important to continue the search for variables by which prognosis may be "explained" more precisely.

QUANTIFICATION OF THERAPEUTIC EFFECT IN THE INDIVIDUAL SUBJECT

In a Cox regression model, including the treatment variable together with the prognostic variables, it is possible to estimate for a given subject the PI for each of the treatments (13). Each of the PIs can be translated to an estimate of the median survival time as described. By subtracting the times for the treatments to be compared, the therapeutic effect can be expressed as the time (in months or years) by which survival may be prolonged (or shortened) by the one treatment compared to the other (13).

VARIABLES ASSOCIATED WITH THERAPEUTIC EFFECT

Just as almost any biological variable shows a spectrum of variation between individuals, the relative hazard under various treatments may show a similar variation, depending on the characteristics of the subject. To identify variables influencing the magnitude of the therapeutic effect, one may use a Cox model which allows for the

TABLE 6. Final Cox regression analysis of the data in Table 1 supplemented with treatment allocation data from Table 5

χ^2 model	d.f.	p model	R^2	Variables included	Regression coefficient b	SE(b)	N.D. [b/SE(b)]	p coefficient
46.59	4	<0.0001	0.37	Albumin	−0.34	0.11	−3.01	<0.0001
				Log_{10} bilirubin	3.61	1.24	2.91	<0.0001
				Alcoholism	1.86	0.78	2.39	0.005
				Treatment	2.18	0.78	2.78	0.0006

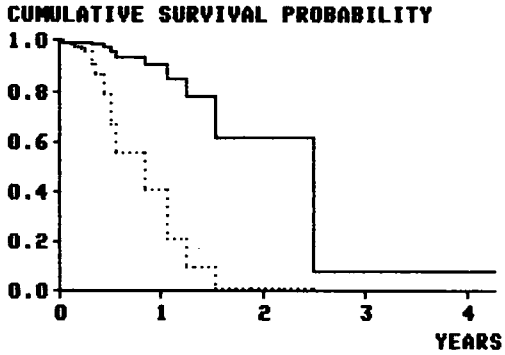

FIG. 11. Adjusted estimated survival curves for prednisone (——) and placebo (· · · ·)-treated subjects based on the Cox model in Table 6 for mean values of albumin, log_{10} bilirubin and absent alcoholism (p = 0.006).

treatment given, other predictor variables characterizing each patient and the interaction between the treatment variable and the other variables (13, 17, 32, 33).

One useful way to design such a model is in addition to the treatment variable (e.g., scored as 0 for Treatment A and 1 for Treatment B) to include two regression terms for each of the other predictor variables, one for each treatment (17, 23). For each pair of regression terms, the predictor variable should be scored by its usual scoring in the term corresponding to the treatment given to the subject and by zero in the other term. This leads to estimation of two regression coefficients b_A and b_B for each variable. If the difference between the two coefficients is significant by comparison with its standard error (estimated from the variance-covariance matrix of the coefficients), one may assume that the treatment effect depends on the value of the variable, which should be considered "therapeutic" and maintained in the model with one coefficient for each treatment (17, 23). Otherwise, one coefficient b_i common to the two treatments may be used instead.

The standard error of the difference $b_A - b_B$, i.e., $SE(b_A - b_B) = \sqrt{var(b_A) + var(b_B) - 2 \times covar(b_A b_B)}$, var meaning the variance and covar the covariance being obtained from the variance-covariance matrix of the b coefficients (which is different from the variance-covariance matrix of the predictor variables). If the ratio $|(b_A - b_B)|/SE(b_A - b_B)$ is higher than 1.96, b_A and b_B are considered significantly different at the 5% level (17, 23). If, for example, $b_A = 1.4$, $b_B = 0.6$, $var(b_A) = 0.7$, $var(b_B) = 0.4$ and $covar(b_A b_B) = 0.5$, then $|(b_A - b_B)|/SE(b_A - b_B) = 0.8/0.32 = 2.5$, and hence b_A and b_B are significantly different. These estimations are not included in the available standard computer program but need to be done by the investigator. The variance-covariance matrix of the b coefficients can be obtained from the computer printouts.

The significance of the interaction between a variable and the treatment may also be tested using the likelihood ratio test (described previously) comparing the model, including both b_A and b_B with the model having only one coefficient b for the variable in question.

