Table 3.3 *Survival times of 36 patients classified according to age group and whether or not they have had a nephrectomy.*

No nephrectomy			Nephrectomy		
<60	60–70	>70	<60	60–70	>70
9	15	12	104*	108*	10
6	8		9	26	9
21	17		56	14	18
			35	115	6
			52	52	
			68	5*	
			77*	18	
			84	36	
			8	9	
			38		
			72		
			36		
			48		
			26		
			108		
			5		

and that due to nephrectomy status is denoted by ν_k, $k = 1, 2$. The terms α_j and ν_k may then be included in proportional hazards models for $h_i(t)$, the hazard function for the ith individual in the study. Five possible models are as follows:

$$\text{Model (1):} \quad h_i(t) = h_0(t);$$

$$\text{Model (2):} \quad h_i(t) = \exp\{\alpha_j\} h_0(t);$$

$$\text{Model (3):} \quad h_i(t) = \exp\{\nu_k\} h_0(t);$$

$$\text{Model (4):} \quad h_i(t) = \exp\{\alpha_j + \nu_k\} h_0(t);$$

$$\text{Model (5):} \quad h_i(t) = \exp\{\alpha_j + \nu_k + (\alpha\nu)_{jk}\} h_0(t).$$

Under Model (1), the hazard of death does not depend on either of the two factors and is the same for all 36 individuals in the study. In Models (2) and (3), the hazard depends on either the age group or on whether a nephrectomy was performed, but not on both. In Model (4), the hazard depends on both factors, where the impact of nephrectomy on the hazard is independent of the age group of the patient. Model (5) includes an interaction between age group and nephrectomy, so that under this model the effect of a nephrectomy on the hazard of death depends on the age group of the patient.

To fit the term α_j, two indicator variables A_2 and A_3 are defined with values shown in the following table.

Age group	A_2	A_3
<60	0	0
60–70	1	0
>70	0	1

The term ν_k is fitted by defining a variable N which takes the value zero when no nephrectomy has been performed and unity when it has. With this choice of indicator variables, the baseline hazard function will correspond to an individual in the youngest age group who has not had a nephrectomy.

Models that contain the term α_j are then fitted by including the variables A_2, A_3 in the model, while the term ν_k is fitted by including N. The interaction is fitted by including the products $A_2N = A_2 \times N$ and $A_3N = A_3 \times N$ in the model. The explanatory variables fitted, and the values of $-2\log\hat{L}$ for each of the five models under consideration, are shown in Table 3.4. Some computer software for modelling survival data enables factors to be included in a model without the user having to define appropriate indicator variables. The values of $-2\log\hat{L}$ in Table 3.4 could then have been obtained directly using such software.

Table 3.4 *Values of $-2\log\hat{L}$ on fitting five models to the data in Table 3.3.*

Terms in model	Variables in model	$-2\log\hat{L}$
null model	none	177.667
α_j	A_2, A_3	172.172
ν_k	N	170.247
$\alpha_j + \nu_k$	A_2, A_3, N	165.508
$\alpha_j + \nu_k + (\alpha\nu)_{jk}$	A_2, A_3, N, A_2N, A_3N	162.479

The first step in comparing these different models is to determine if there is an interaction between nephrectomy status and age group. To do this, Model (4) is compared with Model (5). The reduction in the value of $-2\log\hat{L}$ on including the interaction term in the model that contains the main effects of age group and nephrectomy status is $165.508 - 162.479 = 3.029$ on 2 d.f. This is not significant ($P = 0.220$) and so we conclude that there is no interaction between age group and whether or not a nephrectomy has been performed.

