ANALYSIS OF FAILURE TIME DATA WITH ORDINAL CATEGORIES OF RESPONSE

DAMON M. BERRIDGE AND JOHN WHITEHEAD
Department of Applied Statistics, University of Reading, Whiteknights, P.O. Box 217, Reading RG6 2AN, U.K.

SUMMARY
When failure times are observed, additional information concerning the type of failure is often recorded. A method which simultaneously models the failure times and additional information in the form of ordinal categories is discussed. An application to clinical trial data, in which the failure times are times of onset of headache, and the headaches are classified into the ordinal categories mild, moderate and severe, illustrates how this method may be used and how the final model can be interpreted. The continuation ratio model, which is used in this method, is described in detail.

1. INTRODUCTION
In failure time studies, further information on each individual may be available at time of failure. This information may be in the form of discrete and non-ordered categories, such as cause of death, in which case competing risks methodology may be used to analyse the data. However, instances occur where this additional information is ordered according to some criterion. For example, severity of lesions may be recorded at time of death in animal toxicity experiments. Lesions may be absent, or present and graded on a scale from 1 to 5, say. As another example, the nature of adverse events (ranging from dizziness to death) may be observed on patients during a clinical trial. In both of these examples, an analysis by competing risks would not take the ordinality of categories into account.

A method which does utilize the ordinal structure is described in this paper. The method combines a proportional hazards model for the failure times with a continuation ratio model for the ordinal categories. As the latter is less familiar, it will be reviewed in Section 2. The composite model is introduced in Section 3.

2. CONTINUATION RATIOS
Consider the following situation. In order to assess the effect of various treatments (such as diet or chemotherapy), individuals are allocated to one of b treatments $B_1, \ldots, B_b$. The effect of the treatments is determined by comparing treatment groups with respect to some prespecified response (such as severity of some condition) observed on each individual at the end of the study. If response is graded on a scale of c categories $C_1, \ldots, C_c$ in which category $C_j$ is less severe, say, than $C_{j+1}, j = 1, \ldots, c - 1$, then the classification is ordinal.

An analysis which describes the effect of treatments on response should take this ordinality into account. Category $C_1$ is often a definite baseline (such as condition absent) with which all other
categories can be compared, and so in the appropriate analysis $C_1$ will play a special role. One method which satisfies the above requirements is to use the continuation ratio model.\(^3\)\(^4\)

Let $\pi_{ij}$ denote the probability that the response of an individual on treatment $B_i$ is classified as $C_j$, and let $\gamma_{ij} = \pi_{i1} + \ldots + \pi_{ij}$. The continuation ratio corresponding to an individual on treatment $B_i$ with category $C_j$ is defined as $\pi_{ij}/(1 - \gamma_{ij})$. Then the continuation ratio model is specified by the equation

$$\log \left( \frac{\pi_{ij}}{1 - \gamma_{ij}} \right) = -\theta_j - \alpha_i, \quad i = 1, \ldots, b, \quad j = 1, \ldots, c - 1,$$

where $\theta_j$ is the so-called 'cut-point' between category $C_j$ and those categories more severe than $C_j$, $j = 1, \ldots, c - 1$, and $\alpha_i$ is the effect of treatment $B_i$ ($\alpha_1$ assumed zero), $i = 1, \ldots, b$. This model is based on the assumption that the same set of $\alpha_i$ values is valid for all $j, j = 1, \ldots, c - 1$.

For ease of interpretation, the log continuation ratio corresponding to category $C_j$ for an individual on treatment $B_i$, $\log \left( \frac{\pi_{ij}}{1 - \gamma_{ij}} \right)$, may be expressed as logit $\{\pi_{ij}/(1 - \gamma_{ij} + \pi_{ij})\}$; this is the log odds that, given his category is $C_j$ or more severe, an individual on treatment $B_i$ will have the mildest of these categories, $C_j, j = 1, \ldots, c - 1$. Hence a positive value of $\alpha_i$ ($i = 2, \ldots, g$) implies that individuals on $B_i$ are likely to have more severe categories than individuals on $B_1$, that is that treatment $B_i$ is worse than $B_1$.