For a model including one or more "therapeutic" variables, the PI may be estimated for each of the treatment alternatives as described previously. The difference $PI_{tr.A} - PI_{tr.B}$ may be considered an estimate of the therapeutic effect. As described in more detail in Ref. (23), it is possible to test if the difference (which may be considered as a *therapeutic index*) is significantly different from zero by comparing the difference with its standard error (estimated from the variance-covariance matrix of the coefficients) (23). In this way, it may be possible to identify patients having a significantly beneficial effect, a significantly harmful effect and no significant effect of the treatment given (23).

TIME-DEPENDENT COX REGRESSION MODEL

The preceding methods have utilized covarying variables at the beginning of follow-up, usually at the time of diagnosis or randomization in a controlled clinical trial. However, in a patient with cirrhosis, for example, the clinical situation may rapidly change for better or worse, e.g., prognosis may improve if the patient stops drinking alcohol (22, 34–36) or it may become worse if gastrointestinal bleeding occurs (4, 30). Estimates of prognosis seem to be improved if such changes occurring during the course of the disease are taken into account (22).

The Cox regression model for time-fixed variables previously described can be generalized to the case of time-dependent variables (6), where each variable z_i is no longer constant (equal to the value at the start of the study), but is allowed to vary as a function of time t after entry into the study: $z_i(t)$. Consequently, the hazard of a given patient is allowed to vary corresponding to the variation in time of the predictor variables. If, for example, a patient develops gastrointestinal bleeding, hazard is likely to increase; if the bleeding can be effectively treated, hazard is likely to decrease (22).

A PI based on the time-dependent model may be estimated repeatedly during the course of the disease to update estimates of the prognosis (22). It may be well-suited for close monitoring, e.g., before liver transplantation to insure optimal timing of the procedure (37). For further details on the time-dependent Cox model, the reader is referred to Ref. (22).

PERFORMING COX REGRESSION ANALYSES—HOW AND WHEN

The Cox regression model and other related methods for analysis of survival data have been reviewed by Lee (38) and Kalbfleisch and Prentice (7), who also present computer programs in FORTRAN to perform time-fixed

and time-dependent Cox regression analyses. A number of commercial standard programs for Cox regression analysis are available, for example BMDP 2L (39) and PROC PHGLM (16). However, because of the complexity of the Cox regression analyses and the necessity of carefully checking model assumptions, the analyses should preferably be performed with the close cooperation of a qualified statistician.

The power of a Cox analysis depends largely on the number of events or endpoints in the data set (7). To reduce the risk of finding spurious associations, one should limit the number of predictor variables to be investigated. Harrell (16) recommends that the number of predictor variables examined should not be more than about 5 to 10% of the number of endpoints. Thus, for a data set as that presented in this paper, one should only examine 1 to 2 predictor variables to insure that the results would be reasonably confident. Therefore, the analyses presented in this paper should be taken for nothing more than illustrations of the "dynamics" of Cox analyses.

It must be realized that the variables finally included in a Cox regression model is dependent on the way and order of selection (e.g., forward selection or backward elimination), and the variables which have actually been recorded and thus are available for analysis. This may explain why models on comparable groups of patients may differ in regard to which variables have been found significant and which have not. Such differences are only to be expected. Thus, the regression model finally obtained is not unique. In many instances, slightly different models might have been obtained with nearly the same degree of prognostic information.

In multivariate analyses, including Cox regression analyses, some variables may randomly be found to have a significant prognostic or therapeutic association (40). This problem will increase with the number of variables analyzed and hence the number of statistical tests performed. Therefore, it is important to evaluate the results in the light of common clinical knowledge and biologic principles to see if the results are "reasonable."

The usefulness of the model obtained should be judged from its predictive power and the ease with which the variables included can be obtained. The latter will highly influence its applicability in clinical practice.

The Cox model should be used only if the number of subjects having an endpoint is sufficient in relation to the number of predictor variables being planned to analyze and if the assumption of proportional hazards can be fulfilled. If the latter is not the case, one should not use the Cox model but simpler, less restrictive methods based on stratification (9), although such methods can accommodate only a few predictor variables at a time. When reporting Cox analyses, one should describe carefully how model assumptions were checked. Reports omitting these details should be considered with skepticism.

