STATISTICS IN MEDICINE
Statist. Med. 2005; **24**:2317–2334
Published online in Wiley InterScience (www.interscience.wiley.com). DOI: 10.1002/sim.2122

A multistate Markov chain model for longitudinal, categorical quality-of-life data subject to non-ignorable missingness

Bernard F. Cole[1,*,†], Marco Bonetti[2,3], Alan M. Zaslavsky[4] and Richard D. Gelber[5]

[1]*Department of Community and Family Medicine, Section of Biostatistics and Epidemiology, Dartmouth College Medical School, Lebanon, NH 03756, U.S.A.*
[2]*Istituto di Metodi Quantitativi, Universitá Bocconi, Viale Isonzo, 25, Milano, Italy*
[3]*Department of Biostatistics, Harvard School of Public Health, 655 Huntington Avenue, Boston, MA 02115, U.S.A.*
[4]*Department of Health Care Policy, Harvard Medical School, 180 Longwood Avenue, Boston, MA 02115, U.S.A.*
[5]*Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, 44 Binney Street, Boston, MA 02115, U.S.A.*

SUMMARY

Quality-of-life (QOL) is an important outcome in clinical research, particularly in cancer clinical trials. Typically, data are collected longitudinally from patients during treatment and subsequent follow-up. Missing data are a common problem, and missingness may arise in a non-ignorable fashion. In particular, the probability that a patient misses an assessment may depend on the patient's QOL at the time of the scheduled assessment. We propose a Markov chain model for the analysis of categorical outcomes derived from QOL measures. Our model assumes that transitions between QOL states depend on covariates through generalized logit models or proportional odds models. To account for non-ignorable missingness, we incorporate logistic regression models for the conditional probabilities of observing measurements, given their actual values. The model can accommodate time-dependent covariates. Estimation is by maximum likelihood, summing over all possible values of the missing measurements. We describe options for selecting parsimonious models, and we study the finite-sample properties of the estimators by simulation. We apply the techniques to data from a breast cancer clinical trial in which QOL assessments were made longitudinally, and in which missing data frequently arose. Copyright © 2005 John Wiley & Sons, Ltd.

KEY WORDS: generalized logit model; incomplete data; informative missing data; logistic regression; proportional odds model; repeated measurements

*Correspondence to: Bernard F. Cole, Department of Community and Family Medicine, Dartmouth Medical School, 7927 Rubin Building, Dartmouth-Hitchcock Medical Center, Lebanon, NH 03756, U.S.A.
†E-mail: bernard.cole@dartmouth.edu

Contract/grant sponsor: National Cancer Institute; contract/grant numbers: CA-06516, CA-75362
Contract/grant sponsor: National Institute of Allergy and Infectious Diseases; contract/grant number: 1 R01 AI51164

Received March 2004
Accepted October 2004

1. INTRODUCTION

Quality-of-life (QOL) is an important outcome in clinical research. Typically, QOL data are collected longitudinally from patients at prespecified time points during treatment and follow-up, and missing data are a common problem. The probability a patient misses an assessment may be related to the patient's QOL at the scheduled assessment time [1]. Patients who are very sick, and presumably have low QOL, may miss assessments because of the burden in completing them. On the other hand, a patient who is doing well may skip a clinic appointment and thus miss the corresponding assessment. When the probability of missing an observation is a function of the unobserved variable, the data are said to be affected by non-ignorable missingness, as defined by Little and Rubin [2].

The International Breast Cancer Study Group (IBCSG) conducted a randomized clinical trial in 1212 postmenopausal women with node-positive breast cancer to evaluate chemoendocrine therapies [3, 4]. In this study, patients completed a QOL questionnaire at the beginning of treatment and at various time points during and after treatment. Many patients missed assessments for reasons that may be related to their QOL at the scheduled assessment time. Motivated by this issue, we developed a model for longitudinal QOL data subject to non-ignorable missingness. Our goal was to evaluate the treatment effects while accounting for the possibly informative missing data.

