STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:2807–2821
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2143

Long-term survival with non-proportional hazards: results from the Dutch Gastric Cancer Trial

H. Putter[1,*,†], M. Sasako[2], H. H. Hartgrink[3], C. J. H. van de Velde[3] and J. C. van Houwelingen[1]

[1]*Department of Medical Statistics and Bioinformatics, Leiden University Medical Centre, University of Leiden, The Netherlands*
[2]*Department of Surgery, National Cancer Centre Hospital, Tokyo, Japan*
[3]*Department of Surgery, Leiden University Medical Centre, University of Leiden, The Netherlands*

SUMMARY

Randomized clinical trials with long-term survival data comparing two treatments often show Kaplan–Meier plots with crossing survival curves. Such behaviour implies a violation of the proportional hazards assumption for treatment. The Cox proportional hazards regression model with treatment as a fixed effect can therefore not be used to assess the influence of treatment of survival. In this paper we analyse long-term follow-up data from the Dutch Gastric Cancer Trial, a randomized study comparing limited (D1) lymph node dissection with extended (D2) lymph node dissection. We illustrate a number of ways of dealing with survival data that do not obey the proportional hazards assumption, each of which can be easily implemented in standard statistical packages. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: long-term survival; non-proportional hazards; time-dependent covariate effects

1. INTRODUCTION

Many randomized clinical trials in oncology concern long-term survival data, comparing an experimental treatment with a standard treatment or control. To test for equality of the survival rates of the treatments, the log-rank test is used [1]. Often in these trials, characteristics of the patient and of the tumour that are known before treatment are also recorded. The Cox proportional hazards regression model is the most popular choice to study the effect of those prognostic factors on survival [2]. One of the assumptions underlying the Cox regression model is the assumption of proportional hazards, meaning that the ratio of the hazard rates for different levels of the prognostic factor or for treatment versus control is constant over

*Correspondence to: Hein Putter, Department of Medical Statistics and Bioinformatics, Leiden University Medical Centre, University of Leiden, P.O. Box 9604, Leiden, 2300 RC, The Netherlands.
†E-mail: h.putter@lumc.nl

Received 3 June 2004
Accepted 9 December 2004

time. This ratio can then be expressed as a single number, the hazard ratio or relative risk of treatment over control, or of one level of a prognostic factor over another. Although not as implicitly assumed as in the Cox regression model, the validity of the log-rank test is also sensitive to the assumption that the hazard rates for different levels of the prognostic factor or for treatment versus control do not change appreciably over time [3]. When studying survival data over a short period of time, the proportional hazards assumption is often a reasonable one. In clinical trials with long-term follow-up however it often happens that the Kaplan–Meier survival curves cross. In the beginning of the study for instance, the experimental treatment may yield better survival, but this effect may be reversed after some time. Under the proportional hazards assumption, crossing of the survival curves is impossible. Thus, in a study where the patient groups do not differ between the treatments, a crossing of the survival curves implies a violation of the proportional hazards assumption. In such a case, the log-rank test for the difference in survival rates between the treatments will most likely not be significant, because of the contrasting early and late effects of the treatments. If the proportional hazards assumption fails to hold for the treatment or for one or more of the covariates, the results of a multivariate Cox regression will be misleading. The hazard ratio for the particular covariate cannot be interpreted as a hazard ratio or relative risk; moreover, the regression coefficients of the other covariates (including the treatment) may be biased as a result [4].

The aim of this paper is to illustrate a number of ways of dealing with long-term follow-up data which do not obey the proportional hazards assumption. We shall do this using the Dutch Gastric Cancer Trial as an example. Section 2 will discuss the background of this trial and describe the most important characteristics of the patients involved. Section 3 will focus on the effect of treatment without considering any covariates; the issue of how to deal covariates is studied in Section 4. We conclude in Section 5 with a discussion of the results and the methods and of possible other approaches.

2. THE DUTCH GASTRIC CANCER TRIAL

The data reported here are from the Dutch Gastric Cancer Trial, a prospective, randomized clinical trial, comparing limited (D1) lymph node dissection with extended (D2) lymph node dissection in patients with gastric cancer.

Eligibility criteria were histologically confirmed adenocarcinoma of the stomach without evidence of distant metastasis, and age below 85 years. Exclusion criteria were coexisting cancer or previous gastrectomy for benign tumours. Results of the trial have been presented elsewhere [5]. We limit our analysis to the 711 eligible patients who underwent a curative dissection, 380 of which had a D1-dissection and 331 a D2-dissection. The median follow-up of the patients in the analysis is 9.1 years. The baseline characteristics of the patients are summarized in Table I. Prognostic variables studied in this analysis are gender, lymph node involvement, age, residual tumour, resection type, tumour location and T-stage. Although T-stage was originally recorded in 5 categories, T-stage 0 up to 4, in view of the small sample sizes for T-stages 0 and 4, we have recoded T-stage 0 into T-stage 1 and T-stage 4 into T-stage 3. Patient groups were comparable between D1- and D2-dissection, with the exception of resection type.

Table I. Patient characteristics.

	D1 ($n=380$)	D2 ($n=331$)	p-value
Gender			0.92
Male	215 (57)	186 (56)	
Female	165 (43)	246 (44)	
Age			0.30
$\leqslant 65$ years	178 (47)	168 (51)	
>65 years	202 (53)	163 (49)	
T-stage			0.83
T1	100 (26)	88 (27)	
T2	181 (48)	152 (46)	
T3	97 (26)	91 (27)	
Unknown	2	0	
Type of resection			0.03
Total	115 (30)	126 (38)	
Partial	265 (70)	205 (62)	
Tumour location			0.99
CMA	37 (10)	32 (10)	
C	40 (10)	34 (10)	
M or A	303 (80)	265 (80)	
Residual tumour			0.71
R0	340 (89)	299 (90)	
R1	40 (11)	32 (10)	
Lymph node involvement			0.70
Negative	171 (45)	144 (44)	
Positive	209 (55)	187 (56)	

An earlier report of this trial showed a higher postoperative mortality of D2-dissection compared to D1-dissection [6]. The Kaplan–Meier survival curves of D1- and D2-dissection crossed after about 4 years. No significant difference in survival was found between the D1-group and the D2-group with a median follow-up of 6 years. However, this is most probably due to the initially unfavourable and subsequently advantageous effect of D2 over D1 cancelling out. Results of a (multivariate) Cox regression were presented in this paper as well as in a subsequent paper describing the long-term results of the trial [7], although the violation of the proportional hazards assumption was noted.