VALIDITY OF RESULTS OBTAINED BY COX REGRESSION ANALYSIS

Like other powerful multivariate statistical analyses, the Cox regression analyses should be considered explorative or heuristic. Usually, the results obtained need to be validated before they can be considered "proved" (41). The best way of validating the results is to demonstrate that the obtained statistical model can predict prognosis correctly in independent subjects (13). If an independent group of subjects is not available, a validation in a more limited sense may be performed using a split-sample testing technique (13, 17, 22, 42). With this method, the obtained statistical model is estimated using one portion of the subjects. With this model, the survival for the remaining subjects is predicted. Then, the predicted survival is compared with the survival actually observed for that portion to see if prediction is satisfactory or not (17, 22). However, the ultimate test is the correct prediction of prognosis in new subjects.

CONCLUSIONS

The Cox regression model is a powerful statistical tool for analysis of censored survival data. With this model, new information on variables associated with prognosis and therapeutic effect in chronic liver disease has been obtained (13, 17, 22, 23, 42, 43). These results allow a more precise estimation of the therapy-dependent prognosis in the individual patient. Thus, a more individual treatment strategy based on the characteristics of the patients becomes possible.

The Cox model is complex and may be difficult to understand, but this should not lead to its abandonment and replacement by univariate analyses which disregard the pattern of covariation of other variables with prognosis and therapeutic effect. Instead, every effort should be taken to "translate" the results of analyses into simpler forms, such as "pocket charts" and diagrams to enable easy estimation of the therapy-dependent prognosis at the bedside in new patients.

To promote understanding and an increased use of the Cox regression model, a closer cooperation between doctors and statisticians is necessary. This will also stimulate the development of new statistical tools with a high degree of utility in clinical practice.

Acknowledgments: I am grateful to my coworkers in many previous studies in which some of the ideas presented here were developed; in particular, Per Kragh Andersen and Douglas G. Altman, and to the reviewers of this paper.

REFERENCES

1. Starmer CF, Lee KL, Harrell FE, et al. On the complexity of investigating chronic illness. Biometrics 1980; 36:333–335.
2. Christensen E, Schlichting P, Fauerholdt L, et al. Changes of laboratory variables with time in cirrhosis: prognostic and therapeutic significance. Hepatology 1985; 5:843–853.
3. Christensen E, Crowe J, Doniach D, et al. Clinical pattern and course of disease in primary biliary cirrhosis based on an analysis of 236 patients. Gastroenterology 1980; 78:236–246.
4. Christensen E, Fauerholdt L, Schlichting P, et al. Aspects of the natural history of gastrointestinal bleeding in cirrhosis and the effect of prednisone. Gastroenterology 1981; 81:944–952.
5. Armitage P. Statistical methods in medical research. Oxford, England: Blackwell Scientific Publications, 1971.
6. Cox DR. Regression models and life tables (with discussion). J R Statist Soc B 1972; 34:187–220.