Several authors have considered models for univariate categorical outcomes subject to non-ignorable missingness. Nordheim [5] considered the problem of estimating the prevalence rate for a genetic abnormality given fixed missingness probabilities. Fay [6] and Baker and Laird [7] introduced a class of log-linear models that allow for non-ignorable missingness. They used the EM algorithm to compute maximum likelihood estimates of the model parameters. Park and Brown [8] and Green and Park [9] extended the log-linear modelling approach for contingency tables within a Bayesian framework. Bonetti *et al.* [10] proposed a method-of-moments estimation procedure that can be used in some situations where maximization of the likelihood is problematic. Additional literature considered longitudinal models for binary outcomes that account for non-ignorable missingness. Fitzmaurice *et al.* [11] described the bias that can result from non-random drop-out using various estimators based on generalized estimating equations (GEE). Ten Have *et al.* [12] proposed mixed effects logistic regression models for longitudinal binary response data with informative drop-out. In a later paper, Ten Have *et al.* [13] extended the binary models to ordinal response data with multiple causes of informative drop-out. Huang and Brown [14] proposed a Markov chain model for longitudinal categorical data subject to non-ignorable missingness. Their model accommodates intermittent missingness in addition to monotone missingness (dropout), but it does not provide for the inclusion of covariates. Troxel *et al.* [15] considered longitudinal, continuous data with non-ignorable missingness. Their model involves a Markov assumption regarding the correlation structure for the longitudinal outcomes. Rotnitzky *et al.* [16] developed a class of semi-parametric marginal regression models involving non-ignorable missingness. Their methods treat the non-response model parameter as known and allow it to vary over its range, thus producing a sensitivity analysis. Fairclough [17] described multiple imputation techniques for non-ignorable missing data from longitudinal QOL studies.

In this paper, we develop a Markov chain model for the analysis of longitudinal categorical QOL measures subject to non-ignorable missingness, and we fit this model to the IBCSG data. We model transitions between QOL states with a Markov chain having transition probabilities

that may depend on (possibly time-varying) covariates through generalized logit models or proportional odds models [18]. Logistic regression is used to model the conditional probability of observing a measurement, given the actual value. Estimation is by maximum likelihood, summing over all possible values of the missing measurements. The main advantage of the model over existing methods is that it allows two or more QOL states, while accommodating both intermittent, informative missingness and covariate effects.

Other authors have analysed the IBCSG data under different assumptions, and these analyses serve as useful comparisons to the present analysis. Hürny *et al.* [3] used analysis of variance to compare the treatment groups at each assessment time, finding that patients undergoing chemotherapy tended to have lower QOL, and that QOL tended to improve over time. Bonetti *et al.* [10] applied a method-of-moments estimation procedure to the data, assuming the same missing data mechanism but different distributions of the QOL scores across the treatment groups. At each time point separately, they estimated parameters associated with non-ignorable missingness as well as the probability of each QOL state. Their results also showed that QOL scores tended to increase over time and tended to be lower for patients undergoing chemotherapy.

In Section 2, we describe the model for the transition probabilities of the Markov chain as well as the non-ignorable missingness mechanism. Maximum likelihood estimation of the model parameters is described in Section 3. Section 4 presents simulation studies of the finite sample properties of the model estimators. In Section 5, we fit the model to the IBCSG QOL data. Lastly, Section 6 discusses advantages and limitations of the model along with possible extensions.

2. THE MODEL

Let $Y_t,\ t=1,\ldots,T$ denote a Markov chain with J states.

Define the transition probabilities by

$$p_{ij}(\mathbf{x}_t)=\Pr(Y_t=j|Y_{t-1}=i,\mathbf{x}_t)$$

where $i,j=1,\ldots,J$, and $\mathbf{x}_t$ is a vector of possibly time-varying covariates. We assume that $\Pr(Y_1=j)$, the likelihood for the initial state, does not depend on any of the parameters associated with the transition probabilities. Additionally, we assume that the initial state is always observed.

Let $R_t,\ t=2,\ldots,T$ denote observation indicators for the Markov chain. That is,

$$R_t=\begin{cases}1 & \text{if } Y_t \text{ is observed}\\ 0 & \text{otherwise}\end{cases}$$

For $t=2,\ldots,T$, we define the conditional probability that Y_t is observed given that $Y_t=j$ by $q_j(\mathbf{z}_t)=\Pr(R_t=1|Y_t=j,\mathbf{z}_t)$, where $\mathbf{z}_t$ is a vector of covariates, possibly time-varying and including any or all of the covariates in $\mathbf{x}_t$.