3. TREATMENT EFFECT IN NON-PROPORTIONAL HAZARDS

The primary endpoint is overall survival in months, starting from the day of surgery. Figure 1 shows a Kaplan–Meier plot of the survival curves for each of the treatment groups. It shows an initial survival advantage of D1 over D2 in the first 53 months, followed by a disadvantage. The p-value of the log-rank test is 0.71, suggesting no difference between the D1- and D2-groups.

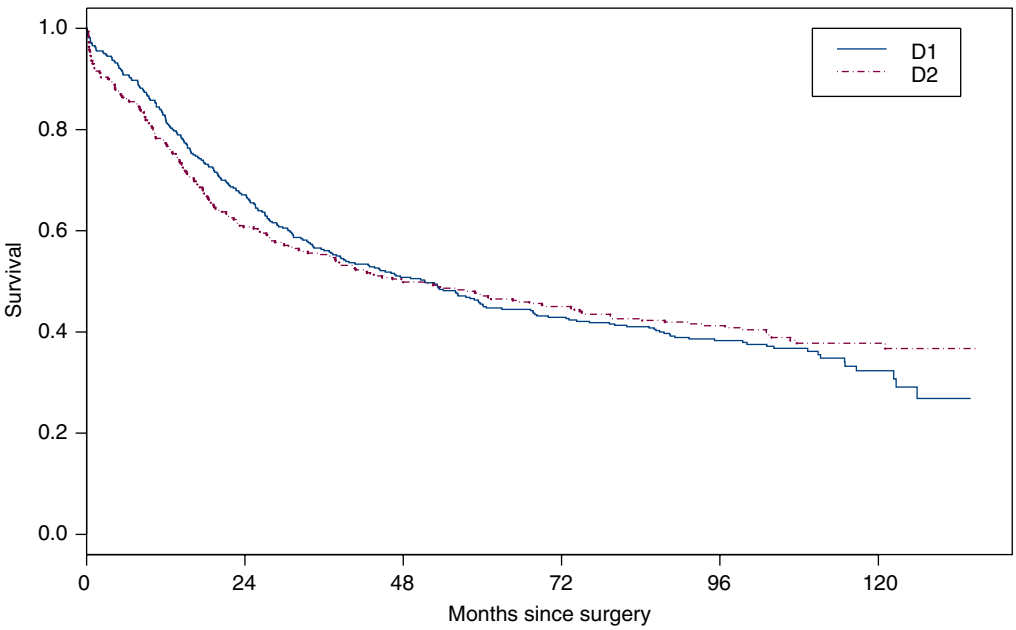

Figure 1. Kaplan–Meier plots of the survival curves for D1- and D2-dissection. The survival curves cross after 53 months.

The Cox regression with only randomization as a time-fixed effect gives an estimated hazard ratio of 0.97 of D2 dissection compared to D1-dissection, with a p-value of 0.73. The survival curves resulting from this univariate Cox regression are depicted in Figure 2. The higher post-operative mortality in the D2 group is not visible from this plot, nor is the crossing of the survival curves, so clearly Figure 2 does not give a realistic picture of the effect of treatment.

One way of studying how the effect of treatment changes over time is by using the life-table method. This method was used by epidemiologists long before the Cox regression model became popular. Divide time into a number of disjoint intervals $I_1, \ldots, I_m$. The hazard h_k of dying in interval I_k is then given by the number of deaths in that interval (d_k) divided by the number of person years in that interval (y_k). The number of person years is the sum over all patients still alive at the beginning of the interval (at risk) of the number of years alive during that interval. The standard error of h_k, based on a Poisson approximation, is $\sqrt{d_k}/y_k$. If h_{k1} and h_{k2} denote the estimated hazards at I_k for D1 and D2, respectively, and d_{k1} and d_{k2} the number of deaths at I_k for D1 and D2, respectively, then the delta-method implies that

$$\widehat{\text{se}}^2 \log\left(\frac{h_{k1}}{h_{k2}}\right) \approx \frac{\widehat{\text{se}}^2(h_{k1})}{h_{k1}^2} + \frac{\widehat{\text{se}}^2(h_{k2})}{h_{k2}^2} = \frac{1}{d_{k1}} + \frac{1}{d_{k2}}$$

The left plot of Figure 3 shows the estimated hazards on a yearly basis using the life-table method for each of the treatment groups separately. The plot on the right shows the resulting hazard ratio and associated error bars. The initial advantage and subsequent disadvantage of

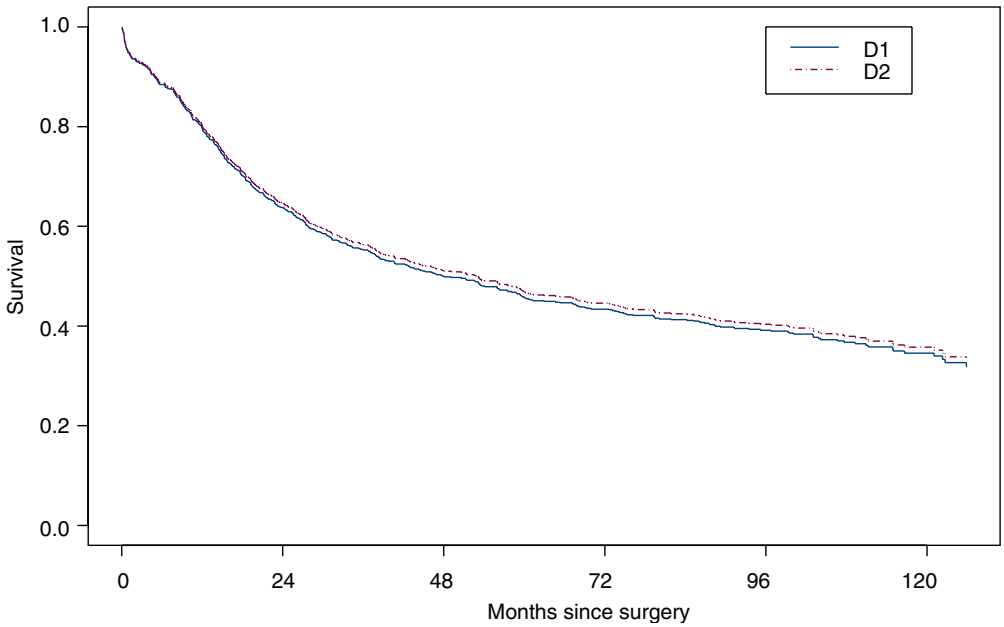

Figure 2. Cox regression plots of the survival curves for D1- and D2-dissection with time-fixed treatment effect.