Let $n_{ij}$ be the observed number of individuals on treatment $B_i$ who respond with category $C_j$. Let $C_j$ and $\bar{C}_j$ denote the events that an individual responds and does not respond with category $C_j$ respectively. Then the probability of responding with $C_j$ is the product of the probabilities of the events $C_1, C_2, \ldots, C_{j-1}$, given $C_1, \ldots, C_{j-1}$ and $C_j$, given $C_1, \ldots, C_{j-1}$, and $C_j$. Thus $\pi_{ij}$ may be expressed as the product of $j$ terms

$$\left( \frac{1 - \pi_{i1}}{1 - \gamma_{i1} + \pi_{i1}} \right) \left( \frac{1 - \pi_{i2}}{1 - \gamma_{i2} + \pi_{i2}} \right) \ldots \left( \frac{1 - \pi_{ij-1}}{1 - \gamma_{ij-1} + \pi_{ij-1}} \right) \left( \frac{\pi_{ij}}{1 - \gamma_{ij} + \pi_{ij}} \right),$$

for $j = 1, \ldots, c - 1$. Thus $\pi_{ic}$ may be expressed as

$$\left( \frac{1 - \pi_{i1}}{1 - \gamma_{i1} + \pi_{i1}} \right) \left( \frac{1 - \pi_{i2}}{1 - \gamma_{i2} + \pi_{i2}} \right) \ldots \left( \frac{1 - \pi_{ic-1}}{1 - \gamma_{ic-1} + \pi_{ic-1}} \right) \left( \frac{\pi_{ic}}{1 - \gamma_{ic} + \pi_{ic}} \right).$$

Consequently, the multinomial likelihood can be written as the product of $(c - 1)$ binomial likelihoods, the $j$th of which has sample sizes $n_{ij} + \ldots + n_{ic}$ and cell probabilities $\pi_{ij}/(1 - \gamma_{ij} + \pi_{ij}), i = 1, \ldots, b$.\(^5\) In the case of two treatment groups, the multinomial likelihood can be expressed as a product of $(c - 1)$ binary likelihoods

$$\prod_{j=1}^{c-1} \left\{ \prod_{i=1}^{b} \left[ \left( \frac{\pi_{ij}}{1 - \gamma_{ij} + \pi_{ij}} \right)^{n_{ij}} \left( \frac{\pi_{ij}}{1 - \gamma_{ij} + \pi_{ij}} \right)^{n_{ij+1} + \ldots + n_{i}} \right] \right\}.$$

In order to fit a continuation ratio model, it is assumed that all $(c - 1)$ logistic models modelling these binary likelihoods share the same odds ratio between treatments, namely $\exp(\alpha_2)$. We now explain how a continuation ratio model can be fitted.

A row of data is generated for each individual. This row includes an outcome variable IND with value 0 if the response is $C_1$ and value 1 otherwise, and a variable CUTPT at level 1 for all individuals. A second row of data is included for each individual that responds with $C_2$ or a more severe category, again including outcome variable IND with value 0 if the response is $C_2$ and value 1 if it is more severe than $C_2$, and variable CUTPT at level 2. This data generation continues up to the $(c - 1)$th row of data for those individuals who respond with $C_{c-1}$ or $C_c$, with IND equal to 0 if the response is $C_{c-1}$ and equal to 1 if it is $C_c$. In the $(c - 1)$th row of data,
CUTPT equals c - 1. D. M. Berridge has written an SAS macro which performs the above operations. One version of this macro is given in the Appendix.

The transformed data set can be used either with the binomial error structure and logit link facilities in GLIM, or in conjunction with SAS procedures LOGIST\textsuperscript{6} or CATMOD,\textsuperscript{7} to fit the \((c - 1)\) binary logistic models simultaneously. How the SAS macro can be implemented with PROC LOGIST in the CMS operating system is explained in the Appendix.