Under the above assumptions, we have for all i, j, r and t,

$$\Pr(Y_t=j,R_t=r|Y_{t-1}=i,\mathbf{x}_t,\mathbf{z}_t)=p_{ij}(\mathbf{x}_t)q_j(\mathbf{z}_t)^r\{1-q_j(\mathbf{z}_t)\}^{1-r} \quad (1)$$

If the response categories are not ordered, the transition probabilities can be modelled using generalized logits. Using the Jth category as a reference, the logits can be parameterized as

$$\log\left\{\frac{p_{ij}(\mathbf{x}_t)}{p_{iJ}(\mathbf{x}_t)}\right\} = \beta_{ij}'\mathbf{x}_t, \quad j = 1,\ldots,J-1 \tag{2}$$

where each β_{ij} is a vector of parameters. Under the model defined by (1) and (2), the transition probabilities are given by

$$p_{ij}(\mathbf{x}_t) = \frac{\exp(\beta_{ij}'\mathbf{x}_t)}{\sum_{\ell=1}^{J}\exp(\beta_{i\ell}'\mathbf{x}_t)}$$

where we set $\beta_{iJ} \equiv 0$ for all i to ensure identifiability. We refer to (1) and (2) as the *generalized logit model.*

The *proportional odds model* is appropriate for ordinal outcomes. Defining the cumulative probabilities, $w_{ij}(\mathbf{x}_t) = p_{i1}(\mathbf{x}_t) + \cdots + p_{ij}(\mathbf{x}_t)$, the cumulative logits can be parameterized as

$$\log\left\{\frac{w_{ij}(\mathbf{x}_t)}{1-w_{ij}(\mathbf{x}_t)}\right\} = \alpha_{ij} + \beta_i'\mathbf{x}_t, \quad j = 1,\ldots,J-1 \tag{3}$$

where each α_{ij} is an intercept parameter, and each β_i is a vector of slope parameters. The α_{ij} must be non-decreasing in j for fixed i to ensure monotonicity among the cumulative probabilities $w_{ij}(\mathbf{x}_t)$. Under (3) the transition probabilities are

$$p_{ij}(\mathbf{x}_t) = \frac{1}{1+\exp(-\alpha_{ij}-\beta_i'\mathbf{x}_t)} - \frac{1}{1+\exp(-\alpha_{i,j-1}-\beta_i'\mathbf{x}_t)}$$

where we define $\alpha_{i0} \equiv -\infty$ and $\alpha_{iJ} \equiv +\infty$.

For both models, we can further model the probabilities of observing the outcomes $Y_2,\ldots,Y_T$ using logistic regression. That is, we let

$$\log\left\{\frac{q_j(\mathbf{z}_t)}{1-q_j(\mathbf{z}_t)}\right\} = \eta_j'\mathbf{z}_t$$

where η_j is a vector of parameters, $j = 1,\ldots,J$. Equivalently,

$$q_j(\mathbf{z}_t) = \frac{1}{1+\exp(-\eta_j'\mathbf{z}_t)}$$

3. INFERENCE

3.1. Parameter estimation

Estimation of the model parameters can be achieved by maximizing the full likelihood via the Newton–Raphson algorithm and enumeration of all possible complete-data chains. Let $\{y_{it}\}, i = 1,\ldots,n, t = 1,\ldots,T_i$, denote the observations from n independent Markov chains, where T_i denotes the length of chain i (including missing elements). Let r_{it} denote the observation indicator for y_{it}, and let $\mathbf{x}_{it}$ and $\mathbf{z}_{it}$ denote the corresponding covariate vectors, which we assume are always observed. To simplify notation, we let $\mathbf{y}_i$ denote the ith chain of data, $\mathbf{y}$

denote all of the chains of data, and θ denote the vector of all model parameters (i.e. all α, β and η parameters). Finally, we let $\mathbf{y}_i^{\text{Obs}}$ and $\mathbf{y}_i^{\text{Miss}}$ denote the observed and missing components of $\mathbf{y}_i$, respectively.

