Marginal Modelling of Categorical Data from Crossover Experiments

By CECILE C. BALAGTAS

Warner–Lambert Company, Ann Arbor, USA

MARK P. BECKER†

University of Michigan, Ann Arbor, USA

and JOSEPH B. LANG

University of Iowa, Iowa City, USA

SUMMARY

Marginal models provide a useful framework for the analysis of crossover experiments when the response variable is categorical. In this paper we use the three-treatment, three-period crossover experiment with a binary outcome variable to demonstrate how marginal models can be used to perform a likelihood-based analysis of multiple-period crossover experiments. Other designs are discussed in less detail. Maximum likelihood estimation is performed using a constraint equation specification of the marginal model. Data from a crossover trial comparing treatments for primary dysmenorrhoea are used to demonstrate the utility of marginal models in analysing crossover data.

Keywords: Carry-over effects; Constraint equations; Lagrange multipliers; Maximum likelihood; Repeated measures; Treatment-by-period interaction

1. Introduction

The crossover design is frequently used to compare treatments for medical conditions that are chronic in nature and subject to temporary relief; see, for example, Jones and Kenward (1989). The most commonly used design is the two-treatment, two-period design, but there are some limitations to this design. The main criticism usually levelled against the two-period design is that three potentially important effects, namely sequence group, carry-over and treatment-by-period interaction, are aliased with one another. Several options are available for reducing or eliminating the problem of aliasing. These include using more than two periods, using non-crossover or partial crossover treatment sequences and taking run-in and wash-out base-line measurements (Matthews (1988) and Jones and Kenward (1989), chapters 4–5). The first option, using more than two treatment periods, is what we focus on in this paper.

The purpose of the present paper is to demonstrate how marginal models and

†*Address for correspondence*: Department of Biostatistics, University of Michigan, Ann Arbor, MI 48109-2029, USA.
E-mail: mbecker@umich.edu

likelihood-based inference can be applied to the analysis of crossover experiments where the response variable is categorical. The three-treatment, three-period, binary response experiment is used as the example for describing the models and the estimation procedure. In this design patients are randomized to one of six possible sequence groups: ABC, ACB, BAC, BCA, CAB and CBA. At the end of each period, the response of each subject is classified as favourable (1) or unfavourable (2). Let n_{ijkl} denote the number of subjects in sequence group l having response i, j, k in periods 1, 2 and 3 respectively. Within each group, there are eight possible response patterns over the three periods — (1, 1, 1), (1, 1, 2), (1, 2, 1), (1, 2, 2), (2, 1, 1), (2, 1, 2), (2, 2, 1) and (2, 2, 2).

The data in Table 1 are from a three-period crossover trial which compared three treatments (A, B and C) for the relief of primary dysmenorrhoea. A first step in the analysis of such data is to compare the proportion of subjects who obtained relief according to the treatment received. The relevant (univariate marginal) sample proportions are given in Table 2. The values in Table 2 suggest that treatment A is not as effective at providing relief from primary dysmenorrhoea as either treatments B or C. Treatments B and C appear to be roughly similar. However, it is not immediately clear from Tables 1 and 2 whether there are important associations, sequence group effects, period effects, carry-over effects or other forms of treatment-by-period interactions that need to be taken into account in analysing these data.

Becker and Balagtas (1993) have shown that marginal models provide a useful framework for the analysis of two-treatment, two-period, binary response crossover experiments. They have shown that marginal models address important questions directly, and they have shown that standard likelihood-based inference applied to marginal models performs well when compared with so-called standard procedures for the analysis of two-period studies. There are, however, modelling issues in the analysis of studies with more than two periods that do not arise with two-period studies, and the likelihood maximization algorithm employed by Becker and Balagtas is not applicable to studies with more than two periods. In Section 2, we describe models for analysing data from a three-treatment, three-period, binary response crossover experiment, and we present a brief description of a procedure for fitting the models by the method of maximum likelihood. Asymptotic theory and a more detailed description of the model fitting algorithm are given in Appendix

TABLE 1

Data from a three-treatment, three-period crossover trial on primary dysmenorrhoea†

Group	Response profile (period 1, period 2, period 3)								
	111	*112*	*121*	*122*	*211*	*212*	*221*	*222*	*Total*
ABC	7	2	4	1	23	7	5	4	53
ACB	6	3	2	0	22	12	6	2	53
BAC	4	1	14	7	5	3	14	5	53
BCA	4	17	6	9	3	4	1	4	48
CAB	4	2	29	5	1	0	6	7	54
CBA	2	23	2	4	2	8	1	5	47

†1, some relief; 2, no relief.

TABLE 2
Observed univariate marginal proportions

| Group | Results for the following treatments: | | |
	A	B	C
ABC	0.26	0.74	0.74
ACB	0.21	0.68	0.81
BAC	0.25	0.49	0.70
BCA	0.29	0.75	0.58
CAB	0.13	0.74	0.74
CBA	0.15	0.74	0.66

A. The data in Table 1 are analysed by using marginal models in Section 3. We conclude this paper with a discussion of designs other than the three-treatment, three-period, binary response design.

2. Marginal Models for Three-period Crossover Design

In the three-treatment, three-period, binary response design there are six sequence groups, seven $(2^3 - 1)$ linearly independent probabilities within each sequence group and hence a total of 42 degrees of freedom (estimable parameters). In a marginal model for this experimental design, 18 degrees of freedom pertain to univariate marginal distributions (three in each of the six sequence groups), 18 degrees of freedom pertain to bivariate associations (again, three per sequence group) and the remaining six degrees of freedom pertain to trivariate associations (one per sequence group). The models presented here will be in terms of linear models for univariate marginal logits, linear models for bivariate marginal log-odds ratios and a linear model for a log-odds ratio contrast that measures three-factor association. In presenting each set of models we begin with a saturated model and then proceed to consider more parsimonious models that are of practical interest.

