STATISTICS IN MEDICINE
Statist. Med. (in press)
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2339

The sequential analysis of repeated binary responses: A score test for the case of three time points

Marina Roshini Sooriyarachchi[1], John Whitehead[2,*,†], Anne Whitehead[2] and Kim Bolland[2]

[1]*Department of Statistics and Computer Science, University of Colombo, Sri Lanka*
[2]*Medical and Pharmaceutical Statistics Research Unit, University of Reading, U.K.*

SUMMARY

In this paper a robust method is developed for the analysis of data consisting of repeated binary observations taken at up to three fixed time points on each subject. The primary objective is to compare outcomes at the last time point, using earlier observations to predict this for subjects with incomplete records. A score test is derived. The method is developed for application to sequential clinical trials, as at interim analyses there will be many incomplete records occurring in non-informative patterns. Motivation for the methodology comes from experience with clinical trials in stroke and head injury, and data from one such trial is used to illustrate the approach. Extensions to more than three time points and to allow for stratification are discussed. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: binary data; longitudinal data; repeated observations; score test; sequential clinical trial

1. INTRODUCTION

In many clinical trials the outcome of therapy cannot be assessed until some considerable time after randomization. For example, in trials in stroke a follow-up of three months is felt necessary before clinicians feel able to assess a definitive measure of recovery. The primary endpoint of such studies is usually a measure of functional status three months after treatment [1]. In trials in severe head injury the minimum follow-up time is usually taken to be six months, and the Glasgow Outcome Scale six months after treatment is used to assess the patient's condition [2]. Despite the length of follow-up felt necessary for individual patients, there is a need to monitor the accumulating results of studies in conditions as serious as stroke or head injury with a view to stopping recruitment as soon as it is clear that the

*Correspondence to: John Whitehead, MPS Research Unit, The University of Reading, P.O. Box 240, Earley Gate, Reading RG6 6FN, U.K.
†E-mail: j.r.whitehead@reading.ac.uk

Contract/grant sponsor: Wellcome Trust

Received 10 December 2004
Accepted 14 June 2005

experimental treatment either does or does not work. Sequential designs have been constructed for stroke trials based on ordinal measures of functional status at three months [3, 4]. Once the stopping criterion is reached, recruitment is stopped, but for three months (or in practice slightly longer) valid assessments of patients already randomized continue to be made. A final 'overrunning analysis' is performed when all of the data are finally in [5, 6]. The advantages of reduction of sample size and trial duration normally associated with sequential designs are reduced because of this delay. In trials in severe head injury, the longer six months follow-up means that sequential methods have little opportunity to reduce sample size. In the trial of eliprodil in severe head injury a sample size review and a safety monitoring procedure were used in conjunction with a fixed sample design in preference to taking a sequential approach [7].

This paper describes an approach in which intermediate responses from each patient are used in interim analyses in addition to the responses that are felt to be definitive. In stroke trials intermediate responses might come from assessments made at day 30 and day 60, prior to the definitive observation at day 90. In head injury the intermediate timings might be days 21 and 90, with the definitive response being observed at day 180. For the majority of patients, the outcome is set early on. For example, in the trial of eliprodil in severe head injury fewer than 50 per cent of patients changed their GOS between days 21 and 90, and fewer than 25 per cent changed between days 90 and 180 (Bolland, University of Reading PhD thesis, 2003). Although it is wise to set day 180 as the definitive time point, reasonably accurate forecasts of the day 180 outcome can be made at interim analyses on the basis of day 21 or 90 data from patients still undergoing follow-up. Similar considerations apply in stroke trials. The principle used in this paper is to make such forecasts, basing them only on the data from the trial itself, and dealing with the experimental and control groups separately. A score test will be derived directly from the likelihood, and it will be seen that the score statistic is of exactly the same form as that based only on the definitive data except that forecasts replace those definitive responses not yet observed. The score statistic is plotted against Fisher's information in order to conduct a sequential test, and the information statistic is reduced appropriately to allow for using forecast data rather than the real thing. This approach makes sequential designs in stroke more efficient and makes possible their use in head injury.

In this paper the case of binary outcomes assessed at three time points will be considered. Extension to more time points should be straightforward but complicated, while extension to ordinal data with more than two outcome categories might prove more challenging. In practice outcome scales used in stroke and head injury trials are often dichotomized, and in that case the methods developed here could be applied directly. However, we are not advocating dichotomization as it is an inefficient use of data. Instead, we see this work as a step towards methodology that will also deal with true ordinal outcomes. The settings of stroke and severe head injury will continue to be used as a background to the development of methods, although we are aware that the approach might be useful in other therapeutic areas such as Alzheimer's disease, multiple sclerosis or the assessment of quality of life following cancer treatment. The method derived reverts to a sequential Pearson χ^2-test [8, Section 3.2] once the data are complete for every patient.

In the next section, we will survey existing approaches that could be used to incorporate intermediate responses into interim analyses. In Section 3 the principal example of the paper, a trial in head injury, will be introduced. Section 4 reviews the sequential approach that is adopted here, as it would be applied to the definitive outcomes only, and in Section 5 the

extensions allowing incorporation of intermediate values are developed. These are applied to the data in Section 6, and Section 7 is a discussion of the potential and limitations of the new approach.

2. EXISTING APPROACHES TO THE INCORPORATION OF INTERMEDIATE OUTCOMES

A generally accepted approach to the sequential analysis of repeated binary or ordinal data from clinical trials involves the use of longitudinal models. Gange and DeMets [9] suggest the use of marginal models fitted at each interim analysis using generalized estimating equations, while Spiessens *et al*. [10] used the subject-specific model of Hedeker and Gibbons [11, 12] for their interim analyses. The sequential software package EaSt includes the sequential analysis of longitudinal models as a standard option. However, in her University of Reading PhD thesis of 2003, Bolland found problems of lack-of-fit when the subject-specific model was applied to stroke and head injury data, and both she and Spiessens *et al*. [10] have reported serious difficulties in fitting the model. In short, small changes in the number of quadrature points used in the numerical integrations can have substantial effects on the estimates of parameters. The marginal model, on the other hand, suffers from serious difficulties in interpretation [13].

A more fundamental problem in using either of the longitudinal approaches to data analysis concerns the nature of the model itself. The hypothesis being tested in these models is not whether there is an advantage in outcome at day 90 in a stroke trial or day 180 in a serious head injury trial. Instead, the question is whether there is some consistent advantage over the whole follow-up period, interpreted as a difference in intercept or in slope or in both, in some linear model. In a sense, an average benefit over the follow-up period is being sought. A treatment that achieves benefit sooner, but not necessarily in a greater number of patients, would come out as successful in such a comparison (although a subsidiary analysis might show lack-of-fit). In early interim analyses in a sequential setting, the majority of the data available would consist of intermediate patient assessments, and these would be extrapolated according to the linear model. Thus better outcomes at day 21 in a head injury trial (say) might dominate, and appear to imply a consistent advantage. The trial might then be stopped, with insufficient data to check the model assumptions, and no prospect of such data being collected. The severity of this problem is greatest when the follow-up time is long and recruitment is slow.

In the approach of this paper, no model linking the successive outcomes will be assumed. Instead, the transitions between them will be estimated from the trial data themselves. Separate estimation within the two treatment groups will ensure maximum robustness. The underlying assumption is that it is of little concern how quickly a good recovery was achieved, the important thing is to achieve it by the time of the definitive observation. Thus the method addresses a question that is fundamentally different from that addressed by the longitudinal approach, and one that in many cases is more appropriate.

The method described by Marschner and Becker [14] is much closer to the approach described here. They too seek a robust test, based on minimal assumptions, for comparing the definitive outcome probabilities. The differences between their approach and ours are matters of detail. They parameterize treatment advantage in terms of a probability difference, whereas we use a log-odds ratio. They use a Wald test, whereas we use a score test. Their paper deals

only with a single intermediate outcome, although it could be generalized to the case of two or more. Wald tests can be derived for either parameterization and use maximum likelihood estimates derived under the alternative hypothesis which are easier to compute than the restricted maximum likelihood estimates required for the score test. However, both approaches involve considerable manipulation in order to derive and evaluate the second derivatives required to find the variance of the test statistic. Nearly all of the material in the appendices is required to derive a sequential Wald test for this case, although we shall not follow that through in this paper. Instead, we pursue the score test because there are more situations in which the Wald test breaks down due to zero counts than is the case for the score test. This important issue is not discussed in Reference [14].

3. EXAMPLE: A TRIAL IN HEAD INJURY

The work described in this paper has been motivated by experience with neurological trials, and in particular with trials of treatments for stroke and head injury. In this section the conduct and results of one particular trial in head injury are described. Unfortunately, no clinical account of this trial has been published.

Between 1993 and 1996, 452 patients who had suffered severe head injury were randomized in a placebo-controlled evaluation of the non-competitive N-Methyl-D-Aspartate receptor antagonist, eliprodil. Administration of treatment, intravenously at first and then by oral or naso-gastric tube, lasted for 20 days. The primary objective of the trial was to assess efficacy in terms of improvement of functional status after six months of follow-up. The primary efficacy criterion was the score on the Glasgow Outcome Scale (GOS) at 6 months after randomization. This scale comprises five ordered response categories: good recovery (GR), moderate disability (MD), severe disability (SD), vegetative state (V) and dead (D). For the primary efficacy analysis the worst three categories of the GOS were combined into a single category, and the resulting trichotomy was modelled as ordered categorical data using the proportional odds regression model [15].