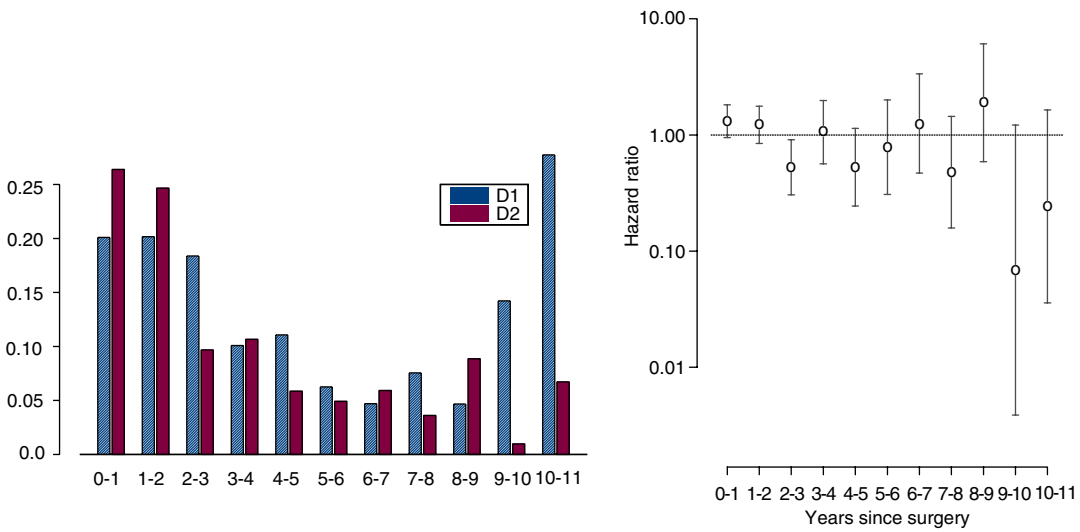

Figure 3. Yearly hazard estimates estimated by the life-table method, and the associated hazard ratios with 95 per cent-confidence intervals.

Table II. Results from Cox regression including treatment and time by treatment interaction. The correlation between the estimated coefficients was −0.93.

	Coefficient	Standard error	*p*-value
Treatment	0.760	0.261	0.0036
Time ∗ treatment	−0.273	0.084	0.0011

D1- compared to D2-dissection can be seen from this plot. Note also that the confidence intervals tend to become wider with time due to the decreasing number of events.

The evolvement of the hazard ratios over time can be modelled in a smooth way using standard statistical software by adding a time-dependent covariate in a Cox proportional hazards model. The most straightforward way to do this is by adding interaction terms of the treatment group and the prognostic variables with $f(t)$, where f is a given function of time, and t is time since surgery (in months in this case). A popular choice is $\log(t)$; we prefer to take $\log(t+1)$ and add an extra month to diminish the influence of very early events on the time-dependent variables. In the results we refer to this as 'time'. A significant effect of that interaction denotes the presence of a time-dependent effect and thus a violation of the proportional hazards assumption. The results of this time-dependent Cox regression applied to the data of the Dutch Gastric Cancer Trial are shown in Table II.

The time by treatment interaction now gives a hazard ratio of D2 over D1 which will vary over time, following the function $\exp(a + b\log(t+1))$, with a and b the estimated coefficients of treatment and treatment by time interaction in Table II. At time $t=0$, the hazard ratio equals $\exp(a) = \exp(0.760) = 2.14$ (the fact that the hazard ratio at $t=0$ depends only on a is an additional advantage of taking $f(t) = \log(t+1)$ as our function of time), after 5 years ($t=60$) it has decreased to $\exp(0.760 - 0.273 * \log(61)) = 0.69$. The log of the time-dependent hazard ratio (HR) of D2 over D1 is $a + b\log(t+1)$ and its standard error se(log(HR)) is the square root of $\mathrm{var}(a) + \mathrm{var}(b) * \log^2(t+1) + 2\,\mathrm{cov}(a,b) * \log(t+1)$. A pointwise 95 per cent-confidence interval can thus be obtained by taking the exponent of $\log(\mathrm{HR}) \pm 1.96 * \mathrm{se}(\log(\mathrm{HR}))$. Figure 4 shows the resulting hazard ratio of D2 over D1 and the confidence interval for each time-point. A hazard ratio of one indicates equality of the hazard rates of D1 and D2. The hazard ratio of D2 over D1 decreases over time; it is significantly higher than one until 5 months after surgery, and significantly lower than one after 35 months.

A number of comments are in order. The previous analysis, although relatively easy to perform with standard statistical software, has the disadvantage that one has to choose a form of how the effect of treatment will change over time. Here we chose $\log(t+1)$, the logarithm of time in months (plus 1). Other choices such as t, $\sqrt{t}$ are also possible, but such a choice is always arbitrary to some degree. A picture of the hazard ratio over time as in Figure 4 will always have the form that was imposed upon it, not necessarily the true effect evolving over time. Detecting the presence of a time-dependent effect may depend on the choice of $f(t)$, even the direction that this effect will take [8]. A plot like Figure 3 or the function `cox.zph()` in S-plus (see Chapter 6 in Reference [9]) can guide in choosing an appropriate function.

The evolvement over time of the hazard ratio is an important summary of a time-dependent treatment effect, but does not tell the full story. The baseline hazard function is crucial for determining the effect of the hazard ratio on the two survival curves. Before showing why, let us discuss estimation of this baseline hazard function.