Let π_{ijkl} denote the cell probabilities $(\Sigma_{ijk} \pi_{ijkl} = 1)$, α a general effect, ρ_r the effect due to period r, δ_l the effect due to group l and $(\rho\delta)_{rl}$ a parameter corresponding to group-by-period interaction. The group effects represent differential effects which are associated with the order of treatment administration. The treatment effect is aliased in the group-by-period interaction, but, as is shown below, there are reparameterizations of this interaction that allow for the estimation of treatment effects, carry-over effects and treatment-by-period interactions. For the standard two-treatment, two-period crossover design, the group effects are aliased with the carry-over effects and a treatment-by-period interaction. A saturated model for the univariate marginal logits is

$$\left.\begin{array}{c} \log\left(\dfrac{\pi_{1++l}}{\pi_{2++l}}\right) = \alpha + \delta_l + \rho_1 + (\rho\delta)_{1l}, \\[2mm] \log\left(\dfrac{\pi_{+1+l}}{\pi_{+2+l}}\right) = \alpha + \delta_l + \rho_2 + (\rho\delta)_{2l}, \\[2mm] \log\left(\dfrac{\pi_{++1l}}{\pi_{++2l}}\right) = \alpha + \delta_l + \rho_3 + (\rho\delta)_{3l}, \qquad l = 1, \ldots, 6. \end{array}\right\} \qquad (1)$$

Model (1) is overparameterized in that the parameters are not uniquely determined, i.e. there are 28 parameters, but only 18 logits. The so-called box corner (or corner point) restrictions (e.g. $\delta_1 = 0$, $\rho_1 = 0$, $(\rho\delta)_{1l} = 0$, $(\rho\delta)_{r1} = 0$) are used for identifying model parameters in this paper.

More parsimonious models for the marginal logits can be derived by replacing the group-by-period interactions in model (1) by terms corresponding to treatment effects, carry-over effects and treatment-by-period interactions. Let α, δ_l and ρ_r be as defined previously, and let $\tau_{t[r,l]}$ denote the direct effect of the treatment administered in period r of group l, $(\tau\rho)_{t[r,l],r}$ a treatment-by-period interaction and $\gamma_{t[r,l]}$ the first-order carry-over effect of the treatment administered in period r of group l. Formulating these terms into a model for the univariate logits we obtain

$$
\left.\begin{aligned}
\log\left(\frac{\pi_{1++l}}{\pi_{2++l}}\right) &= \alpha + \delta_l + \tau_{t[1,l]} + \rho_1 + (\tau\rho)_{t[1,l],1}, \\[2mm]
\log\left(\frac{\pi_{+1+l}}{\pi_{+2+l}}\right) &= \alpha + \delta_l + \tau_{t[2,l]} + \rho_2 + (\tau\rho)_{t[2,l],2} + \gamma_{t[1,l]}, \\[2mm]
\log\left(\frac{\pi_{++1l}}{\pi_{++2l}}\right) &= \alpha + \delta_l + \tau_{t[3,l]} + \rho_3 + (\tau\rho)_{t[3,l],3} + \gamma_{t[2,l]}.
\end{aligned}\right\} \quad (2)
$$

There are 16 estimable parameters in model (2), and hence 2 residual degrees of freedom corresponding to unspecified marginal logit contrasts. Hereafter, we simplify our notation by dropping the portion of the subscript indicating period and group; i.e. $[r, l]$. The first-order carry-over effect is the treatment effect that lasts no more than one period beyond the period of application of the treatment. In contrast with the standard two-treatment, two-period crossover design, the present design allows the estimation of first-order carry-over effects separately from a general form of treatment-by-period interaction.

An alternative model for the marginal logits includes parameters for a general effect, group effects, treatment effects, period effects and first-order and second-order carry-over effects, as well as treatment-by-first-order carry-over interaction. The second-order carry-over effect is the treatment effect that persists for two periods beyond the application of the treatment. The treatment-by-first-order carry-over interaction accounts for the possibility that the first-order carry-over effect depends on the treatment received in the following period. Note that the second-order carry-over effect cannot be added to model (1) because of an aliasing relationship between this parameter and the first-order carry-over effect and the treatment-by-period interaction (Jones and Kenward, 1987).

A saturated model for bivariate marginal associations can be specified in terms of models for the association between responses in the mth pair of periods ($m = 1$, 2, 3 for periods 1–2, 1–3 and 2–3 respectively). For each value of m there are parameters for

(a)   a general effect ($\psi^{(m)}$),
(b)   the uth pair of treatments ($\xi_u^{(m)}$, $u = 1, 2, 3$, for treatment pairs AB, AC and BC respectively),
(c)   the order of the treatments within each pair of treatments ($\phi_v^{(m)}$, $v = 1, 2$, e.g. $v = 1$ for AB and $v = 2$ for BA) and

(d)   the interaction of treatment pair and order $((\xi\phi)_{uv}^{(m)})$, i.e.