Data from previous studies suggested that the percentage of placebo patients with a 6 month GOS in the GR and MD categories would be 17 and 30 per cent, respectively. For the conclusion of the trial to be clinically relevant it was necessary for the proportion of patients having GR or MD to be increased from 47 per cent on placebo to 62 per cent on eliprodil. This corresponds to a log-odds ratio of 0.61 with, respectively, 27.4 and 34.6 per cent of eliprodil patients in the GR and MD categories. A power of 90 per cent was set to attain significance at the 5 per cent level (two-sided alternative) for this magnitude of treatment advantage. Using the sample size formula of Whitehead [16] it was calculated that 400 patients would be required. A planned sample size review was conducted on outcome data from 93 patients and as a result the sample size was recalculated, stratifying for Glasgow Coma Scale (GCS) score at day 0 and increased to 450 [17]. The GCS is used as a baseline assessment of injury severity, the Glasgow Outcome Scale (GOS) cannot be assessed in an emergency situation.

The primary analysis was on data from 229 patients on eliprodil and 223 on placebo. In the proportional odds regression analysis adjusting for age, GCS score and geographical region, the effect of treatment was found to be non-significant ($p = 0.310$). The estimate of the odds ratio on better outcomes was 1.219 for eliprodil relative to placebo with a 95 per cent

confidence interval (0.832, 1.785). In this paper, we shall show how patient outcomes at 21 and 90 days can be used with those at 180 days in a sequential reconstruction of the trial. In doing so, we will sacrifice the trichotomous outcome and reduce to binary responses representing good recovery: yes or no. For the binary response the full data give $p = 0.799$, an estimated odds ratio of 1.052 and a 95 per cent confidence interval of (0.715,1.547). In terms of log-odds ratios, for direct comparison with the results presented in Section 6, these estimates are 0.051 and $(-0.335, 0.436)$, respectively.

4. SEQUENTIAL THEORY

The specification of a sequential clinical trial design is made in terms of a scalar parameter (θ) quantifying the advantage of the experimental treatment (T_1) over the control (T_2). The hypothesis $H_0 : \theta = 0$ can be tested using a plot of the score statistic Z against its null variance V [8]. If $\theta = \theta_R > 0$, then the power to achieve significance at the two-sided level α is required to be $(1 - \beta)$. A fixed sample design is found setting V equal to $V_{\text{fix}} = \{(w_{\alpha/2} + w_\beta)/\theta_R\}^2$, where w_γ is the upper 100γ percentage point of the standard normal distribution function. Within any family of sequential designs there will be a unique design satisfying the power requirement. A series of interim analyses is conducted at times following some pre-arranged plan, at the ith of which the current values Z_i and V_i of Z and V will be compared with stopping bounds ℓ_i and u_i deduced from the design parameters and the values of $V_1, \ldots, V_i$. The trial continues if $Z_i \in (\ell_i, u_i)$ and stops otherwise. Usually the $Z_i \geqslant u_i$ corresponds to significant evidence that T_1 is better than T_2, while $Z_i \leqslant \ell_i$ indicates either significant evidence that T_1 is worse than T_2 or no significant difference. Most of the systems for creating sequential and group sequential designs, including the α-spending approach [18] and stochastic curtailment [19] can be fitted into this framework. The final analysis should allow for the interim analyses [8, 18].

Here score statistics for the sequential analysis of repeated binary responses will be derived that can be used within any sequential framework: illustrations here will be based on the triangular test. They can also be used in fixed sample tests to allow for missing observations, although assumptions concerning the mechanism resulting in data being missing may be harder to justify.

For a single binary response, probabilities of success on T_1 and T_2 can be denoted by p_1 and p_2, respectively. The advantage of E over C can be parameterized by the log-odds ratio $\theta = \log[\{p_1(1 - p_2)\}/\{p_2(1 - p_1)\}]$. When n_k patients have been treated on T_k, of whom S_k have succeeded and F_k have failed $(k = 1, 2)$, $Z = (n_2 S_1 - n_1 S_2)/n$ and $V = n_1 n_2 S_1 S_2 / n^3$, where $S = S_1 + S_2, F = F_1 + F_2$ and $n = n_1 + n_2$.

5. INCORPORATION OF INTERMEDIATE DATA

Suppose that the progress of patients is observed at three separate time points, t_1, t_2 and t_3, the schedule being the same for everyone. At each time point they are classified as being in category 1 or 2, with category 1 being more desirable. The objective of the study is to assess whether there is a difference between the treatments in terms of the probability of being in category 1 at time t_3. It will be assumed that patients' records are complete up to their last

observation. If a patient is in category i at time t_1, category j at time t_2 and category k at time t_3, then their outcome will be denoted by (i,j,k), where i, j and k can take the values 1, 2 or $*$ (denoting missing).

The number of patients on T_g with outcome (i,j,k) at the interim analysis in question will be denoted by $n_{ijk,g}$, and the probability of such an outcome by $p_{ijk,g}$, $g=1, 2$. The symbol $\circ$ will be used to denote a total over the subscripts 1 and 2, so that $p_{ij\circ,g}=p_{ij1,g}+p_{ij2,g}$ and so on. The term $N_{ijk,g}$ will denote the number of patients on T_g who will eventually have outcome (i,j,k) when all assessments have been completed. As the records might never be complete, the values $N_{ijk,g}$ might remain latent observations. The pattern of incomplete and complete data available when the analysis is actually conducted will be denoted by Π. The predicted value of $N_{ijk,g}$, $e_{ijk,g}=E(N_{ijk,g}\,|\,\Pi)$ can be expressed as

$$e_{ijk,g}=\left(\frac{n_{ijk,g}}{p_{ijk,g}}+\frac{n_{ij*,g}}{p_{ij\circ,g}}+\frac{n_{i**,g}}{p_{i\circ\circ,g}}\right)p_{ijk,g}$$

The parameter of interest, θ, is the log-odds ratio for success at the third time point,

$$\theta=\log\{p_{\circ\circ 1,1}/(1-p_{\circ\circ 1,1})\}-\log\{p_{\circ\circ 1,2}/(1-p_{\circ\circ 1,2})\}$$

It will be shown below that the score statistic takes the form

$$Z=(n_{\circ\bullet\bullet,2}\tilde{e}_{\circ\circ 1,1}-n_{\circ\bullet\bullet,1}\tilde{e}_{\circ\circ 1,2})/n_{\circ\bullet\bullet,\circ} \tag{1}$$

where the $\tilde{e}_{ijk,g}$ are functions of restricted maximum likelihood estimates $\tilde{p}_{ijk,g}$ found under the assumption that $\theta=0$ and $\bullet$ denotes a summation over 1, 2 and $*$. When the data are complete, and all patients provide responses at all three time points, Z will take the form described in Section 4. Derivation of the form of the statistic V is deferred to the Appendix.

In order to derive equation (1), we start with the log-likelihood, ℓ, given by

$$\ell=\sum_{g=1}^{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}n_{ijk,g}\log p_{ijk,g}+\sum_{g=1}^{2}\sum_{i=1}^{2}\sum_{j=1}^{2}n_{ij*,g}\log p_{ij\circ,g}+\sum_{g=1}^{2}\sum_{i=1}^{2}n_{i**,g}\log p_{i\circ\circ,g} \tag{2}$$

'Forwards' conditional probabilities $q_{i,g}^{(1)}=P(\text{category } i \text{ at } t_1;\ T_g)$, $q_{ij,g}^{(2)}=P(\text{category } j \text{ at } t_2\,|\,\text{category } i \text{ at } t_1;\ T_g)$ and $q_{ijk,g}^{(3)}=P(\text{category } k \text{ at } t_3\,|\,\text{categories } i \text{ at } t_1 \text{ and } j \text{ at } t_2;\ T_g)$ can be defined, and it follows that

$$p_{ijk,g}=q_{ijk,g}^{(3)}q_{ij,g}^{(2)}q_{i,g}^{(1)};\quad i,j,k,g=1,2 \tag{3}$$

Expressed in terms of the q's, the log likelihood becomes

$$\ell=\sum_{g=1}^{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2}n_{ijk,g}\log q_{ijk,g}^{(3)}+\sum_{g=1}^{2}\sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{*}n_{ijk,g}\log q_{ij,g}^{(2)}+\sum_{g=1}^{2}\sum_{i=1}^{2}\sum_{j=1}^{*}\sum_{k=1}^{*}n_{ijk,g}\log q_{i,g}^{(1)} \tag{4}$$

where certain sums are over 1, 2 and $*$ (missing). The maximum likelihood estimates of the q's are $\hat{q}_{i,g}^{(1)}=n_{i\bullet\bullet,g}/n_{\circ\bullet\bullet,g}$, $\hat{q}_{ij,g}^{(2)}=n_{ij\bullet,g}/n_{i\circ\bullet,g}$ and $\hat{q}_{ijk,g}^{(3)}=n_{ijk,g}/n_{ij\circ,g}$. The predicted counts, $e_{ijk,g}$, can be expressed in terms of the q's as

$$e_{ijk,g}=(n_{ijk,g}+n_{ij*,g}q_{ijk,g}^{(3)}+n_{i**,g}q_{ij,g}^{(2)}q_{ijk,g}^{(3)}) \tag{5}$$

To derive the score test, restricted maximum likelihood estimates under the null hypothesis are required. It is convenient to introduce the 'backwards' conditional probabilities $r_{k,g}^{(1)}=$

$P(\text{category } k \text{ at } t_3;\ T_g)$, $r^{(2)}_{jk,g} = P(\text{category } j \text{ at } t_2 \mid \text{category } k \text{ at } t_3;\ T_g)$ and $r^{(3)}_{ijk,g} = P(\text{category } i \text{ at } t_1 \mid \text{categories } j \text{ at } t_2 \text{ and } k \text{ at } t_3;\ T_g)$. It follows that

$$p_{ijk,g} = r^{(3)}_{ijk,g}\, r^{(2)}_{jk,g}\, r^{(1)}_{k,g}; \quad i,j,k,g = 1,2 \tag{6}$$

The parameters of interest will be $r^{(1)}_{1,g}$, $r^{(2)}_{1k,g}$ and $r^{(3)}_{1jk,g}$. The vectors with entries consisting of all of the $e_{ijk,g}$ and $N_{ijk,g}$ values, arranged with (ijk) running through $(111),(112),\ldots,(222)$, will be denoted by e_g and N_g, respectively, $g = 1, 2$, and their concatenations over both treatment groups will be denoted by e and N.