We first consider the complete-data likelihood for a single chain. This represents the likelihood under the assumption that all missing elements of a chain are known, as well as the observation indicators. For chain $\mathbf{y}_i$, the conditional likelihood for the complete-data chain given the initial state is

$$L^c(\theta; \mathbf{y}_i) = \prod_{t=2}^{T_i} p_{y_{i(t-1)}, y_{it}}(\mathbf{x}_{it}) q_{y_{it}}(\mathbf{z}_{it})^{r_{it}} \{1 - q_{y_{it}}(\mathbf{z}_{it})\}^{1-r_{it}}$$

The full likelihood for a single chain is

$$\sum_{\mathbf{y}_i^{\text{Miss}}} \Pr(Y_{i1} = y_{i1}) L^c(\theta; \mathbf{y}_i) \tag{4}$$

where the summation is over all possible values for the missing components of $\mathbf{y}_i$. Under the assumption that the initial state is always observed, (4) can be written as

$$\Pr(Y_{i1} = y_{i1}) L(\theta; \mathbf{y}_i^{\text{Obs}}) \tag{5}$$

where

$$L(\theta; \mathbf{y}_i^{\text{Obs}}) = \sum_{\mathbf{y}_i^{\text{Miss}}} L^c(\theta; \mathbf{y}_i)$$

The combined likelihood (5) can then be factored into components for the initial state and the subsequent transitions. Under the assumption that the parameters of these components are distinct, we can ignore the initial-state likelihood when we estimate the parameters for the state transitions, which are those of interest.

The total log-likelihood for the sample of chains is $\ell(\theta; \mathbf{y}^{\text{Obs}}) = \sum_{i=1}^{n} \ln L(\theta; \mathbf{y}_i^{\text{Obs}})$. In Appendix A, we compute the first and second derivatives of $\ell(\theta; \mathbf{y}^{\text{Obs}})$ for the purpose of applying the Newton–Raphson algorithm.

After fitting the model, the parameter estimates can be used to impute missing data using the predictive distribution for each chain, $\Pr(\mathbf{Y}_i^{\text{Miss}} | \mathbf{Y}_i^{\text{Obs}}, \hat{\theta}) = L^c(\hat{\theta}; \mathbf{y}_i)/L(\hat{\theta}; \mathbf{y}_i^{\text{Obs}})$. For example, one could draw from this distribution in a multiple imputation analysis, or if appropriate, the mean of the predictive distribution can be substituted for the missing data items.

It may be appropriate to model missing observations which occur at the end of a chain differently from missing observations that occur between two non-missing observations. Missingness at the end of a chain may be due to loss of follow-up for reasons unrelated to QOL (study dropout). A variety of models representing alternative assumptions can be obtained by varying the way in which the likelihood incorporates these missing observations. Summing over all missing observations following dropout assumes that these outcomes follow the same Markov model and that missingness continues to be related to the actual, unobserved QOL in the same way as prior to dropout. Conversely, if one assumes that the transition mechanisms and observation probabilities following dropout are independent of and unrelated to pre-dropout mechanisms, one would sum only over missing observations occurring prior to the last non-missing observation. An intermediate assumption, that we follow in the analysis of

the IBCSG data, is that the first few missed reports following dropout arise similarly to pre-dropout missing observations, but henceforth a different process leads to permanent dropout. The corresponding likelihood is obtained by summing over a shortened range of post-dropout observations.

We use a robust ('sandwich') estimator of the asymptotic variance matrix of the maximum likelihood estimates [19]. Our model assumes that the successive transitions for a given subject are independent, but in fact there might be some unmodelled dependence. The sandwich estimator corrects for the effects of clustering of transitions within subjects. Let $\mathbf{I}(\hat{\theta})$ denote the observed information matrix evaluated at the maximum likelihood estimate $\hat{\theta}$ of θ, and let $s_i(\hat{\theta})$ denote the score vector for subject i evaluated at $\hat{\theta}$. Both $\mathbf{I}(\hat{\theta})$ and $s_i(\hat{\theta})$ are obtained as byproducts of the maximization of $\ell(\theta; \mathbf{y}^{\text{Obs}})$ via the Newton–Raphson method. Then, the sandwich estimator of the covariance matrix for $\hat{\theta}$ is

$$\hat{\Sigma} = \mathbf{I}(\hat{\theta})^{-1}\mathbf{S}\mathbf{I}(\hat{\theta})^{-1} \quad \text{where } \mathbf{S} = \sum_{i=1}^{n} s_i(\hat{\theta})s_i(\hat{\theta})'$$

3.2. Model selection and model fit

The proposed models require a possibly large number of parameters, but more parsimonious models can be obtained by imposing constraints such as setting a parameter to 0, or constraining two or more parameters to be equal. Such restrictions can be applied to the transition probability models, the missingness mechanism models, or both.