$$
\left.
\begin{aligned}
\log\left(\frac{\pi_{11+l}\pi_{22+l}}{\pi_{12+l}\pi_{21+l}}\right) &= \psi^{(1)} + \xi_u^{(1)} + \phi_v^{(1)} + (\xi\phi)_{uv}^{(1)}, \\
\log\left(\frac{\pi_{1+1l}\pi_{2+2l}}{\pi_{1+2l}\pi_{2+1l}}\right) &= \psi^{(2)} + \xi_u^{(2)} + \phi_v^{(2)} + (\xi\phi)_{uv}^{(2)}, \\
\log\left(\frac{\pi_{+11l}\pi_{+22l}}{\pi_{+12l}\pi_{+21l}}\right) &= \psi^{(3)} + \xi_u^{(3)} + \phi_v^{(3)} + (\xi\phi)_{uv}^{(3)}.
\end{aligned}
\right\}
\tag{3}
$$

Several restrictions can be used to simplify model (2) in practically meaningful ways. For example, in the analysis of Table 1 we consider models where the bivariate marginal association does not depend on the pair of periods under consideration: $\psi^{(1)} = \psi^{(2)} = \psi^{(3)} = \psi$; $\xi_u^{(1)} = \xi_u^{(2)} = \xi_u^{(3)} = \xi_u$; $\phi_v^{(1)} = \phi_v^{(2)} = \phi_v^{(3)} = \phi_v$; $(\xi\phi)_{uv}^{(1)} = (\xi\phi)_{uv}^{(2)} = (\xi\phi)_{uv}^{(3)} = (\xi\phi)_{uv}$. Further, or other, simplifications can be achieved by removing sets of terms from model (3). A particularly interesting model for practical consideration is the model specifying that all bivariate marginal associations can be summarized in terms of the pair of treatments being compared, net of any effects related to group, periods or ordering. This model is given as model A_{IV} in the subsequent analysis of Table 1.

There is not a set of parameters specific to sequence groups in model (3), as the group effects are aliased with the effects for periods, treatments and ordering. Alternative parameterizations of the bivariate marginal associations that do include group effect parameters are possible. For example, where the only tenable model for bivariate associations is the saturated model it is perhaps simpler to define $\psi_l^{(m)}$ as the mth bivariate association in group l, and then to interpret these as well as is possible.

The model for the trivariate associations includes a general effect for within-group association (φ) and parameters that measure departures from this general effect due to the lth group (ζ_l).

$$
\log\left(\frac{\pi_{111l}\pi_{122l}\pi_{212l}\pi_{221l}}{\pi_{112l}\pi_{121l}\pi_{211l}\pi_{222l}}\right) = \varphi + \zeta_l; \qquad l = 1, \ldots, 6.
\tag{4}
$$

This association is interpreted just as a three-factor association is interpreted in a conventional log-linear model; see, for example, Agresti (1990), pages 144–145. It is quite common for the number of individuals per treatment sequence group to be somewhat smaller than the approximately 50 observations per group in our example. The power for detecting trivariate association is likely to be quite low in sparse data situations, and 0s can present problems for estimation (i.e. some parameters are estimated to be $\pm\infty$). Two possibilities to consider when analysing tables where 0s are an issue are

(a)   to add a small positive constant to each cell count (see, for example, Agresti (1990), chapter 7) and/or

(b)   to assume that $\zeta_l = \zeta$ for all l, and possibly 0.

In Becker and Balagtas (1993), hypothesis testing and parameter estimation for the two-period design were facilitated by a formula for inverting the model for marginal logits and log-odds ratios into a model for cell probabilities in the full

cross-classification. Note that that formula is the solution of a quadratic equation. The difficulty in inverting models for the three-period crossover design is that the roots of higher order polynomials are required, and, hence, the estimation procedure is not as straightforward (Liang *et al.*, 1992). An alternative strategy for model fitting can be based on a constraint equations specification of the model.

A marginal model cannot be rewritten as a log-linear model for the vector of cell probabilities π in general, but there is a convenient matrix representation for the marginal model when the models for logits, log-odds ratios and contrasts of log-odds ratios are linear, i.e.

$$C \log(A\pi) = X\beta. \tag{5}$$

Here A is a matrix of 0s and 1s such that the matrix product $A\pi$ is the required set of marginal distributions; C is a matrix of 1s, -1s and 0s such that the matrix product $C \log(A\pi)$, where the logarithms are taken elementwise, is the set of logits and log-odds ratios of interest; X is a model matrix, β is the vector of model parameters (i.e. $\beta^T = (\alpha, \delta_1, \ldots)$) and p is the number of parameters to be estimated. In the example $\pi = (\pi_{1111}, \ldots, \pi_{2226})^T$ is the 48×1 compound vector of cell probabilities, A is a 156×48 matrix of 0s and 1s, C is a 42×156 matrix of contrasts, X is a $42 \times p$ matrix and β is a $p \times 1$ vector.

Assume that $\{n_{ijkl}\} \sim$ ind mult$(n_{+++l}, \{\pi_{ijkl}\})$, $l = 1, \ldots, 6$. Let n and μ denote the corresponding vectors of cell frequencies and expected cell frequencies respectively, i.e. a typical element of n is n_{ijkl} and a typical element of μ is $\mu_{ijkl} = n_{+++l}\pi_{ijkl}$. For the models considered in this paper, we can use the equivalence of likelihood inference between the multinomial and Poisson distributions (Palmgren, 1981; Lang, 1992) to facilitate likelihood-based calculations. The maximum likelihood estimates of π can be obtained by maximizing the kernel of the Poisson log-likelihood

$$f(\pi) = (\log \mu^T)n - \mu^T 1 \tag{6}$$

subject to the constraints

$$h(\pi) = U^T C \log(A\pi) = 0, \tag{7}$$

where U is a matrix whose columns span the subspace orthogonal to the range space of X. Maximum likelihood estimation of parameters subject to constraints has been discussed by Aitchison and Silvey (1958), Gokhale (1973), Wedderburn (1974), Haber (1985), Haber and Brown (1986) and Lang (1992), among others. The maximum likelihood estimate of β can be derived from the maximum likelihood estimate of π as follows:

$$\hat{\beta} = (X^T X)^{-1} X^T C \log(A\hat{\pi}). \tag{8}$$

An algorithm for performing the constrained optimization is given in Appendix A, along with statements of the asymptotic distribution theory for the maximum likelihood estimators of π and of β.