The matrix $R_{h,g}$ is defined as $\mathrm{diag}(r^{(3)}_{h11,g}, r^{(3)}_{h12,g}, r^{(3)}_{h21,g}, r^{(3)}_{h22,g}, r^{(2)}_{h1,g}, r^{(2)}_{h2,g}, r^{(1)}_{h,g})$, $h = 1, 2$, and $R_h = \mathrm{diag}(R_{h,1}, R_{h,2})$. The matrices A_1 and A_2 are defined by

$$A_1 = \begin{pmatrix} 1&0&0&0&0&0&0&0\\ 0&1&0&0&0&0&0&0\\ 0&0&1&0&0&0&0&0\\ 0&0&0&1&0&0&0&0\\ 1&0&0&0&1&0&0&0\\ 0&1&0&0&0&1&0&0\\ 1&0&1&0&1&0&1&0 \end{pmatrix} \quad \text{and} \quad A_2 = \begin{pmatrix} 0&0&0&0&1&0&0&0\\ 0&0&0&0&0&1&0&0\\ 0&0&0&0&0&0&1&0\\ 0&0&0&0&0&0&0&1\\ 0&0&1&0&0&0&1&0\\ 0&0&0&1&0&0&0&1\\ 0&1&0&1&0&1&0&1 \end{pmatrix}$$

so that $A_h e_g = (e_{h11,g}\ \ e_{h12,g}\ \ e_{h21,g}\ \ e_{h22,g}\ \ e_{\circ h1,g}\ \ e_{\circ h2,g}\ \ e_{\circ\circ h,g})'$, and similarly for $A_h N_g$; $h,\ g = 1,2$.

Now, let U_g denote the efficient score vector for treatment group T_g

$$U_g = (\partial\ell/\partial r^{(3)}_{111,g} \quad \partial\ell/\partial r^{(3)}_{112,g} \quad \partial\ell/\partial r^{(3)}_{121,g} \quad \partial\ell/\partial r^{(3)}_{122,g} \quad \partial\ell/\partial r^{(2)}_{11,g} \quad \partial\ell/\partial r^{(2)}_{12,g} \quad \partial\ell/\partial r^{(1)}_{1,g})$$

$g = 1,2$. The derivatives of ℓ with respect to the r's are best found as sums of products of derivatives of ℓ with respect to the p's and derivatives of the p's with respect to the r's. The results can be concisely expressed in terms of e_g, which is in turn a function of the q's, and so we obtain $U_g = (R^{-1}_{1,g}A_1 - R^{-1}_{2,g}A_2)e_g$, $g = 1,2$. As $e_g = E(N_g \mid \Pi)$, it follows that $U_g = E\{(R^{-1}_{1,g}A_1 - R^{-1}_{2,g}A_2)N_g \mid \Pi\}$, $g = 1,2$. In particular, the derivatives with respect to $r^{(1)}_{1,g}$ take the form

$$\frac{\partial\ell}{\partial r^{(1)}_{1,g}} = \sum_{k=1}^{2} (-1)^{k-1} \frac{e_{\circ\circ k,g}}{r^{(1)}_{k,g}}, \quad g = 1,2$$

With these derivatives available, we can proceed to derive the score statistic for the parameter θ of interest, which is the log-odds ratio for success at the third time point, $\theta = \log\{r^{(1)}_{1,1}/(1 - r^{(1)}_{1,1})\} - \log\{r^{(1)}_{1,2}/(1 - r^{(1)}_{1,2})\}$. A nuisance parameter ϕ can be defined by $\phi = \log\{r^{(1)}_{1,1}/(1 - r^{(1)}_{1,1})\} + \log\{r^{(1)}_{1,2}/(1 - r^{(1)}_{1,2})\}$. It follows that

$$\frac{\partial\ell}{\partial\theta} = \sum_{g=1}^{2} \frac{\partial\ell}{\partial r^{(1)}_{1,g}} \frac{\partial r^{(1)}_{1,g}}{\partial\theta} \quad \text{and} \quad \frac{\partial\ell}{\partial\phi} = \sum_{g=1}^{2} \frac{\partial\ell}{\partial r^{(1)}_{1,g}} \frac{\partial r^{(1)}_{1,g}}{\partial\phi}$$

As $\partial r^{(1)}_{1,g}/\partial\theta = (-1)^{g-1} r^{(1)}_{1,g} r^{(1)}_{2,g}/2$ and $\partial r^{(1)}_{1,g}/\partial\phi = r^{(1)}_{1,g} r^{(1)}_{2,g}/2$, $g = 1, 2$

$$\frac{\partial \ell}{\partial \theta} = \frac{1}{2}\{e_{\circ\circ 1,1} r^{(1)}_{2,1} - e_{\circ\circ 2,1} r^{(1)}_{1,1} - e_{\circ\circ 1,2} r^{(1)}_{2,2} + e_{\circ\circ 2,2} r^{(1)}_{1,2}\}$$

and

$$\frac{\partial \ell}{\partial \phi} = \frac{1}{2}\{e_{\circ\circ 1,1} r^{(1)}_{2,1} - e_{\circ\circ 2,1} r^{(1)}_{1,1} + e_{\circ\circ 1,2} r^{(1)}_{2,2} - e_{\circ\circ 2,2} r^{(1)}_{1,2}\}$$

Under the null hypothesis $r^{(1)}_{1,1} = r^{(1)}_{1,2}$. Putting $\partial\ell/\partial\phi = 0$ and $\theta = 0$ gives $\tilde{r}^{(1)}_{1,1} = \tilde{r}^{(1)}_{1,2} = \tilde{e}_{\circ\circ 1,\circ}/n_{\circ\bullet\bullet,\circ}$ as the common restricted maximum likelihood estimate, where $\tilde{e}_{\circ\circ 1,\circ} = \tilde{e}_{\circ\circ 1,1} + \tilde{e}_{\circ\circ 1,2}$ and so on. This is the null expected number of successes, divided by the total number of patients: a very intuitive result. Thus the score statistic takes the form presented in (1) above.

To compute Z, restricted maximum likelihood estimates of all parameters are required, and these must satisfy $\partial\ell/\partial r^{(3)}_{1jk,g} = 0$ and $\partial\ell/\partial r^{(2)}_{1k,g} = 0$, so that $\tilde{r}^{(3)}_{1jk,g} = \tilde{e}_{1jk,g}/\tilde{e}_{\circ jk,g}$ and $\tilde{r}^{(2)}_{1k,g} = \tilde{e}_{\circ 1k,g}/\tilde{e}_{\circ\circ k,g}$ for $g, j, k = 1, 2$. The following iterative scheme is used: (i) the q's are estimated using their unrestricted maximum likelihood estimates, (ii) the e's are deduced from equation (5), (iii) the r's are then found using the restricted maximum likelihood equations above, (iv) the e's are then found from the r's using equations (6), (3) and (5) in turn, and steps (iii) and (iv) are iterated to a solution. In the analyses for this paper we have found this scheme to converge within eight iterations.

Some of the counts used in the calculations may be equal to zero, but in most cases this will cause no difficulty. Zeros need careful consideration if they lead to estimates of q's or of r's being indeterminate. First consider indeterminate q's, and in particular the case $n_{ij\circ,g} = 0$ for some i, j and g. In that case straightforward use of the ratios following equation (4) leads to estimates $\hat{q}^{(3)}_{ij1,g}$ and $\hat{q}^{(3)}_{ij2,g}$ that are both indeterminate. In fact, examination of the log-likelihood given in equation (4) shows that the parameters $q^{(3)}_{ij1,g}$ and $q^{(3)}_{ij2,g}$ both disappear. No patients on T_g have been observed to move from category i to j and then on to either 1 or 2, and thus the final status of the $n_{ij*,g}$ patients with incomplete records beginning with i and j cannot be predicted. Now the value of $e_{ijk,g}$ is given by $n_{ijk,g} + n_{ij*,g} q^{(3)}_{ijk,g} + n_{i**,g} q^{(2)}_{ij,g} q^{(3)}_{ijk,g}$, but in this case, the second term cannot be used for $k = 1$ or 2. Furthermore, the third term cannot be used, as even if moves from category i to j can be predicted, the subsequent move to 1 or 2 cannot. In the algorithm, this is achieved automatically by setting both $\hat{q}^{(3)}_{ij1,g}$ and $\hat{q}^{(3)}_{ij2,g}$ to 0, that is: interpreting 0/0 as 0. If $n_{i\circ\bullet,g} = 0$, then $\hat{q}^{(2)}_{i1,g}$ and $\hat{q}^{(2)}_{i2,g}$ should be set to 0 for similar reasons, although they will be removed from the e's anyway due to their multiplication by $\hat{q}^{(3)}_{ijk,g}$ terms that will logically also be set to 0 in these circumstances. Once a q term has been set to 0 in the first step of the iterative scheme, it will remain equal to 0 in subsequent steps. If $n_{\circ\bullet\bullet,g} = 0$, then $\hat{q}^{(1)}_{1,g}$ and $\hat{q}^{(1)}_{2,g}$ are indeterminate, but in this circumstance there are no patients at all on treatment T_g and so the whole method is bound to fail.