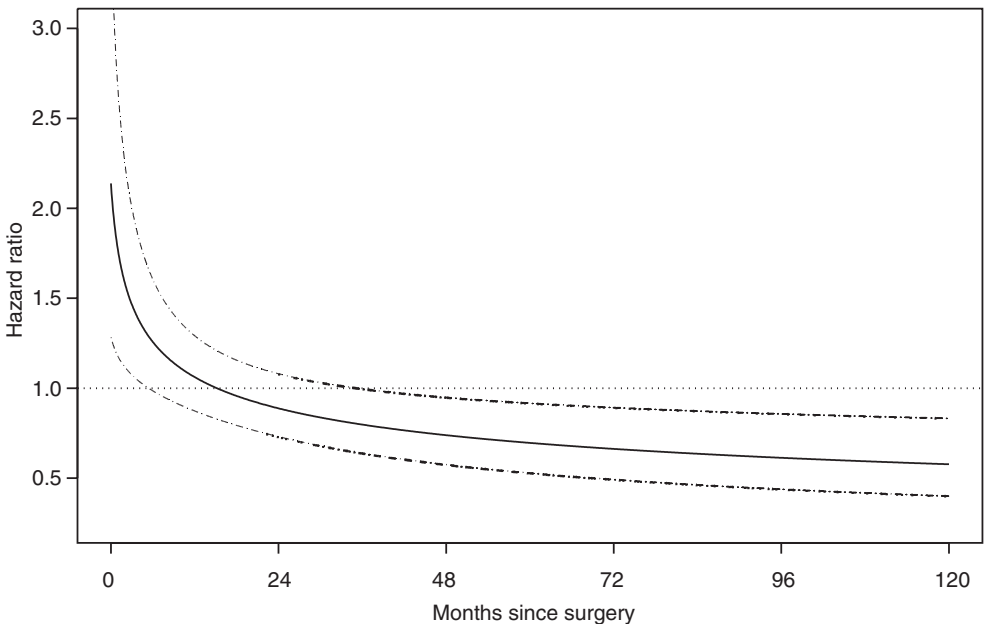

Figure 4. The estimated hazard ratio with 95 per cent confidence intervals based on Cox regression with treatment as time-dependent effect. A hazard ratio of one indicates equality of the hazard rates of D1 and D2.

Standard statistical packages like SPSS, SAS and S-plus are able to perform Cox regression with time-dependent covariates (although for S-plus and R the original data needs to be expanded), but most of them do not return the baseline hazard functions automatically in the presence of time-dependent covariates. The survival library in S-plus and R contains a function `basehaz()` to obtain an estimate of the baseline hazard. To show how this is done, we focus on the situation of a single covariate Z given by two values, 0 and 1. The time-dependent treatment effect is modelled by a function $f(t)$. The Cox proportional hazards model states that the hazard rate of an individual with covariate Z is given by

$$h(t) = h_0(t) \exp(\beta_F Z + \beta_T Z f(t)) \tag{1}$$

where β_F and β_T denote the fixed and time-dependent regression coefficients, respectively. Here h_0 is the baseline hazard corresponding to $Z = 0$, and if we denote the hazard function corresponding to $Z = 1$ by h_1, then this means that $h_1(t) = h_0(t) \exp(\beta_F + \beta_T f(t))$ and $\exp(\beta_F + \beta_T f(t))$ is the hazard ratio varying over time. The regression coefficients are estimated by an extension of the well known partial likelihood (see e.g. Section 9.2 of Klein and Moeschberger [3]). With estimated regression coefficients $\hat{\beta}_F$ and $\hat{\beta}_T$ obtained in this way, the baseline cumulative hazard rate $H_0(t)$ is estimated by Breslow's estimator, given by

$$\hat{H}_0(t) = \sum_{t_i \leqslant t, t_i \in D} \frac{1}{\sum_{j \in R(t_i)} \exp(\hat{\beta}_F Z_j + \hat{\beta}_T Z_j f(t_j))} \tag{2}$$

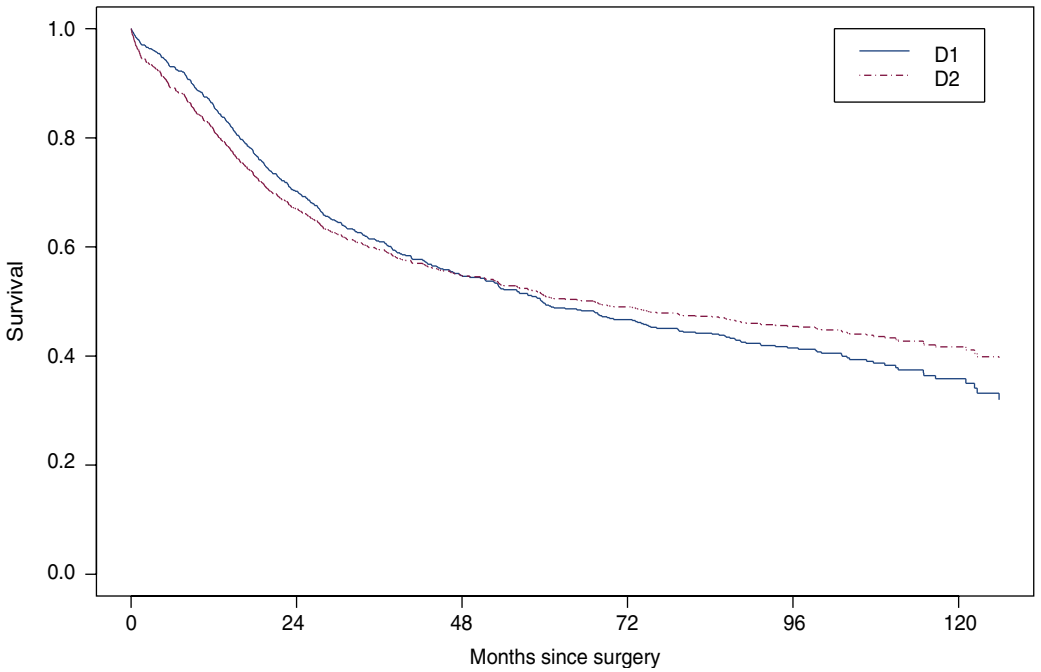

Figure 5. Cox regression plots of the survival curves for D1- and D2-dissection with time-dependent treatment effect. The survival curves cross after 47.6 months, with survival probability 0.548.