In the two-treatment case, IND is the binary response regressed on CUTPT and treatment main effect, the latter of which is fixed over all \((c - 1)\) binary logistic models so as to simulate the common log odds ratio. The assumption of a common log odds ratio is checked by testing for the interaction between CUTPT and the treatment main effect.

3. COMBINING PROPORTIONAL HAZARDS AND CONTINUATION RATIOS

Consider the situation in Section 2, in which each individual on one of \(b\) treatments \(B_i, i = 1, \ldots, b\), has an ordinal category of response \(C_j, j = 1, \ldots, c\). Category of response corresponding to each individual may be observed at either a time of failure or a censoring time. Assume here and in the following section that all responses are observed at failure times. The situation in which responses may be observed at censoring times (for example, when an individual reaches the end of the study without responding) is covered in Section 5.

It may be shown\textsuperscript{8} that the hazard of a particular category of response at a given point in time for an individual on a specific treatment can be expressed as the product of two components:

(a) The hazard \(h_i(t)\) of an individual on treatment \(B_i\) responding at time \(t\), modelled via proportional hazards as

\[
h_i(t) = h_1(t) \exp(\beta_i), \quad i = 1, \ldots, b, \quad t \geq 0,
\]

where \(h_1(t)\) is the underlying (assumed unknown) hazard at time \(t\) corresponding to an individual allocated to treatment \(B_1\), and \(\beta_i\) is the effect on the underlying hazard of treatment \(B_i (\beta_1 = 0), i = 1, \ldots, b\);

(b) The probability \(n_{ij}(t)\) of an individual on treatment \(B_i\) responding with category \(C_j\), given that he responds at time \(t\).

Let \(\gamma_{ij}(t) = \pi_{1i}(t) + \ldots + \pi_{ij}(t)\). Section 2 describes how continuation ratios can model \(\pi_{ij}\). In order to model \(\pi_{ij}(t)\), the continuation ratio model in Section 2 is adjusted for time as

\[
\log\left(\frac{\pi_{ij}(t)}{1 - \gamma_{ij}(t)}\right) = -\theta_j - \alpha_i - \delta f(t), \quad i = 1, \ldots, b, \quad j = 1, \ldots, c - 1, \quad t \geq 0,
\]

where \(\alpha_i\) and \(\theta_j\) are as defined in Section 2, and \(\delta\) is the effect on the log continuation ratio of some function of time \(f(t)\). This function may be a polynomial in \(t\), or could be a transformation of \(t\), for example \(\log(t)\) if \(t > 0\) or \(\log(t + 1)\) if \(t \geq 0\).

The study period can be divided into intervals of time and a log continuation ratio for each cut-point calculated over each interval. The most appropriate form for \(f(t)\) may then be indicated in a plot of log continuation ratio against interval, with one plot per cut-point. Thus time is represented as a factor with discrete levels, each corresponding to such an interval. This representation could be used as a mechanism for choosing \(f(t)\), or used in the continuation ratio model in its own right, especially if the above plots reveal no tangible form for \(f(t)\).
Hence, the joint hazard \( h_{ij}(t) \) of an individual on treatment \( B_i \) who responds with category \( C_j \) at failure time \( t \) may be written as

\[ h_{ij}(t) = h_i(t) \pi_{ij}(t). \]

How this expression can be used to analyse data is illustrated by way of the example presented in the following section.

4. EXAMPLE

The data set used in this section has been extracted from data on 620 patients entered into a clinical trial. The objective of the trial was to evaluate the safety and efficacy of a drug used by many patients suffering a chronic condition. After a two-week run-in period, patients received the drug for ten weeks according to one of the two different dosing schedules \( B_1 \) and \( B_2 \).