A different set of zero problems concern the r's. If $n_{\circ jk,g}$ is zero, then the estimated probabilities of reaching category k from any starting category followed by a move to j will be 0, that is $\hat{q}^{(3)}_{ijk,g} = 0$ for $i = 1, 2$. Equation (5) shows that $\tilde{e}_{\circ jk,g}$ will be equal to 0 too. This means that $\tilde{r}^{(3)}_{ijk,g} = \tilde{e}_{ijk,g}/\tilde{e}_{\circ jk,g}$ is indeterminate for $i = 1, 2$. In the first sum of the log-likelihood given in equation (2) the terms $p_{ijk,g}$ $(i = 1, 2)$ will disappear. These same terms also appear implicitly in the terms $p_{ij\circ,g}$ and $p_{i\circ\circ,g}$ $(i = 1, 2)$ that appear in the other two sums. The

log-likelihood will be maximized when the $p_{ijk,g}$ $(i=1,2)$ are equal to zero, and so these terms can be set equal to zero from the outset. Now, derivatives of ℓ with respect to the r's are found as sums of products of derivatives of ℓ with respect to the p's and derivatives of the p's with respect to the r's. The p's are related to the r's through $p_{ijk,g}=r^{(3)}_{ijk,g}\,r^{(2)}_{jk,g}\,r^{(1)}_{k,g}$, and all of the p's relating to $\tilde{r}^{(3)}_{1jk,g}$ and $\tilde{r}^{(3)}_{2jk,g}$ have been removed so that $\partial\ell/\partial r^{(3)}_{1jk,g}=0$. It is appropriate to set $\tilde{r}^{(3)}_{1jk,g}$ and $\tilde{r}^{(3)}_{2jk,g}$ equal to 0 in this case: again in the algorithm we force $0/0=0$. This in turn makes $\tilde{p}_{ijk,g}$ and $\tilde{q}^{(3)}_{ijk,g}$ equal to 0 for $i=1,2$, and has the desired effect of making the final estimate $\tilde{e}_{\circ jk,g}$ equal to zero. This amounts to foreseeing that when the data are complete there will still be no patients with outcomes j and k at the second and third time points, respectively.

If $n_{\circ\circ k,g}$ is zero, then the estimated probabilities of reaching category k from any of the earlier states will be zero, that is $\hat{q}^{(3)}_{ijk,g}=0$ for all i, j. An argument similar to that in the previous paragraph shows that it is appropriate to set $\tilde{r}^{(2)}_{1k,g}$ and $\tilde{r}^{(2)}_{2k,g}$ equal to 0 in this case making the final estimate $\tilde{e}_{\circ\circ k,g}$ equal to zero. The score statistic can still be calculated under these circumstances. If both $\tilde{e}_{\circ\circ k,1}$ and $\tilde{e}_{\circ\circ k,2}$ become equal to 0, then Z will exist with value zero, unless $(n_{\circ\bullet\bullet,1}+n_{\circ\bullet\bullet,2})=0$, in which case Z will be indeterminate and the test will fail.

6. APPLICATION TO THE HEAD INJURY DATA

In the head injury trial described in Section 3, the GOS score was assessed at days 21, 90 and 180, and in this illustration the score has been dichotomized with success being GR. The first patient was recruited on 30 April 1993. Suppose that the first interim analysis takes place one year later on 30 April 1994 and that subsequent interim analyses are planned every nine months. A design was created using PEST 4 [20] as if for a standard binary response with a power of 0.90 to detect as significant at the 5 per cent level (two-sided) an improvement in the good recovery rate at 180 days from $p_{\circ\circ 1,2}=0.17$ on placebo to $p_{\circ\circ 1,1}=0.25$ on eliprodil. The corresponding log-odds ratio is 0.487, so that this design is more powerful than that actually used. For this power requirement, a fixed sample design would require information corresponding to $V_{\text{fix}}=44.31$. The triangular test satisfying these requirements has upper boundary $Z=10.129+0.148V$ and lower boundary $Z=-10.129+0.444V$ (see Figure 1).

The data at each of the first three interim analyses and the final analysis are given in Table I. Table II presents two pairs of test statistics for each interim data set, and these are plotted together with the triangular stopping region in Figure 1. The first is computed using the repeated binary method introduced in this paper. The second pair of test statistics (denoted with the subscript C) are calculated using only data from patients with responses at 180 days. The figure also shows an internal jagged boundary known as the 'Christmas tree correction' which allows for the amount of information accrued since the previous interim analysis. It is sufficient to reach this less stringent criterion in order to stop [8]. Final analyses for the two methods are presented in Table III.

At each interim analysis there is more information available (a higher value of V) in the repeated binary analysis than in the analysis that is restricted to patients who have completed 180 days in the study. The increases in V are 26, 37, 24 and 7 per cent, respectively, at the

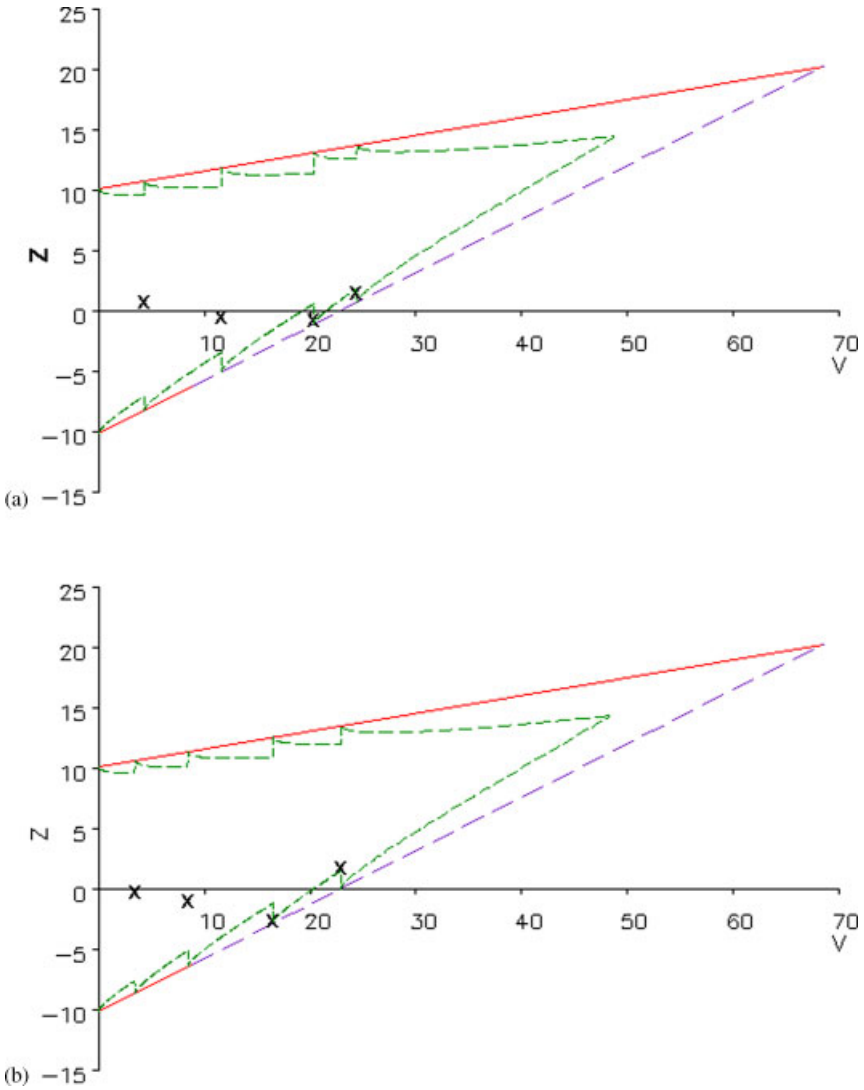

Figure 1. The completed sequential trial: (a) based on complete and partial data from all subjects; and (b) based on complete data only.

four analyses. These repeated binary analyses take account of 80, 68, 48 and 10 per cent more patients than in those based only on completers: this illustrates the way in which partial data from incomplete patients contribute. The data presented in this example are from a real clinical trial, although they are extreme in that no-one who has a good recovery at day 21 ever fails to have that status at day 180. This means that the partial data become very informative. At the first interim, the eliprodil group contains a larger proportion than the placebo group of patients with partial follow-up who are currently in state 1 (8/22 *versus* 5/22). As these patients are likely to remain in state 1, their inclusion turns an apparent (small) disadvantage

Table I. Data from the trial of eliprodil in head injury: each row concerns a different pattern of responses at days 21, 90 and 180, with outcome 1 denoting a good recovery at this time and outcome 2 indicating that the patient is dead or has a poor recovery.