3. Analysis

Our analysis proceeds by exploring the association structure for the repeated measurements, and then exploring models for univariate marginal logits by using the association structure elicited in the first part of the analysis.

A summary of some of the association models that can be estimated with the data in Table 1, along with the residual degrees of freedom and goodness-of-fit likelihood ratio statistics ($G^2 = -2 \times$ logarithm of the likelihood ratio test statistic for testing that the model is correct *versus* the unrestricted alternative) for each model, is presented in Table 3. The models are specified in terms of the restrictions placed on the association parameters in equations (3) and (4). In comparing models, (asymptotically) size $\alpha = 0.01$ tests are used to account for the fact that a fairly large number of comparisons are performed.

In model A_I, the trivariate log-odds ratios measuring the association of responses for the three periods are assumed to be uniform across groups. The likelihood ratio statistic for this model indicates that this assumption is tenable. In model A_{II}, all trivariate log-odds ratios are set equal to 0. This model implies that the bivariate log-odds ratios measuring the association between any two periods (i.e. [1, 2], [1, 3] or [2, 3]) are uniform across the levels of the third period. The likelihood ratio statistic comparing models A_I and A_{II} indicates that this additional assumption is also tenable.

In addition to the restrictions of model A_{II}, model A_{III} assumes equality of marginal log-odds ratios for periods (1, 2), (1, 3) and (2, 3). The comparison of models A_{II} and A_{III} shows that this additional assumption is tenable. Model A_{IV} contains the restrictions of model A_{III} and, in addition, sets the log-odds ratios measuring the association between treatments in two different periods to be the same for the two possible orders of administration of treatment. For example, the association of responses for treatment A followed by treatment B is the same as the association when treatment B precedes treatment A. The comparison of models A_{III} and A_{IV} indicates that the additional assumption of model A_{IV} is tenable.

The last model in Table 3 sets all bivariate and trivariate log-odds ratios equal

TABLE 3
Fit of various association models to the data in Table 1†

Association model	Restrictions	Degrees of freedom	G^2
A_I	$\zeta_l = 0$	5	4.85
A_{II}	$\varphi = 0,\ \zeta_l = 0$	6	7.70
A_{III}	$\varphi = 0,\ \zeta_l = 0,$ $\psi^{(1)} = \psi^{(2)} = \psi^{(3)} = \psi,$ $\xi_u^{(1)} = \xi_u^{(2)} = \xi_u^{(3)} = \xi_u,$ $\phi_v^{(1)} = \phi_v^{(2)} = \phi_v^{(3)} = \phi_v,$ $(\xi\phi)_{uv}^{(1)} = (\xi\phi)_{uv}^{(2)} = (\xi\phi)_{uv}^{(3)} = (\xi\phi)_{uv}$	18	16.43
A_{IV}	$\varphi = 0,\ \zeta_l = 0,$ $\psi^{(1)} = \psi^{(2)} = \psi^{(3)} = \psi,$ $\xi_u^{(1)} = \xi_u^{(2)} = \xi_u^{(3)} = \xi_u,$ $\phi_v^{(1)} = \phi_v^{(2)} = \phi_v^{(3)} = \phi_v = 0,$ $(\xi\phi)_{uv}^{(1)} = (\xi\phi)_{uv}^{(2)} = (\xi\phi)_{uv}^{(3)} = (\xi\phi)_{uv} = 0$	21	17.80
A_V	$\varphi = 0,\ \zeta_l = 0,$ $\psi^{(1)} = \psi^{(2)} = \psi^{(3)} = \psi = 0,$ $\xi_u^{(1)} = \xi_u^{(2)} = \xi_u^{(3)} = \xi_u = 0,$ $\phi_v^{(1)} = \phi_v^{(2)} = \phi_v^{(3)} = \phi_v = 0,$ $(\xi\phi)_{uv}^{(1)} = (\xi\phi)_{uv}^{(2)} = (\xi\phi)_{uv}^{(3)} = (\xi\phi)_{uv} = 0$	24	21.44

†$l = 1, \ldots, 6;\ u = 1, 2, 3;\ v = 1, 2.$

to 0, which is a model of statistical independence. The likelihood ratio statistic comparing this model with model A_{IV} suggests that the assumption of statistical independence is tenable for these data. Assuming that model A_V is an adequate representation of the association structure, the results of fitting various marginal logit models to the data are summarized in Table 4. Models are specified in terms of the restrictions placed on the marginal logit parameters in model (2), and the last two columns in Table 4 pertain to modifications that are described in the following paragraph.

The fit of model (2) suggests that use of this model as a base-line model is tenable, but the comparison of model (2) with model A_V ($G^2_{(2)} - G^2_{(Av)} = 13.22$; $26 - 24 = 2$ degrees of freedom) indicates that there is some lack of fit in model (2) relative to the saturated model for univariate marginal logits. There are several ways of parameterizing the two univariate marginal (logit) contrasts unaccounted for by model (2), but potential sources of the lack of fit are easily traced by examining the sample proportions in Table 2. For example, we can parameterize the two unused degrees of freedom in terms of

(i) a parameter to account for the fact that the probability of success for treatment A is consistently lower in groups where treatment C is administered first (i.e. the CAB and CBA groups) and

(ii) a parameter to account for the fact that the probability of success for treatment B is much lower in the BAC group.