where D is the set of all event time-points and $R(t_i)$ is the risk-set at time t_i, consisting of all patients still alive at t_i. In the case of two treatments, the other hazard function is estimated by

$$\hat{H}_1(t) = \sum_{t_i \leqslant t, t_i \in D} \frac{\exp(\hat{\beta}_F + \hat{\beta}_T f(t_i))}{\sum_{j \in R(t_i)} \exp(\hat{\beta}_F Z_j + \hat{\beta}_T Z_j f(t_j))}$$

and the cumulative survival functions are given by $\hat{S}_j(t) = \exp(-\hat{H}_j(t))$, for $j = 0, 1$.

Figure 5 shows the survival curves that result from this procedure for the data of the Dutch Gastric Cancer Trial. The survival curves are in reasonable agreement with the Kaplan–Meier plot.

Why does the baseline hazard function influence the effect of the hazard ratio on the two survival curves in the presence of time-dependent covariates? Essentially, this is caused by the fact that the hazard at any time-point is a conditional probability of an event, given that it has not occurred yet. Thus, it concerns only subjects that are still at risk at that time-point. Consider the Dutch Gastric Cancer Trial, where the hazard ratio of D2 relative to D1 is initially larger than one, then below one. If the baseline risk is very high initially, most of the patients will have died before the hazard ratio becomes less than one and the effect of the subsequent lower hazard rate for the D2 group will be negligible. Conversely, if the baseline risk is very low, most of the patients will die after the hazard rates have crossed, so the

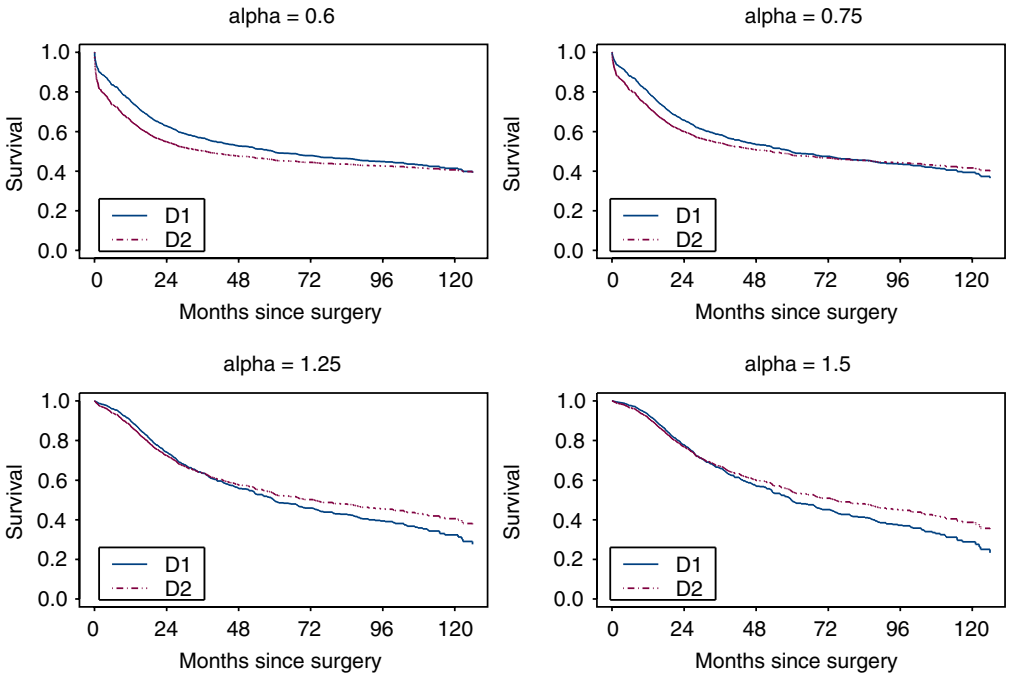

Figure 6. Cox regression plots for the Dutch Gastric Cancer Trial, using the estimated coefficients of Table II. The estimated baseline hazard of Figure 5 has been modified as in equation (3) with shape parameters 0.6, 0.75, 1.25 and 1.5, respectively.

treatment effect will be dominated by the higher hazard rate of D1 later in time. To illustrate this point, we modified the baseline cumulative hazard estimator of the D1-group $\hat{H}_0(t)$ of equation (2) by adjusting a shape parameter similar to that of the Weibull distribution, while ensuring that the baseline survival probability at 60 months remained identically equal to 0.497. More specifically, with $H_0 = \hat{H}_0(60) = 0.699$ the baseline cumulative hazard estimate at 60 months, for a given shape parameter $\alpha > 0$, we set

$$\tilde{H}_0(t) = H_0 \left(\frac{\hat{H}_0(t)}{H_0} \right)^{\alpha} \tag{3}$$

Clearly, $\alpha = 1$ returns $\hat{H}_0(t)$; for $\alpha < 1$, this has the effect of making the modified baseline cumulative hazard increase more rapidly for $t < 60$ and less rapidly for $t > 60$. For $\alpha > 1$, the effect is reversed. The estimates from Table II were then used to construct the survival curve of the D2-group using equation (1). The result is shown in Figure 6, for $\alpha = 0.6, 0.75, 1, 25, 1.5$. The survival curves cross at $t = 126, 85, 36, 29$ with survival probabilities $0.394, 0.453, 0.637$, 0.717, respectively. With the same regression coefficients of treatment and treatment by time interaction, the first plot would clearly favour D1-dissection, while the last plot would favour D2-dissection.

4. PROGNOSTIC FACTORS AND NON-PROPORTIONAL HAZARDS

The Cox regression presented in Section 3 incorporated a time-dependent effect of treatment. Prognostic factors can be easily added to this regression, as extra covariates in a multivariate time-dependent Cox regression.