The number of intercept parameters in the proportional odds model can also be reduced. For example, we can assume that each intercept follows the model

$$\alpha_{ij} = \delta_i + \varepsilon_j \tag{6}$$

where $0 = \varepsilon_1 < \varepsilon_2 < \cdots < \varepsilon_{J-1}$. Under this model, the spacings of the cutpoints (for an underlying continuous variable) are the same for each state being exited. Ideally, the choice to restrict model parameters should have a scientific rationale. However, it is also possible to make judicious use of hypothesis tests (e.g. Wald tests) to choose among several models, as we illustrate in Section 5.

To evaluate the Markovian dependence assumption, we examine lagged values of the QOL states as predictors of future transitions. This represents a semi-Markov model. In this case, missing QOL states will lead to missing covariate values; however, these can be accommodated by replacing the missing state with its enumerated value when computing the likelihood.

4. SIMULATION STUDY OF THE ESTIMATORS

We performed a simulation study of the estimators to examine how well they perform in finite samples and when the missingness model is mis-specified. In all cases, we simulated data from a binary Markov chain with length 10. For the transition probabilities, we included a single covariate and used the parameter values $\alpha_{11} = 0.7$, $\alpha_{21} = -0.4$, $\beta_1 = 0.2$ and $\beta_2 = -0.3$ (based on the proportional odds model notation described in Section 2). The covariate was generated from a standard normal distribution. For the observation probabilities, we first considered an intercept-only model and three sets of parameter values: [$\eta_1 = 0.0$, $\eta_2 = 0.5$],

$[\eta_1 = 0.5, \eta_2 = 0.7]$, $[\eta_1 = 1.5, \eta_2 = 0.7]$, representing a range of (non-ignorable) missingness probabilities. We also performed simulations under a mis-specified model for the observation probabilities, where the observation probability depends on the current state and the missingness of the previous state. In particular,

$$\Pr(Y_t \text{ is observed}) = \frac{1}{1 + \exp(-\eta_j - \gamma R_{t-1})}$$

where we set $\gamma = 0.3$.

Sample sizes (total number of chains) were 60, 125, 250, 500, 1000 and 2000 for the properly specified model. For the mis-specified model, we simulated sample sizes of 250, 500, 1000 and 2000. For each combination of parameter values and sample sizes, we generated 1000 samples and estimated the coverage probability for the 95 per cent confidence interval corresponding to each parameter.

Table I shows the simulation results. When the model is properly specified (i.e. $\gamma = 0$), adequate coverage probabilities (at least 94 per cent for nominal 95 per cent intervals) were obtained for sample sizes of 125 or more. Under the mis-specified model, coverage probabilities were variable. As one would expect, coverage probabilities for the parameters pertaining to the observation probabilities were poor. However, confidence interval coverage associated with the effect of the covariate on transition probabilities was very good (i.e. generally above 93 per cent), especially as the amount of missingness decreased. This was observed even for the largest sample sizes, when any bias due to mis-specification becomes more noticeable.

5. ANALYSIS OF THE IBCSG DATA

We fit the model to the QOL data from the IBCSG trial. The trial randomized 1212 postmenopausal women with node-positive breast cancer into a 2×2 factorial study of chemoendocrine treatment. The four treatment groups were: tamoxifen alone for 5 years (Tam only, $n = 306$), tamoxifen plus three early single cycles of cyclophosphamide, methotrexate, and fluorouracil (CMF) on months 1, 2, and 3 (Tam+early CMF, $n = 302$); tamoxifen plus delayed single courses of CMF on months 9, 12, and 15 (Tam+delayed CMF, $n = 308$); and tamoxifen plus early and delayed CMF on months 1, 2, 3, 9, 12, and 15 (Tam+early+delayed CMF, $n = 296$). A full description of the trial and its findings is published elsewhere [3, 4].