	30 April 1994			31 January 1995			31 October 1995			30 April 1996		
	Eliprodil	Placebo	Total	Eliprodil	Placebo	Total	Eliprodil	Placebo	Total	Eliprodil	Placebo	Total
(1,1,1)	4	3	7	9	7	16	23	18	41	36	30	66
(2,1,1)	6	6	12	14	18	32	29	37	66	46	44	90
(1,2,1)	0	0	0	0	1	1	1	1	2	1	1	2
(2,2,1)	4	4	8	5	8	13	7	12	19	11	18	29
(1,1,2)	0	0	0	0	0	0	0	0	0	0	0	0
(2,1,2)	0	0	0	0	0	0	0	0	0	0	0	0
(1,2,2)	0	0	0	0	0	0	0	0	0	0	0	0
(2,2,2)	15	13	28	36	39	75	69	67	136	87	93	180
(1,1,∗)	2	1	3	8	3	11	8	8	16	2	3	5
(2,1,∗)	3	2	5	7	9	16	10	6	16	6	5	11
(1,2,∗)	0	0	0	0	0	0	0	0	0	0	0	0
(2,2,∗)	7	12	19	12	11	23	11	16	27	7	3	10
(1,∗,∗)	3	2	5	7	6	13	11	6	17	3	0	3
(2,∗,∗)	7	5	12	15	15	30	31	21	52	5	3	8
Total	51	48	99	113	117	230	200	192	392	204	200	404

Table II. Values of Z and V for two methods.

	Repeated binary		Complete binary	
Date	Z	V	Z_C	V_C
30-04-94	0.716	4.300	−0.236	3.426
31-01-95	−0.528	11.611	−0.964	8.449
31-10-95	−0.702	20.361	−2.546	16.476
30-04-96	1.456	24.431	1.774	22.925

Table III. Final analyses for the two methods for Example 1.

	p	θ_M	(θ_L, θ_U)
Repeated binary	0.755	0.063	$(-0.335, 0.465)$
Complete binary	0.709	0.078	$(-0.332, 0.488)$

of eliprodil, based on complete patient records alone, to an apparent (small) advantage for the drug.

In both analyses, the lower boundary is reached at the third interim analysis. Recruitment then stops, and the final 'overrunning' analysis is completed six months later. In principle, all data records should be complete by this date, but in the real data on which this example is based some missing data persist and so the repeated binary analysis never quite matches the completers only analysis. The overrunning analysis shows a small rally in the fortunes of the eliprodil patients. In the repeated binary analysis, the final point remains in the stopping region, but in the completers only analysis the continuation region is re-entered. The issue of re-entering the continuation region in an overrunning analysis is discussed at length in Reference [6]. The effect is less in the repeated binary analysis because a greater proportion of the final information is accounted for at the interim analysis leading to stopping. The final analyses reported in Table III are similar, with the repeated binary method yielding the narrower 95 per cent confidence interval for θ.

7. DISCUSSION

For sequential clinical trials in therapeutic areas such as stroke and head injury, the method derived in this paper has the advantage of making the maximum use of data available at interim analyses without the imposition of strong assumptions that may not be verifiable. The pattern of the results derived holds true for the case of two repeated responses and is likely to extend to more than three.

It is straightforward to apply the method of this paper to stratified data following the approach of Chapter 7.2 of [8]. The statistics Z and V are computed separately within each stratum, and then the Z's are summed to form the score statistic for plotting on the sequential scheme, and the V's are summed to form the corresponding information measure. Covariate adjustment via linear modelling would be a greater challenge.

It is assumed that data missing from the incomplete data sets can be treated as 'missing at random', as defined by Rubin [21]. Provided that the patients recruited are homogeneous over time this will be true. In practice, some of the incomplete data records might never be completed, and such long-term missing values may not be missing at random. If there is a substantial proportion of such records, then more elaborate approaches might become necessary. If the method is used in a non-sequential setting, the assumption that data are missing at random might be difficult to justify, although the approach would certainly improve on an analysis based on simple methods such as last-observation-carried-forward.

In the head injury data analysed in Section 6, all data records included the day 21 observation, but some lacked the day 90 record and yet included that at day 180. The method described does not deal with such records, and in the illustrative analysis they were excluded. As patients were mostly still in hospital on day 21, and as the outcome on day 180 was the primary endpoint for the trial, compliance in reporting these scores was good. If the day 90 score had been identified as important in the analysis, then it is likely that fewer of these scores would have been missing. It is possible to add a term to the log-likelihood used here to incorporate interrupted observation records: the transition probabilities would be combinations of those already used in the model. Such an approach would be worthwhile if the proportion of interrupted records was non-negligible, although it would be more complicated and would rest on the missing at random assumption.

The score test formulation with a log-odds ratio parameterization has been adopted here because, once the data are complete, it reverts to familiar statistics consistent with Pearson's χ^2-test. Furthermore, with complete data, the score test tends to be more accurate than equivalent Wald tests. The approach taken by Marschner and Becker [14] to the same problem, but with only two time points, is based on the Wald test for a probability difference. In a subsequent paper, we will present simulation results comparing the score and Wald approaches and the log-odds ratio and probability difference parameterizations for the two time point case.

APPENDIX A: SECOND DERIVATIVES OF THE LOG-LIKELIHOOD

Second derivatives of ℓ with respect to the r's will now be presented, but first some additional notation is required. Let

$$c_{i,(jk),(j'k'),g} = (n_{ij*,g}q^{(3)}_{ijk,g} + n_{i**,g}q^{(2)}_{ij,g}q^{(3)}_{ijk,g})\delta_{jj'}\delta_{kk'} - n_{ij*,g}q^{(3)}_{ijk,g}q^{(3)}_{ijk',g}\delta_{jj'} - n_{i**,g}q^{(2)}_{ij,g}q^{(2)}_{ij',g}q^{(3)}_{ijk,g}q^{(3)}_{ij'k'g}$$

where $\delta_{jj'} = 1$ if $j = j'$ and 0 otherwise. The interpretation of these quantities is that $c_{i,(jk),(j'k'),g} = \mathrm{cov}(N_{ijk,g}, N_{ij'k',g} \mid \Pi)$. The matrix of all $c_{i,(jk),(j'k'),g}$, with (ijk) running down the rows and $(i'j'k')$ across the columns in the order $(111),(112),\ldots,(222)$, will be denoted by C_g, $g = 1,2$; any entry with $i \neq i'$ is taken to have value zero. The matrix C will be set to $\mathrm{diag}(C_1, C_2)$. Now let H_g denote the Hessian matrix of second derivatives of log-likelihood with respect to the elements of the vector made up of the entries of the matrix $R_{1,g}$, in the order in which they appear. The matrix H_g can be expressed as a sum of two matrices $(H_g = H_{1,g} + H_{2,g})$; the first is diagonal comprising functions of e_g that are present only in the

non-mixed second derivatives. Later the notation $H = \mathrm{diag}(H_1, H_2)$ and $H^{(h)} = \mathrm{diag}(H_{h,1}, H_{h,2})$, so that $H = H^{(1)} + H^{(2)}$, will be used. Appendix C gives details of the derivation of one of the second derivatives of ℓ, the others proceed in a similar manner. After considerable work of this nature, it follows that

$$H_{1,g} = -(R_{1,g}^{-1}\mathrm{diag}(A_1 e_g)R_{1,g}^{-1} + R_{2,g}^{-1}\,\mathrm{diag}(A_2 e_g)R_{2,g}^{-1}) \tag{A1}$$

$$= -E(R_{1,g}^{-1}\,\mathrm{diag}(A_1 N_g)R_{1,g}^{-1} + R_{2,g}^{-1}\,\mathrm{diag}(A_2 N_g)R_{2,g}^{-1}\,|\,\Pi) \tag{A2}$$

for $g = 1, 2$, where diag represents a diagonal matrix with entries the same as its column vector argument. Furthermore,

$$H_{2,g} = (R_{1,g}^{-1}A_1 - R_{2,g}^{-1}A_2)C_g(R_{1,g}^{-1}A_1 - R_{2,g}^{-1}A_2)' \tag{A3}$$

$$= D\{(R_{1,g}^{-1}A_1 - R_{2,g}^{-1}A_2)N_g|\Pi\}, \quad g = 1, 2 \tag{A4}$$

where D denotes a dispersion (variance–covariance) matrix.