At some point during the study, patients may suffer such adverse events as headaches or dizziness. For illustrative purposes, attention is concentrated on one type of adverse event, headache, graded into one of three ordinal categories (with a definite baseline category \( C_1 \)): mild \( (C_1) \), moderate \( (C_2) \) and severe \( (C_3) \). Day of onset \( t \) and sex (amongst other prognostic factors) are also recorded and, along with the ordinal response, could be analysed using the technique described in Section 3. The 35 patients who suffered a headache are uncensored in the proportional hazards component. Those patients who suffered either no adverse events or other forms of adverse event are treated as being censored in the proportional hazards component and are excluded from the continuation ratio component.

Consider all patients who suffered a headache. Define \( h_i(t; z) \) as the hazard of a patient with prognostic factor \( z \) (for example sex with female coded 0 and male 1) on schedule \( B_i \), \( i = 1, 2 \), suffering a headache on day \( t \). Define \( \pi_{ij}(t; z) \) as the probability that this same patient suffers a headache of severity \( C_j \), \( J = 1, 2, 3 \), given that day of onset is \( t \). Then the hazard \( h_{ij}(t; z) \) of this patient suffering a headache of severity \( C_j \), \( j = 1, 2, 3 \), on day \( t \) could be modelled as

\[ h_{ij}(t; z) = h_i(t; z) \pi_{ij}(t; z). \]

The empirical survivor function can be estimated for each schedule separately, and a log(-log) transformation of these estimates plotted against log(t). If there is a constant vertical difference between the two curves over time, then the proportional hazards assumption of a constant hazard ratio over time is satisfied. In the present example, this underlying assumption is deemed appropriate. A test of the cut-point by schedule interaction will check the assumption underlying the continuation ratio model.

In the proportional hazards component, sex and schedule main effects and sex by schedule interaction, fitted using SAS procedure PHGLM,\(^6\) are all non-significant. However, for illustrative purposes, the schedule main effect is included in the final model of this component

\[ h_i(t) = h_1(t) \exp(\beta_i), \quad i = 1, 2, \quad t > 0, \quad (1) \]

where \( h_1(t) \) and \( \beta_i \) are as defined in Section 3, item (a).

Define \( \gamma_j \) as the probability that a patient suffers a headache of severity \( C_j \), \( j = 1, 2, 3 \). Let \( \gamma_j = \pi_1 + \ldots + \pi_j \). The null model in the continuation ratio component is then

\[ \log \left( \frac{\pi_j}{1 - \gamma_j} \right) = -\theta_j, \quad j = 1, 2, \]
Table I. Parameter estimates and corresponding standard errors in the combined proportional hazards and continuation ratio model

<table>
<thead>
<tr>
<th>Proportional hazards component</th>
<th>Continuation ratio component</th>
</tr>
</thead>
<tbody>
<tr>
<td>Parameter Estimate</td>
<td>Standard error</td>
</tr>
<tr>
<td>BETA2</td>
<td>-0.098</td>
</tr>
<tr>
<td>THETA2</td>
<td>-1.821</td>
</tr>
<tr>
<td>DELTA</td>
<td>-0.301</td>
</tr>
</tbody>
</table>

where \( \theta_j \) is the cut-point as described in Section 2. If log continuation ratio is not dependent on day of onset, then the composite model reduces to independent proportional hazards and continuation ratio models. However, in this example some evidence of a dependence does exist: \( \log(t) \) is deemed a reasonable form for \( f(t) \), and its effect approaches significance at the 10 per cent level, corresponding to a drop in \(-2(\text{log-likelihood})\) of 2.15. As a result, the composite model can be retained.

The probability that a patient suffers a headache of severity \( C_j \) on day \( t \) is denoted by \( \pi_j(t) \), and \( \gamma_j(t) = \pi_1(t) + \ldots + \pi_j(t) \). The continuation ratio component now becomes

$$
\log\left( \frac{\pi_j(t)}{1 - \gamma_j(t)} \right) = -\theta_j - \delta \log(t), \quad j = 1, 2, \quad t > 0,
$$

where \( \theta_j \) is as defined above and \( \delta \) is the effect of \( \log(t) \).