Patients were asked to complete a QOL questionnaire at the beginning of treatment, 2 months later, then every 3 months for 2 years, and at 1 month and 6 months after disease relapse. The QOL instrument was administered in the clinic prior to chemotherapy. Following Hürny *et al.* [3], we analysed the answers to the perceived adjustment/coping question: 'How much effort does it cost you to cope with your illness?' which was assessed with a single-item linear analog self-assessment scale ranging from zero ('no effort at all') to 100 ('a great deal of effort'). We focused on the data collected during the first 18 months following randomization (a maximum of seven observations) or until disease relapse. This allowed us to evaluate QOL conditional on being free of recurrent disease.

As in previous analyses of these data [10], we used the coping measure defined by three possible values: 'good' (responses < 13), 'medium' (responses $\geqslant 13$ and < 40), and 'poor' (responses $\geqslant 40$). The three categories of QOL were defined so that each would contain about one-third of the observed scores. The rationale for categorizing the coping measure

Table I. Ninety-five per cent confidence interval coverage probabilities based on 1000 simulated datasets*.

Sample size	Model parameters					
	α_{11}	β_1	α_{21}	β_2	η_1	η_2
Correctly specified model ($\gamma = 0.0$)						
$[\eta_1 = 0.0,\ \eta_2 = 0.5]$						
60	0.931	0.937	0.919	0.946	0.937	0.916
125	0.941	0.955	0.942	0.945	0.948	0.950
250	0.947	0.947	0.940	0.943	0.941	0.949
500	0.953	0.948	0.953	0.952	0.949	0.953
1000	0.962	0.944	0.956	0.960	0.948	0.950
2000	0.945	0.955	0.942	0.973	0.951	0.949
$[\eta_1 = 0.5,\ \eta_2 = 0.7]$						
60	0.927	0.946	0.933	0.942	0.930	0.932
125	0.949	0.944	0.942	0.950	0.944	0.945
250	0.939	0.954	0.931	0.948	0.938	0.941
500	0.951	0.947	0.949	0.947	0.953	0.946
1000	0.950	0.942	0.950	0.950	0.960	0.961
2000	0.954	0.948	0.943	0.953	0.958	0.961
$[\eta_1 = 1.5,\ \eta_2 = 0.7]$						
60	0.924	0.951	0.941	0.939	0.951	0.935
125	0.956	0.956	0.943	0.941	0.950	0.953
250	0.957	0.942	0.953	0.952	0.942	0.948
500	0.953	0.932	0.960	0.953	0.948	0.956
1000	0.942	0.940	0.952	0.950	0.961	0.943
2000	0.949	0.945	0.949	0.960	0.951	0.950
Mis-specified model ($\gamma = 0.3$)						
$[\eta_1 = 0.0,\ \eta_2 = 0.5]$						
250	0.816	0.926	0.872	0.944	0.572	0.907
500	0.758	0.931	0.812	0.949	0.390	0.909
1000	0.673	0.924	0.754	0.957	0.224	0.882
2000	0.537	0.909	0.603	0.938	0.041	0.874
$[\eta_1 = 0.5,\ \eta_2 = 0.7]$						
250	0.874	0.945	0.888	0.953	0.736	0.866
500	0.882	0.940	0.902	0.948	0.669	0.844
1000	0.880	0.951	0.895	0.953	0.542	0.859
2000	0.867	0.958	0.873	0.941	0.308	0.824
$[\eta_1 = 1.5, \eta_2 = 0.7]$						
250	0.910	0.955	0.911	0.948	0.939	0.778
500	0.915	0.943	0.904	0.941	0.927	0.670
1000	0.879	0.930	0.871	0.932	0.929	0.492
2000	0.788	0.954	0.821	0.928	0.916	0.223

*In all cases, data were simulated from a binary Markov chain. Transition probabilities involved a single covariate (generated from a standard normal distribution) and used the parameter values $\alpha_{11} = 0.7$, $\alpha_{21} = -0.4$, $\beta_1 = 0.2$ and $\beta_2 = -0.3$ (based on the proportional odds model notation described in Section 2). For the observation probabilities, an intercept-only model was used with the parameter values indicated. A latent covariate, with corresponding parameter γ was added to achieve mis-specified models.