As an aside that will verify the validity of equations (A2) and (A4), note that the general result $D(X) = E\{D(X\,|\,Y)\} + D\{E(X\,|\,Y)\}$ implies that

$$D\{(R_{1,g}^{-1}A_1 - R_{2,g}^{-1}A_2)N_g\} = E\{D((R_{1,g}^{-1}A_1 - R_{2,g}^{-1}A_2)N_g\,|\,\Pi)\} + D\{E((R_{1,g}^{-1}A_1 - R_{2,g}^{-1}A_2)N_g\,|\,\Pi)\}$$
$$= E(H_{2,g}) + D(U_g) \tag{A5}$$

Now

$$D\{(R_{1,g}^{-1}A_1 - R_{2,g}^{-1}A_2)N_g\} = (R_{1,g}^{-1}A_1 - R_{2,g}^{-1}A_2)D(N_g)(R_{1,g}^{-1}A_1 - R_{2,g}^{-1}A_2)' \tag{A6}$$

and $D(N_g) = D_{1,g} + D_{2,g}$, where $D_{1,g}$ is diagonal with entries $n_{\circ\bullet\bullet,g}\,p_{ijk,g}$ and $D_{2,g}$ has entries $-n_{\circ\bullet\bullet,g}\,p_{ijk,g}\,p_{i'j'k',g}$. The term in (A6) corresponding to $D_{2,g}$ is equal to 0, and by careful consideration of matrices of the form $A_h D_{1,g}(A_{h'})'$, $h = h' = 1, 2$, it follows that

$$(R_{1,g}^{-1}A_1 - R_{2,g}^{-1}A_2)D_{1,g}(R_{1,g}^{-1}A_1 - R_{2,g}^{-1}A_2)'$$
$$= R_{1,g}^{-1}A_1 D_{1,g}A_1' R_{1,g}^{-1} - R_{1,g}^{-1}A_1 D_{1,g}A_2' R_{2,g}^{-1} - R_{2,g}^{-1}A_2 D_{1,g}A_1' R_{1,g}^{-1} + R_{2,g}^{-1}A_2 D_{1,g}A_2' R_{2,g}^{-1}$$
$$= E\{R_{1,g}^{-1}\,\mathrm{diag}(A_1 N_g)R_{1,g}^{-1} + R_{2,g}^{-1}\,\mathrm{diag}(A_2 N_g)R_{2,g}^{-1}\} = -E(H_{1,g}), \quad g = 1, 2$$

Thus (A5) can be written $-E(H_{1,g}) = E(H_{2,g}) + D(U_g)$, so that $D(U_g) = -E(H_g)$. The dispersion matrix of the efficient score statistic is equal to minus the expected value of the matrix of second derivatives of log-likelihood. This is a general result, but in this case neither $D(U_g)$ nor $E(H_g)$ can actually be evaluated without a model for the pattern of missingness Π, and no such model is being assumed.

APPENDIX B: FISHER'S INFORMATION, V

To derive Fisher's observed information, V, the parameters $r_{1,1}^{(1)}$ and $r_{1,2}^{(1)}$ need to be replaced in the system of second derivatives by θ and ϕ, and unknown parameters replaced by their restricted maximum likelihood estimates. We have

$$\frac{\partial^2\ell}{\partial\theta^2}=\frac{\partial}{\partial\theta}\left(\sum_{g=1}^{2}\frac{\partial\ell}{\partial r_{1,g}^{(1)}}\frac{\partial r_{1,g}^{(1)}}{\partial\theta}\right)=\sum_{g=1}^{2}\frac{\partial^2\ell}{\partial(r_{1,g}^{(1)})^2}\left(\frac{\partial r_{1,g}^{(1)}}{\partial\theta}\right)^2+\sum_{g=1}^{2}\frac{\partial\ell}{\partial r_{1,g}^{(1)}}\frac{\partial^2 r_{1,g}^{(1)}}{\partial\theta^2}$$

As $\partial r_{1,g}^{(1)}/\partial\theta=(-1)^{g-1}r_{1,g}^{(1)}r_{2,g}^{(1)}/2$, it follows that $\partial^2 r_{1,g}^{(1)}/\partial\theta^2=(r_{2,g}^{(1)}-r_{1,g}^{(1)})r_{1,g}^{(1)}r_{2,g}^{(1)}/4$. Thus, under the null hypothesis, $\partial r_{1,1}^{(1)}/\partial\theta=-\partial r_{1,2}^{(1)}/\partial\theta=\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}/(2n_{\circ\bullet\bullet,\circ}^2)$ and $\partial^2 r_{1,1}^{(1)}/\partial\theta^2=\partial^2 r_{1,2}^{(1)}/\partial\theta^2=(\tilde{e}_{\circ\circ 2,\circ}-\tilde{e}_{\circ\circ 1,\circ})\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}/(4n_{\circ\bullet\bullet,\circ}^3)$. It follows that the second term in the expression for $\partial^2\ell/\partial\theta^2$ is equal to zero, so that

$$\frac{\partial^2\ell}{\partial\theta^2}=\left(\frac{\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}}{2n_{\circ\bullet\bullet,\circ}^2}\right)^2\sum_{g=1}^{2}\frac{\partial^2\ell}{\partial(r_{1,g}^{(1)})^2}$$

It can be seen that $\partial^2\ell/\partial\phi^2$ will take the same value. For the mixed derivative

$$\frac{\partial^2\ell}{\partial\theta\partial\phi}=\frac{\partial}{\partial\theta}\left(\sum_{g=1}^{2}\frac{\partial\ell}{\partial r_{1,g}^{(1)}}\frac{\partial r_{1,g}^{(1)}}{\partial\phi}\right)=\sum_{g=1}^{2}\frac{\partial^2\ell}{\partial(r_{1,g}^{(1)})^2}\frac{\partial r_{1,g}^{(1)}}{\partial\theta}\frac{\partial r_{1,g}^{(1)}}{\partial\phi}+\sum_{g=1}^{2}\frac{\partial\ell}{\partial r_{1,g}^{(1)}}\frac{\partial^2 r_{1,g}^{(1)}}{\partial\theta\partial\phi}$$

Under the null hypothesis, using the forms of the first derivatives given above, $\partial^2 r_{1,1}^{(1)}/\partial\theta\partial\phi=\partial^2 r_{1,2}^{(1)}/\partial\theta\partial\phi=(-1)^{g-1}(\tilde{e}_{\circ\circ 2,\circ}-\tilde{e}_{\circ\circ 1,\circ})\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}/(4n_{\circ\bullet\bullet,\circ}^3)$. Consequently,

$$\frac{\partial^2\ell}{\partial\theta\partial\phi}=\left(\frac{\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}}{2n_{\circ\bullet\bullet,\circ}^2}\right)^2\sum_{g=1}^{2}(-1)^{g-1}\frac{\partial^2\ell}{\partial(r_{1,g}^{(1)})^2}+\frac{(\tilde{e}_{\circ\circ 2,\circ}-\tilde{e}_{\circ\circ 1,\circ})\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}}{4n_{\circ\bullet\bullet,\circ}^2}$$

$$\times\sum_{g,k=1}^{2}(-1)^{g-1}(-1)^{k-1}\frac{\tilde{e}_{\circ\circ k,g}}{\tilde{e}_{\circ\circ k,\circ}}$$

If the second derivatives were being evaluated at the unrestricted maximum likelihood estimates, then $\partial\ell/\partial r_{1,g}^{(1)}$ would be equal to zero for $g=1,2$. However, here evaluation is at the restricted maximum likelihood estimates, and so this does not hold. The second term in both $\partial^2\ell/\partial\theta^2$ and $\partial^2\ell/\partial\phi^2$ reduces to zero anyway, but in $\partial^2\ell/\partial\theta\partial\phi$ it remains non-zero. This odd term turns out to enable simplification to take place later: it will be denoted by y. Other mixed derivatives are given by

$$\frac{\partial^2\ell}{\partial r_{1k,g}^{(2)}\partial\theta}=(-1)^{g-1}\left(\frac{\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}}{2n_{\circ\bullet\bullet,\circ}^2}\right)\frac{\partial^2\ell}{\partial r_{1k,g}^{(2)}\partial r_{1,g}^{(1)}},\quad\frac{\partial^2\ell}{\partial r_{1k,g}^{(2)}\partial\phi}=\left(\frac{\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}}{2n_{\circ\bullet\bullet,\circ}^2}\right)\frac{\partial^2\ell}{\partial r_{1k,g}^{(2)}\partial r_{1,g}^{(1)}}$$

$$\frac{\partial^2\ell}{\partial r_{1jk,g}^{(3)}\partial\theta}=(-1)^{g-1}\left(\frac{\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}}{2n_{\circ\bullet\bullet,\circ}^2}\right)\frac{\partial^2\ell}{\partial r_{1jk,g}^{(3)}\partial r_{1,g}^{(1)}},\quad\frac{\partial^2\ell}{\partial r_{1jk,g}^{(3)}\partial\phi}=\left(\frac{\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}}{2n_{\circ\bullet\bullet,\circ}^2}\right)\frac{\partial^2\ell}{\partial r_{1jk,g}^{(3)}\partial r_{1,g}^{(1)}}$$