Adding a single parameter to modify the effect of treatment A when C is administered first results in $G^2 = 26.47$ (25 residual degrees of freedom), and adding a single parameter to modify the effect of treatment B in the BAC group results in $G^2 = 29.80$ (25 residual degrees of freedom, also). Adding both terms into model (2) is equivalent to fitting the saturated model for univariate marginals, i.e. model A_V. The third and fourth columns of Table 4 report G^2 and residual degrees of freedom for each model when the two additional terms (i) and (ii) are added to the model. Given that the groups were formed by randomization, the fact that the probability of success with treatment B does not appear to be lower when treatment C is administered first, and model (2) provides a reasonable overall summary of the data, it seems reasonable to conclude that the lack of fit of model (2) relative to model A_V is due to sampling variability rather than to some form of systematic

TABLE 4
Fit of various marginal logit models to the data in Table 1†

Marginal logit model	Restrictions	G^2	Degrees of freedom	Modified G^2	Modified degrees of freedom
(2)	—	34.66	26	21.44	24
M_I	$(\tau\rho)_{tr} = 0$	37.35	30	28.54	28
M_{II}	$\gamma_t = 0, (\tau\rho)_{tr} = 0$	41.34	32	29.90	30
M_{III}	$\tau_t = 0, \gamma_t = 0, (\tau\rho)_{tr} = 0$	243.14	34	123.45	32
M_{IV}	$\delta_l = 0, \gamma_t = 0, (\tau\rho)_{tr} = 0$	46.33	37	30.73	35
M_V	$\rho_r = 0, \delta_l = 0, \gamma_t = 0, (\tau\rho)_{tr} = 0$	47.18	39	30.79	37
M_{VI}	$\tau_t = 0, \rho_r = 0, \delta_l = 0, \gamma_t = 0, (\tau\rho)_{tr} = 0$	248.72	41	168.65	39

†$l = 1, \ldots, 6$; $t = 1, 2, 3$; $r = 1, 2, 3$.

variation that should be carried forwards in the analysis. However, both for completeness and to demonstrate that the main conclusions drawn from the data in Table 1 are not sensitive to the aberrations noted, information about the fit of the models accounting for the aberrant observations is provided below.

The comparisons of certain nested models in Table 4 indicate no evidence of treatment-by-period interaction (model (2) *versus* M_I), first-order carry-over effects (model M_I *versus* M_{II}) or period effects (model M_{IV} *versus* M_V). However, the comparison of models M_{II} and M_{III}, as well as M_V and M_{VI}, indicates strong evidence of treatment effects.

Estimates of the treatment effects for the final marginal logit model (M_V) under three different *association* structures,

(a) model A_V (independence model),
(b) model A_{IV} (bivariate marginal association) and
(c) model (3) (saturated model),

are given in Table 5. The parameter estimates and standard errors for all three association models are similar. The estimate of the treatment difference for treatments B and C is identical for association structures A_{IV} and A_V, and only slightly different for the saturated model. Thus, neither parameter estimates nor associated estimated standard errors are sensitive to the assumed association structure in the present example.

The estimates of the treatment effects for the model M_V indicate that the odds of relief under treatment B are eight times ($e^{\hat{\tau}_2} = 8.1$) larger than the odds of relief under treatment A. Furthermore, the odds of relief under treatment C are approximately nine times ($e^{\hat{\tau}_3} = 8.9$) larger than the odds of relief under treatment A. However, the odds of relief under treatment C are not significantly higher than the odds of relief under treatment B ($e^{\hat{\tau}_2 - \hat{\tau}_3} = 0.91$). The estimates of the marginal probabilities of relief are 0.21, 0.69 and 0.71, for treatments A, B and C respectively.

The estimates of α, τ_2 and τ_3 are -1.09 (0.16), 2.08 (0.21) and 1.98 (0.20) respectively, when model M_V is modified to account for the two aberrations noted previously; i.e. (i) and (ii). The estimates of the marginal probabilities of relief are 0.25, 0.73 and 0.71, for treatments A, B and C respectively. Hence, even though accounting for the aberrations results in large changes in G^2, the estimated

TABLE 5
Parameter estimates for selected models†

| Parameter | Results for the following association models: | | |
	Independence	A_{IV}	Saturated
$\hat{\alpha}$	-1.299 (0.139)	-1.291 (0.136)	-1.322 (0.139)
$\hat{\tau}_2$	2.092 (0.186)	2.082 (0.190)	2.121 (0.192)
$\hat{\tau}_3$	2.184 (0.187)	2.175 (0.186)	2.197 (0.182)
$\hat{\tau}_2 - \hat{\tau}_3$	-0.093 (0.169)	-0.093 (0.169)	-0.077 (0.166)
G^2	47.18	43.69	25.11
Degrees of freedom	39	36	15

†Standard errors are given in parentheses.

marginal probabilities change very little when the aberrations are taken into account. Note that the estimated value for the change in the effect of treatment A when it follows treatment C is −0.73 (0.33), and the estimated value of the parameter accounting for the aberration noted for treatment B in the BAC group is −1.03 (0.31).