If the effect of such a covariate does not vary over time and is identical for the different treatments, then adding this covariate in a multivariate Cox regression may affect the baseline hazard and the estimates of treatment and treatment by time interaction. After having obtained the two survival curves for D1- and D2-dissection for a reference value of that covariate, the survival curves for D1- and D2-dissection for a different value of that covariate will be similar to the survival curves for D1 and D2 for the reference value, but shifted upwards or downwards in a proportional way, depending on the direction of the effect. With a proportional shift we mean a shift as is seen in the familiar context of time-fixed proportional hazards regression, i.e. the survival curves for D1 (and D2) for different values of the covariate will be parallel on the log(-log)-probability scale.

But things start to become more complicated when the effect of a covariate may also depend on time or on treatment or on both time and treatment. When the effect of a covariate itself is time-dependent, the baseline hazard function and the estimates of treatment and treatment by time interaction may again change. In this case, however, the survival curves for D1 (D2) for different values of the covariate will not change in a proportional way; the shift may be negligible early in time and quite large for larger values of t, or the other way around, depending on the direction of the time-dependent effect of the covariate.

When the effect of a covariate depends on treatment, the situation is again different. Then both survival curves are shifted upwards or downwards, each in a proportional way. The size of the shifts however will differ for the different treatments. As a result, the survival curves may cross at different points in time, for different values of the covariate. The most complicated situation occurs when the effect of the covariate changes both with time and treatment; then any combination of the individual effects on the survival curves of covariate by time and covariate by treatment interaction is conceivable.

To conduct a multivariate analysis, we took the practical approach of first studying time-dependent effects for each covariate separately. The effect of each covariate was first evaluated univariately, then in a bivariate Cox regression with the covariate and its interaction with time, again taking $f(t) = \log(t+1)$, with t indicating time since surgery in months. Treatment was not included at this stage. The result of this analysis is shown in Table III.

Each of the covariates, with the exception of gender, has a significant effect on survival. Time-dependent effects were only observed for residual tumour ($p = 0.023$) and resection type ($p = 0.025$).

We proceeded with the multivariate Cox regression analysis in three steps. First we used a forward selection procedure in which each of the covariates and their interaction with time was included. A covariate by time interaction was only considered significant if the likelihood-ratio test for the model with both covariate and covariate by time indicated a significantly better fit compared to the model without. In the second step, treatment and interaction of treatment with time was entered. In the last step, all interactions of covariates with treatment and of covariates with treatment and time were entered in a forward selection procedure. A covariate by treatment interaction was only considered significant if the likelihood-ratio test for the model with both covariate and covariate by treatment indicated a significantly better

Table III. Results from uni- and bivariate Cox regressions including covariate and covariate by time effects.

	Covariate effect			Time effect			Overall test
Variable	Coefficient	SE	p-value	Coefficient	SE	p-value	p-value
Gender							0.31
Female	−0.148	0.096	0.12	0.005	0.083	0.95	
Lymph node involvement							<0.001
Positive	1.100	0.104	<0.001	0.051	0.091	0.57	
Age							<0.001
>65 years	0.549	0.096	<0.001	−0.069	0.085	0.41	
Residual tumour							<0.001
R1	1.266	0.135	<0.001	0.273	0.120	0.023	
T-stage			<0.001			0.74	<0.001
T2	0.871	0.138	<0.001	−0.093	0.128	0.47	
T3	1.543	0.145	<0.001	−0.051	0.134	0.71	
Resection type							<0.001
Partial	−0.617	0.096	<0.001	0.191	0.085	0.025	
Tumour location			<0.001			0.71	<0.001
C	−0.638	0.188	0.001	0.133	0.165	0.42	
M or A	−0.952	0.140	<0.001	0.049	0.118	0.68	

The regression coefficient for a specific covariate in the column 'Covariate effect' refers to the regression coefficient in a univariate Cox regression with only that covariate as time-fixed effect.

fit compared to the model without. A covariate by treatment by time interaction was only considered significant if the likelihood-ratio test for the model with both covariate, covariate by treatment and covariate by treatment by time indicated a significantly better fit compared the model without. The results of this multivariate analysis is shown in Table IV.

In the first step, each of the covariates was selected that was found to be significant in the uni-bivariate analysis reported in Table III, i.e. all covariates with the exception of gender. The only covariate by time interaction that remained significant was residual tumour by time. The hazard ratio for patients with residual tumour is 1.11 initially and increases with time; after 60 and 120 months it equals 4.22 and 5.26, respectively. In the second step, both treatment by time interaction and treatment were found to be significant ($p<0.001$ and $p=0.002$, respectively). The likelihood ratio test was significant as well ($p=0.001$, 2 degrees of freedom). The estimated coefficients of treatment and treatment by time in this second step did not differ very much from those of Table II. The third step did not result in any addition of covariate by treatment or covariate by treatment by time interaction.

It is important to communicate these results to physicians in a way that is useful and understandable to them. One way of doing this is by defining a prognostic score on the basis of all important prognostic factors that are not time-dependent. To illustrate this we will define a score on the basis of our final model in Table IV as the sum of 0.785 if the patient is lymph-node positive, 0.545 if the patient is above 65 years of age, 0.448 or 0.853 for tumour stage T2 or T3, respectively, and minus 0.347 for partial resection type.

Table IV. Multivariate Cox regression.

	Step 1		Step 2	
Variable	Coefficient (SE)	*p*-value	Coefficient (SE)	*p*-value
Lymph node				
Negative				
Positive	0.776 (0.113)	<0.001	0.785 (0.113)	<0.001
Age				
$\leqslant 65$				
>65	0.556 (0.097)	<0.001	0.545 (0.097)	<0.001
Residual tumour				
R0				
R1	0.117 (0.346)	0.73	0.106 (0.353)	0.76
T-stage				
T1		<0.001		
T2	0.441 (0.146)	0.002	0.448 (0.146)	0.002
T3	0.855 (0.164)	<0.001	0.853 (0.164)	<0.001
Resection type				
Total				
Partial	-0.343 (0.101)	0.001	-0.347 (0.101)	0.001
Residual tumour $*$ time				
R1$*\log(t+1)$	0.307 (0.121)	0.012	0.324 (0.125)	0.009
Treatment				
D1				
D2	–	–	0.819 (0.262)	0.0018
Treatment $*$ time				
D2$*\log(t+1)$	–	–	-0.308 (0.084)	<0.001

Step 1: all covariates and their interaction with time in months, forward selection; step 2: treatment and treatment by time interaction included; step 3: interactions of all covariates with treatment and with treatment by time, forward selection, not included here because it was identical to step 2.