Let H^* be the Hessian matrix with respect to θ, ϕ, the $r^{(3)}_{1jk,1}$, the $r^{(2)}_{1k,1}$, the $r^{(3)}_{1jk,2}$, and the $r^{(2)}_{1k,2}$ in that order. Let W be the 14×14 diagonal matrix $\mathrm{diag}(w, w, 1, \ldots, 1)$ with $w = \tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}/(2n^2_{\circ\bullet\bullet,\circ})$, W_1 be the 2×14 matrix comprising the first 2 rows of W, W_2 the 6×14 matrix comprising the next 6 rows of W and W_3 the 6×14 matrix comprising the last 6 rows of W. Let B be the 14×14 matrix of the form

$$B = \begin{pmatrix} 0 & 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & 0 & 0 & 1 \\ I_4 & 0 & 0 & 0 & 0 & 0 \\ 0 & I_2 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & I_4 & 0 & 0 \\ 0 & 0 & 0 & 0 & I_2 & 0 \end{pmatrix}$$

where I_h is the $h \times h$ identity matrix. Let Y be the 14×14 matrix with entry y in positions $(1,2)$ and $(2,1)$ and zeros elsewhere, and Y_2 the 2×2 matrix comprising the leading entries of Y. Then, with the parameters inherent in the H matrices replaced by their restricted maximum likelihood estimates

$$\begin{aligned} H^* &= WBHB'W' + Y \\ &= \begin{pmatrix} W_1B(H^{(1)} + H^{(2)})B'W_1' + Y_2 & W_1BH^{(2)}B'W_2' & W_1BH^{(2)}B'W_3' \\ W_2BH^{(2)}B'W_1' & W_2B(H^{(1)} + H^{(2)})B'W_2' & 0 \\ W_3BH^{(2)}B'W_1' & 0 & W_3B(H^{(1)} + H^{(2)})B'W_3' \end{pmatrix} \end{aligned}$$

because $H^{(1)}$ is diagonal, so that $BH^{(1)}B'$ has a 2×2 block in the top left-hand corner and is otherwise diagonal. Furthermore, $H^{(2)}$ is block diagonal, with two 7×7 blocks, and so $BH^{(2)}B'$ has blocks of zeros in the last 6 columns of rows 3–8 and in the last 6 rows of columns 3–8. It follows that $W_2BH^{(2)}B'W_3' = W_3BH^{(2)}B'W_2' = 0$, explaining the zero entries above. Now let $H^*_{11} = W_1BH^{(1)}B'W_1' + Y_2$. Using equation (A1), it can be seen that the diagonal entries of this 2×2 matrix are

$$-w^2 \sum_{g,k=1}^{2} \frac{\tilde{e}_{\circ\circ k,g}}{(\tilde{r}^{(1)}_{k,g})^2} = -\left(\frac{\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}}{2n_{\circ\bullet\bullet,\circ}}\right)^2 \sum_{g,k=1}^{2} \frac{\tilde{e}_{\circ\circ k,g}}{(\tilde{e}_{\circ\circ k,\circ})^2} = -\frac{\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}}{4n_{\circ\bullet\bullet,\circ}}$$

Off-diagonal entries are equal to

$$\begin{aligned} &-w^2 \sum_{g,k=1}^{2} (-1)^{g-1} \frac{\tilde{e}_{\circ\circ k,g}}{(\tilde{r}^{(1)}_{k,g})^2} + y \\ &= -\left(\frac{\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}}{2n_{\circ\bullet\bullet,\circ}}\right)^2 \left\{ \sum_{g,k=1}^{2} (-1)^{g-1} \frac{\tilde{e}_{\circ\circ k,g}}{(\tilde{e}_{\circ\circ k,\circ})^2} - \frac{(\tilde{e}_{\circ\circ 2,\circ} - \tilde{e}_{\circ\circ 1,\circ})}{\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}} \sum_{g,k=1}^{2} (-1)^{g-1}(-1)^{k-1} \frac{\tilde{e}_{\circ\circ k,g}}{\tilde{e}_{\circ\circ k,\circ}} \right\} \\ &= -\left(\frac{\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}}{4n^2_{\circ\bullet\bullet,\circ}}\right)(n_{\circ\bullet\bullet,1} - n_{\circ\bullet\bullet,2}) \end{aligned}$$

and so

$$H_{11}^{*}=-\left(\frac{\tilde{e}_{\circ\circ 1,\circ}\tilde{e}_{\circ\circ 2,\circ}}{4n_{\circ\bullet\bullet,\circ}^{2}}\right)\begin{pmatrix} n_{\circ\bullet\bullet,\circ} & n_{\circ\bullet\bullet,1}-n_{\circ\bullet\bullet,2} \\ n_{\circ\bullet\bullet,1}-n_{\circ\bullet\bullet,2} & n_{\circ\bullet\bullet,\circ}\end{pmatrix}$$

When the data are complete $\tilde{e}_{\circ\circ 1,\circ}=S$, $\tilde{e}_{\circ\circ 2,\circ}=F$, $n_{\circ\bullet\bullet,1}=n_1$, $n_{\circ\bullet\bullet,2}=n_2$ and $n_{\circ\bullet\bullet,\circ}=n$, in the notation of Section 4, which leads to the usual form: $V=n_1n_2SF/n^3$. For incomplete data, V is minus the reciprocal of the leading element of the inverse of H^{*}: that is minus the reciprocal of the leading element of

$$\begin{aligned}&[H_{11}^{*}+W_1BH^{(2)}B'W_1'-W_1BH^{(2)}B'W_2'\{W_2B(H^{(1)}+H^{(2)})B'W_2'\}^{-1}W_2BH^{(2)}B'W_1'\\&\quad -W_1BH^{(2)}B'W_3'\{W_3B(H^{(1)}+H^{(2)})B'W_3'\}^{-1}W_3BH^{(2)}B'W_1']^{-1}\end{aligned}\tag{B1}$$

When computing V, if any of the r's are estimated to be zero or one, then the corresponding row and column in each of the vectors and matrices from which V is found should be removed before computation. For example, in the final analysis of the trial described in the next section, the estimates of $r_{112,2}$, $r_{122,2}$, $r_{12,1}$, $r_{12,2}$ are all zero and those of $r_{212,2}$, $r_{222,2}$, $r_{22,1}$, $r_{22,2}$ are zero or one. Thus, these parameters do not actually appear in the likelihood. The diagonal matrices of r's become $R_{h,1}=\mathrm{diag}(r_{h11,1}^{(3)},r_{h12,1}^{(3)},r_{h21,1}^{(3)},r_{h22,1}^{(3)},r_{h1,1}^{(2)},r_{h,1}^{(1)})$ and $R_{h,2}=\mathrm{diag}(r_{h11,2}^{(3)},r_{h21,2}^{(3)},r_{h1,2}^{(2)},r_{h,2}^{(1)})$, the sixth element of $R_{h,1}$ being removed and the second, fourth and sixth elements of $R_{h,2}$, $h=1,2$. In dealing with group 1, the sixth rows of A_1 and A_2 are removed, while dealing with group 2 the second, fourth and sixth rows of A_1 and A_2 are removed. The vectors e_1 and e_2 and the matrices C_1 and C_2 are unchanged. This allows the calculation of $H_{1,g}$ and $H_{2,g}$ for $g=1,2$. The former will be 6×6 and the latter 4×4. The matrix B has to be modified by removal of columns 6, 9, 11 and 13 and rows 8, 10, 12 and 14. These are column $(0+6)$ and row $(2+6)$ for the first group and columns $(7+2)$, $(7+4)$ and $(7+6)$ and rows $(2+6+2)$, $(2+6+4)$ and $(2+6+6)$ for the second. The matrix W is diagonal, and is reduced to 10×10 by the removal of elements 8, 10, 12 and 14 from the diagonal. The submatrices of W will be as follows: W_1 is the 2×10 matrix with wI_2 at the left-hand side and zeros everywhere else, W_2 be the 5×10 matrix with I_5 in the third to seventh columns and zeros everywhere else and W_3 the 3×10 matrix with I_3 at the right-hand side, and zeros everywhere else. Thus W remains the concatenation of W_1, W_2 and W_3.

APPENDIX C: Derivation of $\partial^2\ell/\partial r_{1jk,g}^{(3)}\partial r_{1k',g}^{(2)}$

From equation (6), $\partial\ell/\partial r_{1k,g}^{(2)}=\sum_j(-1)^{j-1}e_{\circ jk,g}/r_{1k,g}^{(2)}$, and as $p_{ijk,g}=r_{ijk,g}^{(3)}r_{jk,g}^{(2)}r_{k,g}^{(1)}$,

$$\frac{\partial\ell}{\partial r_{1k,g}^{(2)}}=\sum_{i,j=1}^{2}(-1)^{j-1}\left(\frac{n_{ijk,g}}{p_{ijk,g}}+\frac{n_{ij*,g}}{p_{ij\circ,g}}+\frac{n_{i**,g}}{p_{i\circ\circ,g}}\right)r_{ijk,g}^{(3)}\,r_{k,g}^{(1)}$$

Next,

$$\frac{\partial^2\ell}{\partial r_{1jk,g}^{(3)}\partial r_{1k',g}^{(2)}}=\frac{\partial}{\partial r_{1jk,g}^{(3)}}\left(\frac{\partial\ell}{\partial r_{1k',g}^{(2)}}\right)$$

$$= \frac{\partial}{\partial r^{(3)}_{1jk,g}} \left(\sum_{i,j'=1}^{2} (-1)^{j'-1} \left(\frac{n_{ij'k',g}}{p_{ij'k',g}} + \frac{n_{ij'*,g}}{p_{ij'\circ,g}} + \frac{n_{i**,g}}{p_{i\circ\circ,g}} \right) r^{(3)}_{1j'k',g} r^{(1)}_{k',g} \right)$$

$$= \sum_{i,j'=1}^{2} (-1)^{j'-1} \left\{ \frac{\partial}{\partial r^{(3)}_{1jk,g}} \left(\frac{n_{ij'k',g}}{p_{ij'k',g}} + \frac{n_{ij'*,g}}{p_{ij'\circ,g}} + \frac{n_{i**,g}}{p_{i\circ\circ,g}} \right) r^{(3)}_{ij'k',g} r^{(1)}_{k',g} \right.$$

$$\left. + \left(\frac{n_{ij'k',g}}{p_{ij'k',g}} + \frac{n_{ij'*,g}}{p_{ij'\circ,g}} + \frac{n_{i**,g}}{p_{i\circ\circ,g}} \right) \frac{\partial}{\partial r^{(3)}_{1jk,g}} (r^{(3)}_{ij'k',g} r^{(1)}_{k',g}) \right\} \tag{C1}$$