In fact, this is the final model in the continuation ratio component because all subsequent main effects and first-order interactions prove to be non-significant. The continuation ratio assumption is satisfied since the cut-point by schedule interaction is not significant.

Thus the joint hazard \( h_{ij}(t) \) can be expressed as

$$
h_{ij}(t) = h_i(t) \pi_j(t), \quad t > 0,
$$

where \( h_i(t) \) is modelled via (1), and where (2) is used to model \( \pi_j(t) \). The parameter estimates and corresponding standard errors associated with this joint hazard are given in Table I as they are presented in SAS. Notice that, in terms of the original notation, the SAS parameters INTERCEPT and THETA2 correspond to \( \theta_1 \) and \( \theta_2 - \theta_1 \) respectively. These estimates can be interpreted in the following manner. Patients on schedule \( B_2 \) started to suffer a headache almost as early as patients on schedule \( B_1 \). The estimate of \( h_2(t)/h_1(t) \) is 0.907, with 95 per cent confidence interval (0.467, 1.759). At the start of the study, patients were as likely to suffer a moderate headache as either a mild or a severe headache. By day 60, the probability of suffering a moderate headache had changed only slightly and had been overtaken by the chances of suffering a mild headache. Patients had only a 1 in 13 chance of starting to suffer a severe headache by this stage. The estimates of \( \pi_j(1), j = 1, 2, 3, \) are 0.213, 0.493 and 0.294 respectively, while the corresponding estimates of \( \pi_j(60) \) are 0.482, 0.441 and 0.077.

In other examples, \( \pi_j(t) \) may be dependent on treatment and \( h_{ij}(t) \) may depend on one or more prognostic factors. In such cases, a more unified summary of the results may be obtained by plotting \( h_{ij}(t; z)/h_{kj}(t; z) \) against \( j \) and \( t \) simultaneously, for treatments \( B_i, B_k \) \( (i \neq k) \) and vector of prognostic factors \( z \) fixed.
5. DISCUSSION

In this paper, continuation ratios, instead of the more familiar proportional odds, have been used to model the ordinal categories. An ordinal response will often have a definite baseline category, so the appropriate model should differentiate between the cases when the initial category is $C_1$ and when it is $C_2$, that is when the order of the categories has been reversed. The continuation ratio model satisfies this condition, whereas the proportional odds model does not. Also, the continuation ratio is mathematically more convenient than proportional odds for modelling the joint hazard $h_{ij}(t; z)$.

For simplicity, the example presented in Section 4 has been restricted to one prognostic factor, two treatments and three ordinal categories of response observed only at failure times. However, both components in the composite model can handle any number and type of prognostic factors and treatments. For instance, a set of qualitative treatments, such as dietary regimes, is suitably represented as a factor at the requisite number of levels, whereas quantitative treatments like dose levels could be fitted more appropriately as a continuous covariate. The analysis of more than three ordinal categories using the continuation ratio model is straightforward.

Responses observed at censoring times can be accommodated in the continuation ratio component, but the course of action depends on how time is represented in the continuation ratio model and on when censoring occurs. First consider time as being handled as a continuous covariate. If all censoring occurs at the end of the study, then only a binary indicator variable which distinguishes between censored and uncensored individuals has to be fitted in addition to the function of failure times $f(t)$. Otherwise, censoring times that occur at any point during the study can be fitted as a function of time which is distinct from $f(t)$.

Now consider the representation of time as a factor at discrete levels. If censoring occurs only at the end of the study, then the failure times of censored individuals can be thought of as lying in one time interval of effectively infinite length which starts when the study ends. This one interval can then be represented as an extra level of the factor used to represent the failure times, and can be fitted in the continuation ratio model accordingly. Otherwise, censoring times that occur throughout the study have to be grouped into intervals. These intervals are then fitted as levels of a factor in a separate continuation ratio model.