We now determine whether the hazard function is related to neither, one or both of the factors age group and nephrectomy status. The change in the value of $-2\log\hat{L}$ on including the term α_j in the model that contains ν_k is $170.247 - 165.508 = 4.739$ on 2 d.f. This is significant at the 10% level ($P = 0.094$) and so there is some evidence that α_j is needed in a model

that contains ν_k. The change in $-2\log\hat{L}$ when ν_k is added to the model that contains α_j is $172.172 - 165.508 = 6.664$ on 1 d.f., which is significant at the 1% level ($P = 0.010$). Putting these two results together, the term α_j may add something to the model that includes ν_k, and ν_k is certainly needed in the model that contains α_j. This means that both terms are required, and that the hazard function depends on both the patient's age group and on whether or not a nephrectomy has been carried out.

Before leaving this example, let us consider other possible results from the comparison of the five models, and how they would affect the conclusion as to which model is the most appropriate. If the term corresponding to age group, α_j, was needed in a model in addition to the term corresponding to nephrectomy status, ν_k, and yet ν_k was not needed in the presence of α_j, the model containing just α_j, Model (2), is probably the most suitable. To make sure that α_j was needed at all, Model (2) would be further compared with Model (1), the null model. Similarly, if the term corresponding to nephrectomy status, ν_k, was needed in addition to the term corresponding to age group, α_j, but α_j was not required in the presence of ν_k, Model (3) would probably be satisfactory. However, the significance of ν_k would be checked by comparing Model (3) with Model (1). If neither of the terms corresponding to age group and nephrectomy status were needed in the presence of the other, a maximum of one variable would be required. To determine which of the two is necessary, Model (2) would be compared with Model (1) and Model (3) with Model (1). If both results were significant, on statistical grounds, the model that leads to the biggest reduction in the value of $-2\log\hat{L}$ from that for the null model would be adopted. If neither Model (2) nor Model (3) was superior to Model (1), we would conclude that neither age group nor nephrectomy status had an effect on the hazard function.

There are two further steps in the modelling approach to the analysis of survival data. First, we will need to critically examine the fit of a model to the observed data in order to ensure that the fitted proportional hazards model is indeed appropriate. Second, we will need to interpret the model, in order to quantify the effect that the explanatory variables have on the hazard function. Interpretation of parameters in a fitted model is considered in Section 3.7, while methods for assessing the adequacy of a fitted model will be considered in Chapter 4. But first, some general comments are made on possible strategies for model selection.

3.6 Strategy for model selection

An initial step in the model selection process is to identify a set of explanatory variables that have the potential for being included in the linear component of a proportional hazards model. This set will contain those variates and factors that have been recorded for each individual, but additionally terms corresponding to interactions between factors or between variates and factors may also be required.

Once a set of potential explanatory variables has been isolated, the combination of variables that are to be used in modelling the hazard function has to be determined. In practice, a hazard function will not depend on a unique combination of variables. Instead, there are likely to be a number of equally good models, rather than a single "best" model. For this reason, it is desirable to consider a wide range of possible models.

An important principle in statistical modelling is that when a term corresponding to the interaction between two factors is to be included in a model, the corresponding lower-order terms should also be included. This rule is known as the *hierarchic principle*, and means that interactions should not be fitted unless the corresponding main effects are present. Models that are not hierarchic are difficult to interpret.

The model selection strategy depends to some extent on the purpose of the study. In some applications, information on a number of variables will have been obtained, and the aim might be to determine which of them has an effect on the hazard function, as in Example 1.3 on multiple myeloma. In other situations, there may be one or more variables of primary interest, such as terms corresponding to a treatment effect. The aim of the modelling process is then to evaluate the effect of such variables on the hazard function, as in Example 1.4 on prostatic cancer. Since the other variables that have been recorded might also be expected to influence the magnitude of the treatment effect, these variables will need to be taken account of in the modelling process.

3.6.1 Variable selection procedures

We first consider the situation where all explanatory variables are on an equal footing, and the aim is to identify subsets of variables upon which the hazard function depends. When the number of potential explanatory variables, including interactions, non-linear terms and so on, is not too large, it might be feasible to fit all possible combinations of terms, paying due regard to the hierarchic principle. Alternative nested models can be compared by examining the change in the value of $-2\log\hat{L}$ on adding terms into a model or deleting terms from a model.