4. Discussion

Although the focus of the present paper is on the three-treatment, three-period crossover experiment, the methodology proposed is readily extended to other multi-period designs such as the two-treatment, three-period design or the four-treatment, four-period design. In all cases, the first step in model construction is the formulation of models for the univariate marginal logits and the association structure of the data. These models jointly determine the model for the cell probabilities. Covariates, such as treatment centre and base-line measurements, may also be entered into the marginal models through the X-matrix in equation (5). The proposed fitting procedure based on the constraint equations specification of the model can be used to compute the maximum likelihood fit of the model. The analysis presented in Section 3 demonstrates that the full likelihood approach to model fitting is feasible for higher order designs. All the computations for this paper were performed with a Fortran program written by the third author. Anyone interested in obtaining a copy of the program should contact Professor Lang directly. A more complete treatment of the constraint equation approach to fitting marginal models by the method of maximum likelihood is given in Lang (1992).

The marginal modelling framework can also be applied to the analysis of crossover experiments where the response variable takes on more than two levels. In the setting where there are r unordered categories it is reasonable to model a set of $r - 1$ univariate logits that are defined by comparing each response category to a base-line category. The full model is specified by models for the base-line category logits of the univariate marginal probabilities along with models for the association structure of the data. The association structure can be modelled by using local odds ratios (Agresti (1990), p. 18) and contrasts of local odds ratios, much as is done in log-linear modelling. If the r response categories are ordered, we have the option of modelling the univariate marginals by using either cumulative logits, adjacent category logits or continuation ratio logits; see section 9.3 of Agresti (1990) for definitions of each of these logits. The selection of the type of logit used should be based on the nature of the response measured. The association structure for ordinal variables can be modelled by using either local odds ratios or global cross-ratios (see, for example, Dale (1986)). There are close connections between certain models for association based on local odds ratios and the bivariate normal distribution (Goodman, 1981, 1985, 1991; Holland and Wang, 1987; Becker, 1989), which suggests that modelling the association in terms of local odds ratios is sensible when the assumption of an underlying multivariate normal distribution is tenable. No new computational difficulties are introduced by moving from a dichotomous response to a polytomous response, in that the algorithm described in Appendix A can still be used to estimate the models described above by the method of maximum likelihood.

Many other approaches to specifying and estimating marginal models have been

published; see, for example, Liang and Zeger (1986), Prentice (1988), Liang *et al.* (1992), Ratkowsky *et al.* (1993) and Fitzmaurice and Laird (1993). The approaches vary in how associations are measured (correlations or odds ratios, and conditional *versus* marginal), which associations are modelled (bivariate only, or all possible orders of association) and the criteria used in deriving estimators (so-called generalized estimating equations methods or maximum likelihood). The approach taken here is to model all possible marginal associations by using odds ratios (and contrasts of log-odds ratios) as the fundamental units for measuring association, with the method of maximum likelihood being used as the basis for inference. The full likelihood also provides a necessary part of the machinery for Bayesian calculations, should they be desired (Ten Have and Becker, 1994). The practical limiting factor in the constrained maximum likelihood approach to marginal modelling is that the matrices A and C grow rapidly as the number of response categories grows and/or as the number of study periods grows. For example, if the response in our example had three categories (rather than two) the matrices A and C would be of dimensions 378×162 and 156×378 respectively (rather than 156×48 and 42×156 respectively).

Acknowledgements

Appendix A: Distribution Theory and Algorithm for Computing Maximum Likelihood Estimates

The algorithm presented in this appendix is based on the application of Newton's method to an objective function that is simply the log-likelihood function augmented by a matrix product involving Lagrange multipliers and constraints based on the model $C \log(A\pi) = X\beta$. The general algorithm was, to the best of our knowledge, first described by Aitchison and Silvey (1958).

Let

$$
M_1 = \begin{pmatrix}
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\
1 & 1 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 1 & 1 \\
1 & 0 & 1 & 0 & 1 & 0 & 1 & 0 \\
0 & 1 & 0 & 1 & 0 & 1 & 0 & 1
\end{pmatrix}, \quad
M_2 = \begin{pmatrix}
1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 \\
1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \\
1 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 1
\end{pmatrix}
$$

and

$$M_3 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix}.$$

The matrix A can then be written as

$$A = \begin{pmatrix} A_{1(36 \times 48)} \\ A_{2(72 \times 48)} \\ A_{3(48 \times 48)} \end{pmatrix}$$

where

$$A_1 = M_1 \otimes I_6,$$

$$A_2 = M_2 \otimes I_6,$$

$$A_3 = M_3 \otimes I_6$$

and $\otimes$ denotes the Kronecker product $(D \otimes E = \{De_{ij}\})$. The contrast matrix that produces the required marginal logits, bivariate log-odds ratios and trivariate log-odds ratios can be represented as

$$C = \begin{pmatrix} C_{1(18 \times 36)} & 0_{(18 \times 72)} & 0_{(18 \times 48)} \\ 0_{(18 \times 36)} & C_{2(18 \times 72)} & 0_{(18 \times 48)} \\ 0_{(6 \times 36)} & 0_{(6 \times 72)} & C_{3(6 \times 48)} \end{pmatrix}$$

where

$$C_1 = [(1, \ -1) \otimes I_3] \otimes I_6,$$

$$C_2 = [(1, \ -1, \ -1, \ 1) \otimes I_3] \otimes I_6,$$

$$C_3 = [1, \ -1, \ -1, \ 1, \ -1, \ 1, \ 1, \ -1] \otimes I_6.$$

Without loss of generality, assume that X is now of full column rank p, $p < 42$. The model $C \log(A\pi) = X\beta$ can be expressed by using the equivalent specification

$$U^{\mathrm{T}} C \log(A\pi) = 0 \tag{9}$$

where U is a $42 \times q$ matrix whose columns span the subspace orthogonal to the range space of X, i.e. $U^{\mathrm{T}} X = 0$ and $\mathrm{rank}(U) = q = 42 - p$.