7. Kalbfleisch JD, Prentice RL. The statistical analysis of failure time data. New York: John Wiley, 1980.
8. Kaplan E, Meier P. Nonparametric estimation from incomplete observations. J Amer Statist Assoc 1958; 53:457–481.
9. Peto R, Pike MC, Armitage P, et al. Design and analysis of randomized clinical trial requiring prolonged observation of each patient. II. Analysis and examples. Br J Cancer 1977; 35:1–39.
10. Armitage P, Gehan EA. Statistical methods for the identification and use of prognostic factors. Int J Cancer 1974; 13:16–36.
11. Christensen E, Schlichting P, Fauerholdt L, et al. Prognostic value of Child-Turcotte criteria in medically treated cirrhosis. Hepatology 1984; 4:430–435.
12. Breslow NE. Covariance analysis of censored survival data. Biometrics 1974; 30:89–99.
13. Christensen E, Neuberger J, Crowe J, et al. Beneficial effect of azathioprine and prediction of prognosis in primary biliary cirrhosis: final results of an international trial. Gastroenterology 1985; 89:1084–1091.
14. Draper NR, Smith H. Applied regression analysis. New York: John Wiley, 1981.
15. Atkinson AC. A note on the generalized information criterion for choice of a model. Biometrika 1980; 67:413–418.
16. Harrell FE. The PHGLM procedure. In: SAS supplemental user's guide. Cary, North Carolina: SAS Institute, 1983: 267–294.
17. Schlichting P, Christensen E, Andersen PK, et al. Prognostic factors in cirrhosis identified by Cox's regression model. Hepatology 1983; 3:889–895.
18. Elashoff JD. Surviving proportional hazards. Hepatology 1983; 3:1031–1035.
19. Kay R. Proportional hazards regression models and the analysis of censored survival data. Appl Statist 1977; 26:227–237.
20. Andersen PK. Testing goodness-of-fit of Cox's regression and life model. Biometrics 1982; 38:67–77.
21. Kay R. Goodness-of-fit methods for the proportional hazards regression model: a review. Revue de Epidemiologie et de Santé Publique 1984; 32:185–198.
22. Christensen E, Schlichting P, Andersen PK, et al. Updating prognosis and therapeutic effect evaluation in cirrhosis using Cox's multiple regression model for time dependent variables. Scand J Gastroenterol 1986; 21:163–174.
23. Christensen E, Schlichting P, Andersen PK, et al. A therapeutic index that predicts the individual effects of prednisone in patients with cirrhosis. Gastroenterology 1985; 88:156–165.
24. Andersen PK, Christensen E, Fauerholdt L, et al. Measuring prognosis using the proportional hazards model. Scand J Statist 1983; 10:49–52.
25. Brown BW. Designing for cancer clinical trials: selection of prognostic factors. Cancer Treat Rept 1980; 64:499–502.
26. Lachin JM. Statistical analysis of the randomized clinical trial. In: Tygstrup N, Lachin JM, Juhl E, eds. The randomized clinical trial and therapeutic decisions. New York: Marcel Dekker, 1982: 155–197.
27. Lewis JA. Clinical trials: statistical developments of practical benefit to the pharmaceutical industry (with discussion) J R Statist Soc A 1983; 146:362–393.
28. Altman DG. A fair trial? Br Med J 1984; 298:336.
29. Altman DG. Compatibility of randomised groups. Statistician 1985; 34:125–136.
30. The Copenhagen Esophageal Varices Sclerotherapy Project. Sclerotherapy after first variceal haemorrhage in cirrhosis. A randomized multicenter trial. N Engl J Med 1984; 311:1594–1600.
31. Byar DP. Why data bases should not replace randomized clinical trials. Biometrics 1980; 36:337–342.
32. Byar DP, Corle DK. Selecting optimal treatment in clinical trials using covariate information. J Chron Dis 1977; 30:445–459.
33. Byar DP, Green SB. The choice of treatment for cancer patients based on covariate information: application to prostate cancer. Bull Cancer (Paris) 1980; 67:477–490.
34. Tygstrup N, Juhl E, Copenhagen Study Group for Liver Diseases. The treatment of alcoholic cirrhosis: the effect of continued drinking and prednisone on survival. In: Gerok W, Sickinger, K, HH Hennekeuser, eds.) Alcohol and the liver. Stuttgart, West Germany: Schafttaur, 1971: 519–526.
35. Borowsky SA, Strome S, Lott E. Continued heavy drinking and survival in alcoholic cirrhotics. Gastroenterology 1981; 80:1405–1409.
36. Schenker S. Alcoholic liver disease: evaluation of natural history and prognostic factors. Hepatology 1984; 4:36S–43S.
37. Van Thiel DH, Schade RR, Gavaler JS, et al. Medical aspects of liver transplantation. Hepatology 1984; 4:79S–83S.
38. Lee ET. Statistical models for survival data analysis. Belmont, California: Lifetime Learning Publications, 1980.
39. Dixon WJ. BMDP statistical software. Los Angeles, California: University of California Press, 1981.
40. Lee KL, McNeer JF, Starmer CF, et al. Clinical judgment and statistics—Lessons from a simulated randomized trial in coronary artery disease. Circulation 1980; 61:508–515.
41. Peduzzi PN, Detre KM, Chan YK, et al. Validation of a risk function to predict mortality in a VA population with coronary artery disease. Controlled Clinical Trials 1982; 3:47–60.
42. Gines P, Quintero E, Arroyo V, et al. Compensated cirrhosis: natural history and prognostic factors. Hepatology, 1987; 77:122–128.
43. Copenhagen Study Group for Liver Diseases. Testosterone treatment of men with alcoholic cirrhosis: a double-blind study. Hepatology 1986; 6:807–813.