Comparisons between a number of possible models, which need not necessarily be nested, can also be made on the basis of the statistic

$$AIC = -2\log\hat{L} + \alpha q,$$

in which q is the number of unknown β-parameters in the model and α is a predetermined constant. This statistic is known as *Akaike's information criterion*; the smaller the value of this statistic, the better the model. The motivation behind this statistic is that if the only difference between two models is that one includes unnecessary covariates, the values of AIC for the two models will not be very different. Indeed, the value of AIC will tend to increase when unnecessary terms are added to the model.

Values of α between 2 and 6 are generally used in computing the value of the statistic. The choice $\alpha = 3$ is roughly equivalent to using a 5% significance

level in judging the difference between the values of $-2\log L$ for two nested models that differ by between one and three parameters. This value of α is recommended for general use.

Of course, some terms may be identified as alternatives to those in a particular model, leading to subsets that are equally suitable. The decision on which of these subsets is the most appropriate should not then rest on statistical grounds alone. When there are no subject matter grounds for model choice, the model chosen for initial consideration from a set of alternatives might be the one for which the value of $-2\log\hat{L}$ or AIC is a minimum. It will then be important to confirm that the model does fit the data using the methods for model checking described in Chapter 4.

In some applications, information might be recorded on a number of variables, all of which relate to the same general feature. For example, the variables height, weight, body mass index (weight/height^2), head circumference, arm length, and so on, are all concerned with the size of an individual. In view of inter-relationships between these variables, a model for the survival times of these individuals may not need to include each of them. It would then be appropriate to determine which variables from this group should be included in the model, although it may not matter exactly which variables are chosen.

When the number of variables is relatively large, it can be computationally expensive to fit all possible models. In particular, if there is a pool of p potential explanatory variables, there are 2^p possible combinations of terms, so that if $p > 10$, there are more than a thousand possible combinations of explanatory variables. In this situation, automatic routines for variable selection that are available in many software packages might seem an attractive prospect. These routines are based on *forward selection*, *backward elimination* or a combination of the two known as the *stepwise procedure*.

In forward selection, variables are added to the model one at a time. At each stage in the process, the variable added is the one that gives the largest decrease in the value of $-2\log\hat{L}$ on its inclusion. The process ends when the next candidate for inclusion in the model does not reduce the value of $-2\log\hat{L}$ by more than a prespecified amount. This is known as the *stopping rule*. This rule is often couched in terms of the significance level of the difference in the values of $-2\log\hat{L}$ when a variable is added to a model, so that the selection process ends when the next term for inclusion ceases to be significant at a pre-assigned level.

In backward elimination, a model that contains the largest number of variables under consideration is first fitted. Variables are then excluded one at a time. At each stage, the variable omitted is the one that increases the value of $-2\log\hat{L}$ by the smallest amount on its exclusion. The process ends when the next candidate for deletion increases the value of $-2\log\hat{L}$ by more than a prespecified amount.

The stepwise procedure operates in the same way as forward selection. However, a variable that has been included in the model can be considered for exclusion at a later stage. Thus after adding a variable to the model, the procedure then checks whether any previously included variable can now be deleted. These decisions are again made on the basis of prespecified stopping rules.

These automatic routines have a number of disadvantages. Typically, they lead to the identification of one particular subset, rather than a set of equally good ones. The subsets found by these routines often depend on the variable selection process that has been used, that is, whether it is forward selection, backward elimination or the stepwise procedure, and generally tend not to take any account of the hierarchic principle. They also depend on the stopping rule that is used to determine whether a term should be included in or excluded from a model. For all these reasons, these automatic routines have a limited role in model selection, and should certainly not be used uncritically.

Instead of using automatic variable selection procedures, the following general strategy for model selection is recommended.

1. The first step is to fit models that contain each of the variables one at a time. The values of $-2\log\hat{L}$ for these models are then compared with that for the null model to determine which variables on their own significantly reduce the value of this statistic.