Let $\mathscr{L}$ denote the subspace generated by the columns of X and let $\mathscr{L}^{\perp}$ denote its orthogonal complement. In addition, let W be some $42 \times q$ matrix with linearly independent columns. The matrix U can then be obtained as the orthogonal projection of W onto $\mathscr{L}^{\perp}$ along $\mathscr{L}$, i.e.

$$U = (I - P)W \tag{10}$$

where

$$P = X(X^{\mathrm{T}} X)^{-1} X^{\mathrm{T}} \tag{11}$$

is the orthogonal projector which projects onto $\mathscr{L}$ along $\mathscr{L}^{\perp}$ (Rao (1973), pages 46–48).

Recall from Section 2 that the maximum likelihood estimates of π can be obtained by maximizing the kernel of the Poisson log-likelihood

$$f(\pi) = (\log \mu^{\mathrm{T}})n - \mu^{\mathrm{T}}\mathbf{1}$$

subject to the constraints

$$h(\pi) = U^{\mathrm{T}}C \log(A\pi) = 0,$$

where $\{n_{ijkl}\} \sim$ ind mult$(n_{+++l}, \{\pi_{ijkl}\})$, $l = 1, \ldots, 6$, and the μ_{ijkl} are the corresponding expected cell counts.

Following Haber (1985), define the objective function

$$Q(\theta) = f(\pi) - \lambda^{\mathrm{T}} h(\pi), \tag{12}$$

where $\theta^{\mathrm{T}} = (\pi^{\mathrm{T}}, \lambda^{\mathrm{T}})$ and λ is a $q \times 1$ vector of Lagrange multipliers. The constrained maximum likelihood estimates are obtained by solving the system of equations

$$g(\theta) = \frac{\partial Q(\theta)}{\partial \theta} = \begin{pmatrix} F(\pi) - H(\pi)\lambda \\ -h(\pi) \end{pmatrix} = 0$$

where

$$F(\pi) = \frac{\partial f(\pi)}{\partial \pi} = \left(\frac{\partial f(\pi)}{\partial \pi_1}, \ldots, \frac{\partial f(\pi)}{\partial \pi_{48}} \right)^{\mathrm{T}}$$

is the 48×1 vector

$$F(\pi) = \mathrm{diag}(\pi)^{-1}(n - \mu) \tag{13}$$

and

$$H(\pi) = \frac{\partial h(\pi)^{\mathrm{T}}}{\partial \pi} = \left(\frac{\partial h_1(\pi)}{\partial \pi}, \ldots, \frac{\partial h_q(\pi)}{\partial \pi} \right)$$

is the $48 \times q$ matrix

$$H(\pi) = A^{\mathrm{T}} \mathrm{diag}(A\pi)^{-1}C^{\mathrm{T}}U. \tag{14}$$

Let the Hessian matrix be denoted by

$$G(\theta) = \frac{\partial^2 Q(\theta)}{\partial \theta \, \partial \theta^{\mathrm{T}}} = \begin{pmatrix} B & -H \\ -H^{\mathrm{T}} & 0 \end{pmatrix}$$

where $H = H(\pi)$, B is the 48×48 matrix

$$B = \frac{\partial F(\pi)}{\partial \pi^{\mathrm{T}}} - \frac{\partial H(\pi)}{\partial \pi^{\mathrm{T}}} (\lambda \otimes I_{48}).$$

Let $\eta = A\pi$ and $\kappa = C^{\mathrm{T}}U\lambda$. The matrix B can then be written as

$$B = \mathrm{diag}(\zeta) - A^{\mathrm{T}} \mathrm{diag}(\upsilon)A \tag{15}$$

where $\zeta_{(48 \times 1)} = \{-n_{ijkl}/\pi_{ijkl}^2\}$ and $\upsilon_{(156 \times 1)} = \{-\kappa_s/\eta_s^2\}$.

The Newton–Raphson iterative scheme is defined by

$$\theta^{(t+1)} = \theta^{(t)} - G^{-1}(\theta^{(t)}) \, g(\theta^{(t)}) \tag{16}$$

for $t = 0, 1, \ldots$, where $\lambda^{(0)} = 0$ and $\pi^{(0)} = \tilde{\pi}$, $\tilde{\pi}_{ijkl} = n_{ijkl}/n_{+++l}$ can be chosen as convenient initial values. The inverse of the Hessian matrix is obtained as follows (Aitchison and Silvey, 1958):

$$G^{-1}(\theta) = \begin{pmatrix} R & S \\ S^{\mathrm{T}} & T \end{pmatrix}$$

where

$$\left. \begin{aligned} T &= -(H^{\mathrm{T}}B^{-1}H)^{-1}, \\ S &= B^{-1}HT, \\ R &= B^{-1} + SH^{\mathrm{T}}B^{-1}. \end{aligned} \right\} \tag{17}$$

Recall that the maximum likelihood estimate of β can be derived from the maximum likelihood estimate of π through

$$\hat{\beta} = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}C\log(A\hat{\pi}).$$