One could try, by rounding the coefficients, to obtain a simplified score, but we will not pursue this here. The distribution of the prognostic scores in the D1/D2 study population is shown in Figure 7. Based on the score distribution in the D1/D2 study population, we divide the population into three parts of equal size, corresponding to low risk, medium risk and high risk patients. As representative scores for low, medium and high risk, we choose the median scores in these three regions, 0.10, 0.89 and 1.64, respectively. Figure 8 shows, for these three types of patients and for the two different levels R0 and R1 of residual tumour separately, the survival curves of D1 and D2 resection, obtained using the methods described in Section 3. The survival curves of D1 and D2 cross at 38.5 months for R0 (at survival probabilities of 0.798, 0.610 and 0.350 for low, medium and high risk, respectively) and 28.6 months for R1 (at survival probabilities of 0.628, 0.360 and 0.114 for low, medium and high risk, respectively). Because the 'baseline' risk for D1 dissection is progressively higher for higher risk patients, we see a phenomenon similar to that illustrated in Figure 6 occurring: depending on the baseline risk, the same regression coefficients for treatment and treatment

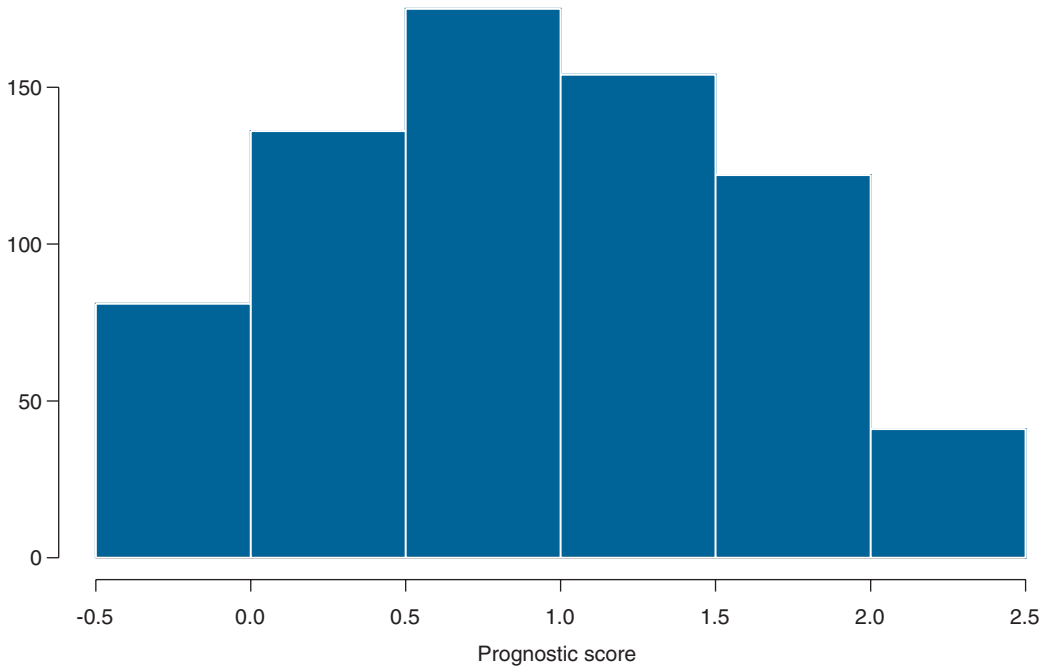

Figure 7. Histogram of the prognostic scores in the D1/D2 study population.

by time interaction lead to different conclusions regarding which treatment is better. For low risk patients, D2 dissection seems to be relatively advantageous, particularly for R0, while for high risk patients, particularly for R1, the reverse is true.

5. DISCUSSION

We have used data from the Dutch Gastric Cancer Trial to illustrate an important issue arising frequently in long-term survival data: violation of the proportional hazards assumption in the Cox regression model. We discussed a practical approach of dealing with this problem that is straightforward to implement in most standard statistical packages, the Cox regression model with time-dependent covariates. We have illustrated that in the context of such a model, a simple summary of the regression coefficients of both time-fixed and time-dependent covariates is not sufficient to fully describe the effect of these covariates on the survival probabilities. For this purpose, the baseline hazard needs to be estimated as well. Estimation of the baseline hazard, however, is still a weak point for most statistical packages, with the exception of S-plus and R.

Other approaches of dealing with time-dependent covariates effects are of course also possible. Some authors [8, 10] have suggested a spline approach to smooth the hazard ratio over time. Frailty models have also been used to model time-dependent effects of covariates, the underlying idea being that deviations from proportional hazards may reflect selection effects in