2. The variables that appear to be important from Step 1 are then fitted together. In the presence of certain variables, others may cease to be important. Consequently, those variables that do not significantly increase the value of $-2\log\hat{L}$ when they are omitted from the model can now be discarded. We therefore compute the change in the value of $-2\log\hat{L}$ when each variable on its own is omitted from the set. Only those that lead to a significant increase in the value of $-2\log\hat{L}$ are retained in the model. Once a variable has been dropped, the effect of omitting each of the remaining variables in turn should be examined.

3. Variables that were not important on their own, and so were not under consideration in Step 2, may become important in the presence of others. These variables are therefore added to the model from Step 2, one at a time, and any that reduce $-2\log\hat{L}$ significantly are retained in the model. This process may result in terms in the model determined at Step 2 ceasing to be significant.

4. A final check is made to ensure that no term in the model can be omitted without significantly increasing the value of $-2\log\hat{L}$, and that no term not included significantly reduces $-2\log\hat{L}$.

When using this selection procedure, rigid application of a particular significance level should be avoided. In order to guide decisions on whether to include or omit a term, the significance level should not be too small; a level of around 10% is recommended.

In some applications, a small number of interactions and other higher-order terms, such as powers of certain variates, may need to be considered for inclusion in a model. Such terms would be added to the model identified in Step 3 above, after ensuring that any terms necessitated by the hierarchic principle have already been included in the model. If any higher-order term leads to a

significant reduction in the value of $-2\log\hat{L}$, that term would be included in the model.

The procedure outlined above is now illustrated in an example.

Example 3.5 Survival of multiple myeloma patients
The analysis of the data on the survival times of multiple myeloma patients in Example 3.2 suggested that not all of the seven explanatory variables, *Age*, *Sex*, *Bun*, *Ca*, *Hb*, *Pcells*, and *Protein*, are needed in a proportional hazards model. We now determine the most appropriate subsets of these variables. In this example, transformations of the original variables and interactions between them will not be considered. We will further assume that there are no medical grounds for including particular variables in a model. A summary of the values of $-2\log\hat{L}$ for all models that are to be considered is given in Table 3.5.

Table 3.5 *Values of* $-2\log\hat{L}$ *for models fitted to the data from Example 1.3.*

Variables in model	$-2\log\hat{L}$
none	215.940
Age	215.817
Sex	215.906
Bun	207.453
Ca	215.494
Hb	211.068
Pcells	215.875
Protein	213.890
Hb + *Bun*	202.938
Hb + *Protein*	209.829
Bun + *Protein*	203.641
Bun + *Hb* + *Protein*	200.503
Hb + *Bun* + *Age*	202.669
Hb + *Bun* + *Sex*	202.553
Hb + *Bun* + *Ca*	202.937
Hb + *Bun* + *Pcells*	202.773

The first step is to fit the null model and models that contain each of the seven explanatory variables on their own. Of these variables, *Bun* leads to the largest reduction in $-2\log\hat{L}$, reducing the value of the statistic from 215.940 to 207.453. This reduction of 8.487 is significant at the 1% level ($P = 0.004$) when compared with percentage points of the chi-squared distribution on 1 d.f. The reduction in $-2\log\hat{L}$ on adding *Hb* to the null model is 4.872, which is also significant at the 5% level ($P = 0.027$). The only other variable that on its own has some explanatory power is *Protein*, which leads to a reduction in $-2\log\hat{L}$ that is nearly significant at the 15% level ($P = 0.152$). Although

this P-value is relatively high, we will for the moment keep *Protein* under consideration for inclusion in the model.

The next step is to fit the model that contains *Bun*, *Hb* and *Protein*, which leads to a value of $-2\log\hat{L}$ of 200.503. The effect of omitting each of the three variables in turn from this model is shown in Table 3.5. In particular, when *Bun* is omitted, the increase in $-2\log\hat{L}$ is 9.326, when *Hb* is omitted the increase is 3.138, and when *Protein* is omitted it is 2.435. Each of these changes in the value of $-2\log\hat{L}$ can be compared with percentage points of a chi-squared distribution on 1 d.f. Since *Protein* does not appear to be needed in the model, in the presence of *Hb* and *Bun*, this variable will not be further considered for inclusion.