Now

$$\frac{\partial}{\partial r^{(3)}_{1jk,g}} \left(\frac{n_{ij'k',g}}{p_{ij'k',g}} + \frac{n_{ij'*,g}}{p_{ij'\circ,g}} + \frac{n_{i**,g}}{p_{i\circ\circ,g}} \right)$$

$$= \sum_{i'',j'',k''=1}^{2} \frac{\partial}{\partial p_{i''j''k'',g}} \left(\frac{n_{ij'k',g}}{p_{ij'k',g}} + \frac{n_{ij'*,g}}{p_{ij'\circ,g}} + \frac{n_{i**,g}}{p_{i\circ\circ,g}} \right) \frac{\partial p_{i''j''k'',g}}{\partial r^{(3)}_{1jk,g}}$$

The first term of each summand will be equal to zero unless $i'' = i$. Also, because $p_{i''j''k'',g} = r^{(3)}_{i''j''k'',g} r^{(2)}_{j''k'',g} r^{(1)}_{k'',g}$, the second term will be zero unless $j'' = j$ and $k'' = k$. Hence,

$$\frac{\partial}{\partial r^{(3)}_{1jk,g}} \left(\frac{n_{ij'k',g}}{p_{ij'k',g}} + \frac{n_{ij'*,g}}{p_{ij'\circ,g}} + \frac{n_{i**,g}}{p_{i\circ\circ,g}} \right) = \frac{\partial}{\partial p_{ijk,g}} \left(\frac{n_{ij'k',g}}{p_{ij'k',g}} + \frac{n_{ij'*,g}}{p_{ij'\circ,g}} + \frac{n_{i**,g}}{p_{i\circ\circ,g}} \right) \frac{\partial p_{ijk,g}}{\partial r^{(3)}_{1jk,g}} \tag{C2}$$

Now, $\partial p_{ijk,g} / \partial r^{(3)}_{1jk,g} = (-1)^{i-1} r^{(2)}_{jk,g} r^{(1)}_{k,g}$, and

$$\frac{\partial}{\partial p_{ijk,g}} \left(\frac{n_{ij'k',g}}{p_{ij'k',g}} + \frac{n_{ij'*,g}}{p_{ij'\circ,g}} + \frac{n_{i**,g}}{p_{i\circ\circ,g}} \right) = -\frac{n_{ijk,g}\delta_{jj'}\delta_{kk'}}{(p_{ijk,g})^2} - \frac{n_{ij*,g}\delta_{jj'}}{(p_{ij\circ,g})^2} - \frac{n_{i**,g}}{(p_{i\circ\circ,g})^2} \tag{C3}$$

Also, $\partial (r^{(3)}_{ij'k',g} r^{(1)}_{k',g}) / \partial r^{(3)}_{1jk,g} = (-1)^{i-1} r^{(1)}_{k,g} \delta_{jj'} \delta_{kk'}$. Substituting this together with (C2) and (C3) into (C1) gives

$$\frac{\partial^2 \ell}{\partial r^{(3)}_{1jk,g} \partial r^{(2)}_{1k',g}} = (-1)^{j-1} \sum_{i,j'=1}^{2} (-1)^{i+j'} \left\{ \left(-\frac{n_{ijk,g}\delta_{jj'}\delta_{kk'}}{(p_{ijk,g})^2} - \frac{n_{ij*,g}\delta_{jj'}}{(p_{ij\circ,g})^2} - \frac{n_{i**,g}}{(p_{i\circ\circ,g})^2} \right) r^{(1)}_{k,g} r^{(1)}_{k',g} r^{(2)}_{jk,g} r^{(3)}_{ij'k',g} \right.$$

$$\left. + \left(\frac{n_{ijk,g}}{p_{ijk,g}} + \frac{n_{ij*,g}}{p_{ij\circ,g}} + \frac{n_{i**,g}}{p_{i\circ\circ,g}} \right) \delta_{jj'} \delta_{kk'} r^{(1)}_{k,g} \right\}$$

$$= (-1)^{j-1} \sum_{i,j'=1}^{2} (-1)^{i} (-1)^{j-j'} \frac{1}{r^{(2)}_{j'k',g} r^{(3)}_{ijk,g}} (n_{ij*,g} q^{(3)}_{ijk,g} \delta_{jj'} \delta_{kk'} - n_{ij*,g} q^{(3)}_{ijk,g} q^{(3)}_{ijk',g} \delta_{jj'}$$

$$+ n_{i**,g} q^{(3)}_{ijk,g} q^{(2)}_{ij,g} \delta_{jj'} \delta_{kk'} - n_{i**,g} q^{(2)}_{ij,g} q^{(2)}_{ij',g} q^{(3)}_{ijk,g} q^{(3)}_{ij'k',g})$$

$$= (-1)^{j-1} \sum_{i,j'=1}^{2} (-1)^{i-1} (-1)^{j-j'} \frac{c_{i,(jk),(j'k),g}}{r^{(2)}_{j'k,g} r^{(3)}_{ijk,g}}$$

ACKNOWLEDGEMENTS

During this research, Dr Sooriyarachchi was in receipt of a Wellcome Trust Short-term Travel Grant. The authors are grateful to Sanofi-Synthelabo for permission to include unpublished details of the head injury study used in illustration.

REFERENCES

1. Lees KR, Asplund K, Carolei A, Davis S, Diener HC, Kaste M, Orgogozo JM, Whitehead J. Glycine antagonist (gavestinel) in neuroprotection (GAIN International) in patients with acute stroke: a randomised controlled trial. *Lancet* 2000; **335**:1949–1954.
2. Marshall LF, Maas AI, Marshall SB, Bricolo A, Fearnside M, Iannotti F, Klauber MR, Lagarrigue J, Lobato R, Perrson L, Pickard JD, Piek J, Servadei F, Wellis GN, Morris GF, Means ED, Musch B. A multicenter trial on the efficacy of using tirilizad mesylate in cases of head injury. *Journal of Neurosurgery* 1998; **89**:519–525.
3. Whitehead J. Application of sequential methods to a phase III clinical trial in stroke. *Drug Information Journal* 1993; **27**:733–740.
4. Bolland K, Weeks A, Whitehead J, Lees KR. How a sequential design would have affected the GAIN international study of gavestinel in stroke. *Cerebrovascular Diseases* 2004; **7**:111–117.
5. Whitehead J. Overrunning and underrunning in sequential clinical trials. *Controlled Clinical Trials* 1992; **13**:106–121.
6. Sooriyarachchi MR, Whitehead J, Matsushita T, Bolland K, Whitehead A. Incorporating data received after a sequential trial has stopped into the final analysis: implementation and comparison methods. *Biometrics* 2003; **59**:701–709.
7. Bolland K, Whitehead J. Formal approaches to safety monitoring of clinical trials in life-threatening conditions. *Statistics in Medicine* 2000; **19**:2899–2917.
8. Whitehead J. *The Design and Analysis of Sequential Clinical Trials* (revised 2nd edn). Wiley: Chichester, 1997.
9. Gange SJ, DeMets DL. Sequential monitoring of clinical trials with correlated responses. *Biometrika* 1996; **83**:157–167.
10. Spiessens B, Lesaffre E, Verbeke G. A comparison of group sequential methods for binary longitudinal data. *Statistics in Medicine* 2003; **22**:501–515.
11. Hedeker D, Gibbons RD. A random-effects ordinal regression model for multilevel analysis. *Biometrics* 1994; **50**:933–944.
12. Hedeker D, Gibbons RD. Mixor: a computer program for mixed effect ordinal regression analysis. *Computer Methods and Programs in Biomedicine* 1996; **49**:157–176 (See also http://tigger.uic.edu/∼hedeker/mixwin.html)
13. Lindsey JK, Lambert P. On the appropriateness of marginal models for repeated measurements in clinical trials. *Statistics in Medicine* 1998; **17**:447–469.
14. Marschner IC, Becker SL. Interim monitoring of clinical trials based on long-term binary endpoints. *Statistics in Medicine* 2001; **20**:177–192.
15. McCullagh P. Regression models for ordinal data. *Journal of the Royal Statistical Society*, *Series B* 1980; **42**:109–142.
16. Whitehead J. Sample size calculations for ordered categorical data. *Statistics in Medicine* 1993; **12**:2257–2271.
17. Bolland K, Sooriyarachchi MR, Whitehead J. Sample size review in a head injury trial with ordered categorical responses. *Statistics in Medicine* 1998; **17**:2835–2847.
18. Jennison CJ, Turnbull BW. *Group Sequential Methods with Applications to Clinical Trials.* Chapman & Hall/CRC: New York, NY, Boca Raton, FL, 2000.
19. Lan KKG, Simon R, Halperin M. Stochastically curtailed tests in long-term clinical trials. *Sequential Analysis* 1982; **1**:207–219.
20. MPS Research Unit. *PEST 4*: *Operating Manual*, The University of Reading, 2000.
21. Rubin DB. Inference and missing data (with discussion). *Biometrika* 1976; **63**:581–592.