Throughout this paper, it has been assumed that response time, that is time to onset of response, and failure time are equivalent. This is reasonable in the cases where a response is 'fatal' and causes the failure to occur immediately, or is directly observable and defined as the failure itself, such as the headaches in the above example. However, situations often arise in which responses are neither fatal nor directly observable, but are only incidental. In animal carcinogenicity studies, for example, occult tumours are detected only when an animal has died. In such a case, time to failure, that is death, will not be an accurate reflection of time of response, which is the onset of tumour. Nevertheless, when the response times themselves are unknown, the only alternative is to adjust the continuation ratio model for failure times.

To conclude, this paper has presented a technique which provides an overall quantitative description of a process whereby individuals respond with an ordinal category of some kind at a particular point in time. To aid interpretation, the relationship between failure times and ordinal categories of response can be displayed graphically with a three-dimensional plot of the hazard ratio between any two treatment groups against time and category, for a fixed combination of prognostic factors.
Save the following statements:

```
%MACRO FAOR;
  %DO I = 1 %TO &N;
    DATA _DATA_; 
      SET ORIGIN;
      CUTPT = &I;
      RESP = 0;
      IF (CAT + 1) > CUTPT THEN RESP = 2;
      IF (CAT + 1) = CUTPT THEN RESP = 1;
      IF RESP = 0 THEN DELETE;
      IND = (RESP = 2);
      DROP RESP;
      PROC APPEND BASE = TRANSFO DATA = DATA&I;
  %END;
%MEND FAOR;
```

as an SAS macro in the file MACR SAS, say.

Save the following statements:

```
OPTIONS LEAVE = 25000;
%INCLUDE MACR;
%GLOBAL N;
DATA ORIGIN;
  SET PERM. ORIGIN;
  %LET N = 2;
  %FAOR
DATA TRANSFO;
  SET WORK. TRANSFO;
  THETA2 = (CUTPT = 2);
  PROC LOGIST DATA = TRANSFO K = 1;
  MODEL IND = THETA2;
ENDSAS;
```

as an SAS main program in the file MAIN SAS, say.

Running MAIN SAS will fix the number of cut-points at 2 in this case and then invoke macro FAOR, which will transform, for example, the following observation from the permanent SAS data set ORIGIN PERM (with categories $C_1$, $C_2$ and $C_3$ coded 0, 1 and 2 respectively to agree with the notation used in PROC LOGIST):

<table>
<thead>
<tr>
<th>Treatment (TRE)</th>
<th>Failure time</th>
<th>Ordinal category (CAT)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>1</td>
</tr>
</tbody>
</table>

into the following observations in temporary SAS data set TRANSFO WORK:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Failure time</th>
<th>CUTPT</th>
<th>IND</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>25</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>1</td>
<td>25</td>
<td>2</td>
<td>0</td>
</tr>
</tbody>
</table>
PROC LOGIST will fit the null continuation ratio model (with only the two cut-points: \( \theta_1 \) fitted as INTERCEPT, \( \theta_2 \) as INTERCEPT + THETA2) by performing binary logistic regression on data set TRANSFO, regarding IND as the response variable. To fit two treatments, say, in the model, add the statement

\[
\text{ALPHA2} = (\text{TRE} = 2);
\]

to the second DATA step and add ALPHA2 to the MODEL statement. Prognostic factors or more treatments will require the generation of more dummy variables, in the same way that more categories will require more cut-points to be generated. These new variables could be created automatically in a further SAS macro which could be included in MACR SAS, along with macro FAOR.

ACKNOWLEDGEMENTS

D. M. Berridge is supported by an Imperial Cancer Research Fund bursarship. The authors wish to thank Jack Cuzick, Peter Lee and Ian Major for their helpful comments on this work.

REFERENCES