If either *Hb* or *Bun* is excluded from the model that contains both of these variables, the increase in $-2\log\hat{L}$ is 4.515 and 8.130, respectively. Both of these increases are significant at the 5% level, and so neither *Hb* nor *Bun* can be excluded from the model without significantly increasing the value of the $-2\log\hat{L}$ statistic.

Finally, we look to see if any of variables *Age*, *Sex*, *Ca* and *Pcells* should be included in the model that contains *Bun* and *Hb*. Table 3.5 shows that when any of these four variables is added, the reduction in $-2\log\hat{L}$ is less than 0.5, and so none of them need to be included in the model. We therefore conclude that the most satisfactory model is that containing *Bun* and *Hb*.

We now turn to studies where there are variables of primary importance, such as a treatment effect. Here, we proceed in the following manner.

1. The important prognostic variables are first selected, ignoring the treatment effect. Models with all possible combinations of the variables can be fitted when their number is not too large. Alternatively, the variable selection process might follow similar lines to those described previously in Steps 1 to 4.

2. The treatment effect is then included in the model. In this way, any differences between the two groups that arise as a result of differences between the distributions of the prognostic variables in each treatment group, are not attributed to the treatment.

3. If the possibility of interactions between the treatment and other explanatory variables has not been discounted, these must be considered before the treatment effect can be interpreted.

It will often be interesting to fit a model that contains the treatment effect alone. This enables the effect that the prognostic variables have on the magnitude of the treatment effect to be evaluated.

In this discussion on strategies for model selection, the use of statistical criteria to guide the selection process has been emphasised. In addition, due account must be taken of the application area. In particular, on subject area grounds, it may be inappropriate to include particular combinations of variables. On the other hand, there might be some variables that it is not sensible to omit from the model, even if they appear not to be needed in modelling a

particular data set. Indeed, there is always a need for non-statistical considerations in model building.

Example 3.6 Comparison of two treatments for prostatic cancer
In the data from Example 1.4 on the survival times of 38 prostatic cancer patients, there are four prognostic variables that might have an effect on the survival times. These are the age of the patient in years (*Age*), serum haemoglobin level (*Shb*), tumour size (*Size*) and Gleason index (*Index*). All possible combinations of these variates are fitted in a proportional hazards model and the values of $-2\log\hat{L}$ computed. These values are shown in Table 3.6, together with the values of Akaike's information criterion, computed with $\alpha = 3$.

Table 3.6 *Values of* $-2\log\hat{L}$ *and AIC for models fitted to the data from Example 1.4.*

Variables in model	$-2\log\hat{L}$	*AIC*
none	36.349	36.349
Age	36.269	39.269
Shb	36.196	39.196
Size	29.042	32.042
Index	29.127	32.127
Age + *Shb*	36.151	42.151
Age + *Size*	28.854	34.854
Age + *Index*	28.760	34.760
Shb + *Size*	29.019	35.019
Shb + *Index*	27.981	33.981
Size + *Index*	23.533	29.533
Age + *Shb* + *Size*	28.852	37.852
Age + *Shb* + *Index*	27.893	36.893
Age + *Size* + *Index*	23.269	32.269
Shb + *Size* + *Index*	23.508	32.508
Age + *Shb* + *Size* + *Index*	23.231	35.231

The two most important explanatory variables when considered separately are *Size* and *Index*. From the change in the value of $-2\log\hat{L}$ on omitting either of them from a model that contains both, we deduce that both variables are needed in a proportional hazards model. The value of $-2\log\hat{L}$ is only reduced by a very small amount when *Age* and *Shb* are added to the model that contains *Size* and *Index*. We therefore conclude that only *Size* and *Index* are important prognostic variables.