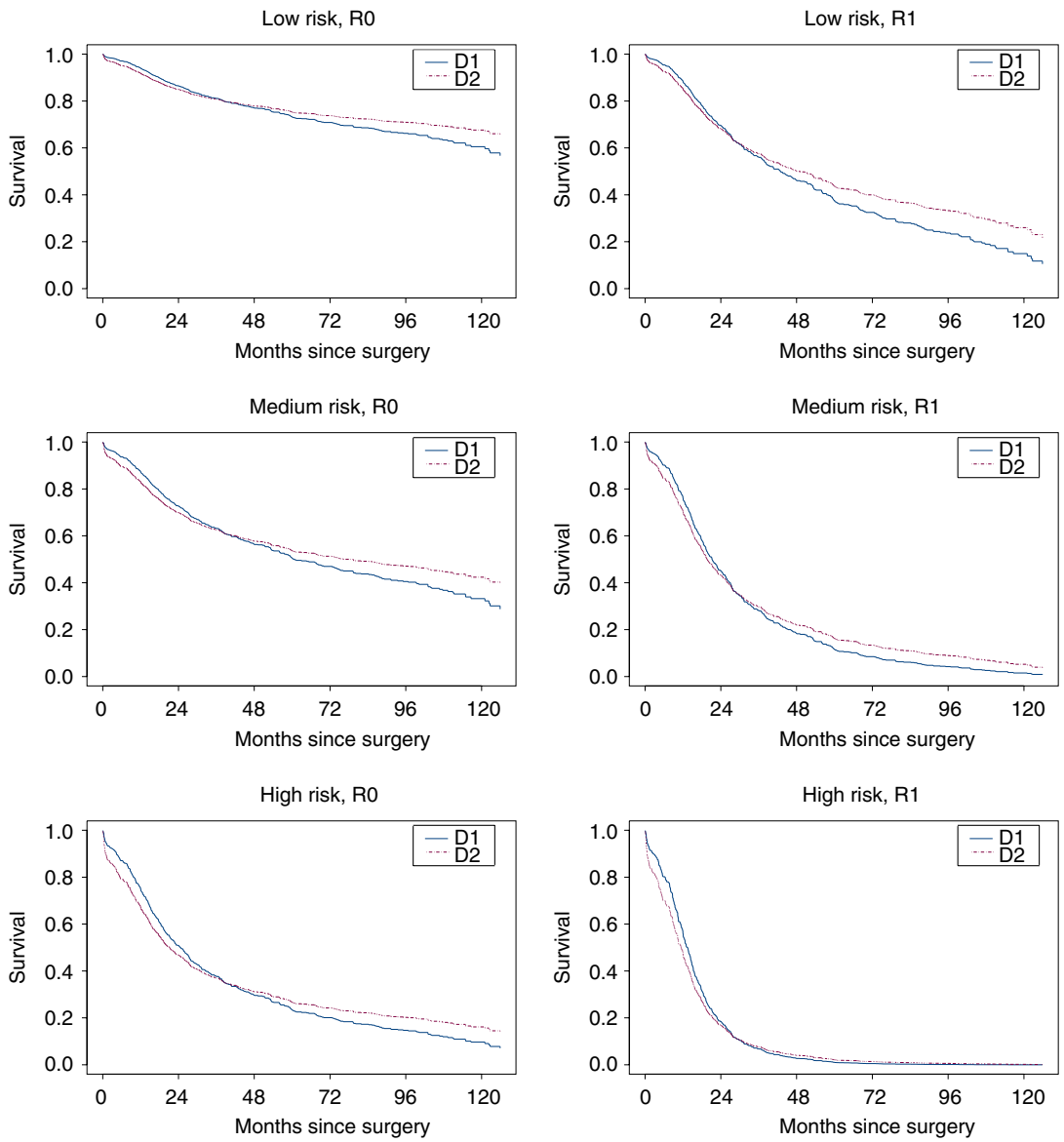

Figure 8. Model-based survival curves for D1- and D2-dissection, for low, medium and high risk patients, for R0 and R1 separately.

a heterogeneous population, due to omitted covariates [11]. Other approaches worth mentioning are additive hazards, accelerated failure time models, and a newly proposed semiparametric hazard regression model [12], containing both the Cox model and the accelerated failure time model as special cases. This model is able to capture certain time-dependent covariate effects in an efficient and elegant way. Two draw-backs of the semiparametric hazard regression

model (at this moment) are the lack of available software to estimate parameters within this model and the difficulty in interpreting the resulting estimates.

REFERENCES

1. Kaplan EL, Meier P. Non-parametric estimation from incomplete observations. *Journal of the American Statistical Association* 1958; **53**:457–481.
2. Cox DR. Regression models and life-tables. *Journal of the Royal Statistical Society B* 1972; **34**:187–220.
3. Klein JP, Moeschberger ML. *Survival Analysis. Techniques for Censored and Truncated Data*. Statistics for Biology and Health. Springer: New York, 1997.
4. Oquigley J, Pessione F. The problem of a covariate time qualitative interaction in a survival study. *Biometrics* 1991; **47**:101–115.
5. Bonenkamp JJ, Songun I, Hermans J, Sasako M, Welvaart K, Plukker JTM, Vanelk P, Obertop H, Gouma DJ, Taat CW, Vanlanschot J, Meyer S, Degraaf PW, Vonmeyenfeldt MF, Tilanus H, Van de Velde CJH. Randomized comparison of morbidity after D1 and D2 dissection for gastric-cancer in 996 Dutch patients. *Lancet* 1995; **345**:745–748.
6. Bonenkamp JJ, Hermans J, Sasako M, van de Velde CJH. Extended lymph-node dissection for gastric cancer. *New England Journal of Medicine* 1999; **340**:908–914.
7. Hartgrink HH, van De Velde CJ, Putter H, Bonenkamp JJ, Klein Kranenbarg E, Songun I, Welvaart K, Van Krieken JH, Meijer S, Plukker JT, Van Elk PJ, Obertop H, Gouma DJ, Van Lanschot JJ, Taat CW, De Graaf PW, Von Meyenfeldt MF, Tilanus H, Sasako M. Extended lymph node dissection for gastric cancer: who may benefit? Final results of the randomized Dutch Gastric Cancer Group trial. *Journal of Clinical Oncology* 2004; **22**:2069–2077.
8. Quantin C, Abrahamowicz M, Moreau T, Bartlett G, Mackenzie T, Tazi MA, Lalonde L, Faivre J. Variation over time of the effects of prognostic factors in a population-based study of colon cancer: comparison of statistical models. *American Journal of Epidemiology* 1999; **150**:1188–1200.
9. Therneau TM, Grambsch PM. *Modeling Survival Data. Extending the Cox Model*. Statistics for Biology and Health. Springer: New York, 2000.
10. Verweij PJM, van Houwelingen JC. Time-dependent effects of fixed covariates in Cox regression. *Biometrics* 1995; **51**:1550–1556.
11. Keiding N, Andersen PK, Klein JP. The role of frailty models and accelerated failure time models in describing heterogeneity due to omitted covariates. *Statistics in Medicine* 1997; **16**:215–224.
12. Chen YQ, Jewell NP. On a general class of semiparametric hazards regression models. *Biometrika* 2001; **88**:687–702.