Under the assumed sampling scheme for the crossover experiment, Lang (1992) showed that as $n_{+++l} \to \infty$ in such a way that $n_{+++l} \sim n_* = \min\{n_{+++1}, \ldots, n_{+++6}\}$,

$$n_*^{1/2}(\hat{\pi} - \pi) \xrightarrow{\mathrm{L}} N\{0, \Sigma(\pi)\},$$

where

$$\Sigma(\pi) = D(\pi) - \bigoplus_{l=1}^{6} \pi_l \pi_l^{\mathrm{T}} - D(\pi)H(H^{\mathrm{T}}D(\pi)H)^{-1}H^{\mathrm{T}}D(\pi),$$

$$D(\pi) = \mathrm{diag}\{[\mathrm{diag}(\pi_l)], \ l = 1, \ldots, 6\},$$

$$D(\pi) - \bigoplus_{l=1}^{6} \pi_l \pi_l^{\mathrm{T}} = \mathrm{diag}\{[\mathrm{diag}(\pi_l) - \pi_l \pi_l^{\mathrm{T}}], \ l = 1, \ldots, 6\},$$

and π_l is the 8×1 vector of cell probabilities for the lth group. The delta method can then be used to show that under the same limit conditions the asymptotic distribution of $n_*^{1/2}(\hat{\beta} - \beta)$ is likewise multivariate Gaussian with mean 0 and covariance matrix

$$\Sigma(\beta) = \left(\frac{\partial \beta}{\partial \pi}\right) \Sigma(\pi) \left(\frac{\partial \beta}{\partial \pi}\right)^{\mathrm{T}}. \tag{18}$$

The required derivative is

$$\frac{\partial \beta}{\partial \pi} = K \, \mathrm{diag}(A\pi)^{-1}A, \tag{19}$$

where $K = (X^{\mathrm{T}}X)^{-1}X^{\mathrm{T}}C$. Thus,

$$\Sigma(\beta) = K \, \mathrm{diag}(A\pi)^{-1}A \, \Sigma(\pi)A^{\mathrm{T}} \, \mathrm{diag}(A\pi)^{-1}K^{\mathrm{T}}. \tag{20}$$

On convergence of the fitting algorithm, one can estimate the asymptotic variance-covariance matrices $\Sigma(\pi)$ and $\Sigma(\beta)$ by replacing the vector π with the consistent estimator $\hat{\pi}$.

References

Agresti, A. (1990) *Categorical Data Analysis*. New York: Wiley.

Aitchison, J. and Silvey, S. D. (1958) Maximum-likelihood estimation of parameters subject to restraints. *Ann. Math. Statist.*, **29**, 813–828.

Becker, M. P. (1989) On the bivariate normal distribution and association models for ordinal categorical data. *Statist. Probab. Lett.*, **8**, 435–440.

Becker, M. P. and Balagtas, C. C. (1993) Marginal modeling for binary cross-over data. *Biometrics*, **49**, 997–1009.

Dale, J. R. (1986) Global cross-ratios for bivariate, discrete, ordered responses. *Biometrics*, **42**, 909–917.

Fitzmaurice, G. M. and Laird, N. M. (1993) A likelihood-based method for analysing longitudinal binary responses. *Biometrika*, **80**, 141–151.

Gokhale, D. V. (1973) Iterative maximum likelihood estimation for discrete distributions. *Sankhya* B, **35**, 293–298.

Goodman, L. A. (1981) Association models and the bivariate normal for contingency tables with ordered categories. *Biometrika*, **68**, 347–355.

—— (1985) The analysis of cross-classified data having ordered and/or unordered categories: association models, correlation models, and asymmetry models for contingency tables with or without missing entries. *Ann. Statist.*, **13**, 10–69.

—— (1991) Measures, models, and graphical displays in the analysis of cross-classified data (with discussion). *J. Am. Statist. Ass.*, **86**, 1085–1138.

Haber, M. (1985) Log-linear models for correlated marginal totals of a contingency table. *Communs Statist. Theory Meth.*, **14**, 2845–2856.

Haber, M. and Brown, M. B. (1986) Maximum likelihood methods for log-linear models when expected frequencies are subject to linear constraints. *J. Am. Statist. Ass.*, **81**, 477–482.

Holland, P. W. and Wang, Y. J. (1987) Dependence functions for continuous bivariate densities. *Communs Statist. Theory Meth.*, **16**, 863–876.

Jones, B. and Kenward, M. G. (1987) Modelling binary data from a three-period cross-over trial. *Statist. Med.*, **6**, 1607–1619.

—— (1989) *Design and Analysis of Cross-over Trials*. London: Chapman and Hall.

Lang, J. B. (1992) On model fitting for multivariate polytomous response data. *PhD Thesis*. Department of Statistics, University of Florida, Gainesville.

Liang, K.-Y. and Zeger, S. L. (1986) Longitudinal data analysis using generalized linear models. *Biometrika*, **73**, 13–22.

Liang, K.-Y., Zeger, S. L. and Qaqish, B. (1992) Multivariate regression analyses for categorical data (with discussion). *J. R. Statist. Soc.* B, **54**, 3–40.

Matthews, J. N. S. (1988) Recent developments in cross-over designs. *Int. Statist. Rev.*, **56**, 117–127.

Palmgren, J. (1981) The Fisher information matrix for log-linear models arguing conditionally in the observed explanatory variables. *Biometrika*, **68**, 563–566.

Prentice, R. L. (1988) Correlated binary regression with covariates specific to each binary observation. *Biometrics*, **44**, 1022–1048.

Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*, 2nd edn. New York: Wiley.

Ratkowsky, D. A., Evans, M. A. and Alldredge, J. R. (1993) *Cross-over Experiments*. New York: Dekker.

Ten Have, T. R. and Becker, M. P. (1994) Random effects models for repeated categorical responses. Unpublished.

Wedderburn, R. W. M. (1974) Generalized linear models specified in terms of linear constraints. *J. R. Statist. Soc.* B, **36**, 449–454.