From the values of Akaike's information criterion in Table 3.6, the model with *Size* and *Index* leads to the smallest value of the statistic, confirming that this is the most suitable model of those tried. Notice also that there are no other combinations of explanatory variables that lead to similar values of

the *AIC*-statistic, which shows that there are no obvious alternatives to using *Size* and *Index* in the model.

We now consider the treatment effect. Let *Treat* be a variable that takes the value zero for individuals allocated to the placebo, and unity for those allocated to DES. When *Treat* is added to the model that contains *Size* and *Index*, the value of $-2\log\hat{L}$ is reduced to 22.572. This reduction of 0.961 on 1 d.f. is not significant ($P = 0.327$). This indicates that there is no treatment effect, but first we ought to examine whether the coefficients of the two explanatory variables in the model depend on treatment. To do this, we form the products *Tsize* = *Treat* × *Size* and *Tindex* = *Treat* × *Index*, and add these to the model that contains *Size*, *Index* and *Treat*. When *Tsize* and *Tindex* are added to the model, $-2\log\hat{L}$ is reduced to 20.829 and 20.792, respectively. On adding both of these mixed terms, $-2\log\hat{L}$ becomes 19.705. The reductions in $-2\log\hat{L}$ on adding these terms to the model are not significant, and so there is no evidence that the treatment effect depends on *Size* and *Index*. This means that our original interpretation of the size of the treatment effect is valid, and that on the basis of these data, treatment with DES does not appear to affect the hazard of death. The estimated size of this treatment effect will be considered later in Example 3.10.

Before leaving this example, we note that when either *Tsize* or *Tindex* is added to the model, their estimated coefficient, and that of *Treat*, become large. The standard errors of these estimates are also very large. In particular, in the model that contains *Size*, *Index*, *Treat* and *Tsize*, the estimated coefficient of *Treat* is −11.28 with a standard error of 18.50. For the model that contains *Size*, *Index*, *Treat* and *Tindex*, the coefficients of *Treat* and *Tindex* are −161.52 and 14.66, respectively, while the standard errors of these estimates are 18476 and 1680, respectively! This is evidence of *overfitting*.

In an overfitted model, the estimated values of some of the β-coefficients will be highly dependent on the actual data. A very slight change to the values of one of these variables could then have a large impact on the estimate of the corresponding coefficient. This is the reason for such estimates having large standard errors.

An overfitted model is one that is more complicated than is justified by the data, and does not provide a useful summary of the data. This is another reason for not including the mixed terms in the model for the hazard of death from prostatic cancer.

3.6.2 Testing for non-linearity

When the dependence of the hazard function on a variate that takes a wide range of values is to be modelled, we should consider whether the variate should be included as a linear term in the proportional hazards model.

For some variates, transformations of their original values might be used in place of the original variate. For example, if a variate takes a wide range of values, that variate might first be transformed by taking logarithms. This is particularly appropriate for variates that are strictly positive. The logarithm

of a variate may also be used when the distribution of the values of the variate is highly positively skew.

When there are no *a priori* reasons for transforming a variate, the assumption of linearity in the variate should at least be critically examined. One possibility is to add quadratic or even cubic terms to the model, and examine the consequent reduction in the value of $-2\log\hat{L}$. If the inclusion of such terms significantly reduces the value of this statistic, we would conclude that there is non-linearity, and incorporate the polynomial terms in the model.

In many situations, non-linearity in an explanatory variate cannot be adequately represented by including polynomial terms in a model, or by transforming the original variable. For this reason, the following procedure is recommended for general use.

The values of the variate are first grouped into four or five categories containing approximately equal numbers of observations. A factor is then defined whose levels correspond to this grouping. For example, a variate reflecting the size of a tumour could be fitted as a factor whose levels correspond to very small, small, medium and large.

More specifically, let A be a factor with m levels formed from a continuous variate, and let X be a variate that takes the value j when A is at level j, for $j = 1, 2, \ldots, m$. Linearity in the original variate will then correspond to there being a linear trend across the levels of A. This linear trend can be modelled by fitting X alone. Now, fitting the $m-1$ terms $X, X^2, \ldots, X^{m-1}$ is equivalent to fitting A as a factor in the model, using indicator variables as in Section 3.2.1. Accordingly, the difference between the value of $-2\log\hat{L}$ for the model that contains X, and that for the model that contains A, is a measure of non-linearity across the levels of A. If this difference is not significant we would conclude that there is no non-linearity and the original variate would be fitted. On the other hand, if there is evidence of non-linearity, the factor which corresponds to the variate is fitted.

The actual form of the non-linearity can be further studied from the coefficients of the indicator variables corresponding to A. Indeed, a plot of these coefficients may help in establishing the nature of any trend across the levels of the factor A.

Example 3.7 Survival of multiple myeloma patients

In Example 3.5, we found that a proportional hazards model that contained the explanatory variables *Bun* and *Hb* appeared to be appropriate for the data on the survival times of multiple myeloma patients. We now consider whether there is any evidence of non-linearity in the values of serum haemoglobin level, and examine whether a quadratic term is needed in the proportional hazards model that contains *Bun* and *Hb*. When the term Hb^2 is added to this model, the value of $-2\log\hat{L}$ is reduced from 202.938 to 202.917. This reduction of 0.021 on 1 d.f. is clearly not significant, which suggests that a linear term in *Hb* is sufficient.

An alternative way of examining the extent of non-linearity is to use a factor to model the effect of serum haemoglobin level on the hazard function.

Suppose that a factor with four levels is defined, where level 1 corresponds to values of *Hb* less than or equal to 7, level 2 to values between 7 and 10, level 3 to values between 10 and 13 and level 4 to values greater than 13. This choice of levels corresponds roughly to the quartiles of the distribution of the values of *Hb*. This factor can be fitted by defining three indicator variables, *Hb*2, *Hb*3 and *Hb*4, which take the values shown in the following table.

Level of factor (X)	Value of *Hb*	*Hb*2	*Hb*3	*Hb*4
1	$Hb \leqslant 7$	0	0	0
2	$7 < Hb \leqslant 10$	1	0	0
3	$10 < Hb \leqslant 13$	0	1	0
4	$Hb > 13$	0	0	1

When a model containing *Bun*, *Hb*2, *Hb*3 and *Hb*4 is fitted, the value of $-2\log\hat{L}$ is 200.417. The change in the value of this statistic on adding the indicator variables *Hb*2, *Hb*3 and *Hb*4 to the model that contains *Bun* alone is 7.036 on 3 d.f., which is significant at the 10% level ($P = 0.071$). However, it is difficult to identify any pattern across the factor levels.

A linear trend across the levels of the factor corresponding to haemoglobin level can be modelled by fitting the variate X, which takes values 1, 2, 3, 4, according to the factor level. When the model containing *Bun* and X is fitted, $-2\log\hat{L}$ is 203.891, and the change in the value of $-2\log\hat{L}$ due to any non-linearity is $203.891 - 200.417 = 3.474$ on 2 d.f. This is not significant when compared with percentage points of the chi-squared distribution on 2 d.f. ($P = 0.176$). We therefore conclude that the effect of haemoglobin level on the hazard of death in this group of patients is adequately modelled by using the linear term *Hb*.

3.7 Interpretation of parameter estimates

When the proportional hazards model is used in the analysis of survival data, the coefficients of the explanatory variables in the model can be interpreted as logarithms of the ratio of the hazard of death to the baseline hazard. This means that estimates of this hazard ratio, and corresponding confidence intervals, can easily be found from the fitted model. The interpretation of parameters corresponding to different types of term in the proportional hazards model is described in the following sections.

3.7.1 Models with a variate

Suppose that a proportional hazards model contains a single continuous variable X, so that the hazard function for the ith of n individuals, for whom X takes the value x_i, is

$$h_i(t) = e^{\beta x_i} h_0(t).$$