Table II. Number (per cent) of transitions between QOL states in the IBCSG clinical trial.

State exited	State entered			
	Poor	Medium	Good	Missing
All patients ($n = 974$)				
Poor	829 (51)	331 (20)	124 (8)	347 (21)
Medium	255 (18)	634 (45)	325 (23)	186 (13)
Good	81 (6)	242 (17)	891 (64)	187 (13)
Missing	162 (22)	171 (23)	154 (21)	254 (34)
Tam only ($n = 238$)				
Poor	122 (42)	67 (23)	31 (11)	72 (25)
Medium	46 (14)	163 (49)	85 (25)	40 (12)
Good	15 (4)	58 (14)	285 (68)	59 (14)
Missing	37 (20)	38 (20)	43 (23)	68 (37)
Tam + early CMF ($n = 245$)				
Poor	226 (54)	77 (18)	27 (6)	90 (21)
Medium	59 (17)	152 (43)	92 (26)	48 (14)
Good	23 (7)	68 (20)	210 (61)	42 (12)
Missing	34 (18)	52 (27)	40 (21)	67 (35)
Tam + delayed CMF ($n = 244$)				
Poor	218 (50)	87 (20)	35 (8)	92 (21)
Medium	72 (20)	158 (45)	70 (20)	53 (15)
Good	24 (8)	50 (16)	200 (64)	38 (12)
Missing	45 (23)	45 (23)	31 (16)	71 (37)
Tam + early + delayed CMF ($n = 247$)				
Poor	263 (54)	100 (21)	31 (6)	93 (19)
Medium	78 (22)	161 (44)	78 (22)	45 (12)
Good	19 (6)	66 (20)	196 (60)	48 (15)
Missing	46 (27)	36 (21)	40 (24)	48 (28)

was to accommodate the distribution of the data. In particular, substantial numbers of patients provided responses at the extreme high and low ends of the scale.

Of the 1212 randomized patients, 225 were excluded from this analysis because they did not complete the initial QOL assessment. An additional 13 patients were excluded due to disease relapse occurring before scheduled follow-up assessments. This analysis includes data from the remaining 974 patients. All types of transitions are well represented in the observed data as shown in Table II. In each group, the most probable transition from each state was to the same state, while transitions from 'good' to 'poor' or the reverse were relatively unusual.

We included the first two time points following study dropout in all calculations (see Section 3.1). The resulting mean per-patient chain length was 6.3. In 58 per cent of the chains at least one outcome was missing. The average number of missing observations in chains having some missing observations was 1.7, and the maximum was 5. Approximately 19 per cent of transitions entered an unknown state. The total number of completed-data chains enumerated for the likelihood estimation was 9174. We also examined the sensitivity of our results to the handling of post-dropout missingness by re-fitting the model after including all missing timepoints following dropout. The parameter estimates for these two cases were very

similar, and our conclusions were not affected. Therefore, we present only results from the model that includes the first two timepoints following dropout.

Because the QOL states are ordinal, we used the proportional odds model for this analysis; however, we also compared these results to those of the generalized logit model. We included chemotherapy as a time-varying covariate in the analysis, coding it as 1 if the QOL assessment occurred at a time when chemotherapy was planned (for an intention-to-treat analysis), and 0 otherwise. An indicator variable was included for the presence of 4 or more affected lymph nodes as a measure of disease severity. Age in years at study entry was included as a continuous covariate. All three covariates were used to model the transition probabilities and the observation probabilities.

We began the analysis by fitting a model that included lagged values of the QOL state as covariates in the transition models. Using Wald tests, we found that first- and second-lagged measures were significant predictors ($\chi^2 = 45.61$ with 3 d.f., $p < 0.0001$ and $\chi^2 = 13.22$ with 3 d.f., $p = 0.004$, respectively). Including a third-lagged measure did not improve model fit ($\chi^2 = 0.72$ with 3 d.f.; $p = 0.9$). Therefore, we fit a semi-Markov model with 2 lags. Table III shows results from the full proportional odds model, including all 33 parameters.

To obtain a parsimonious model, we used a backward-elimination approach with Wald tests and removed parameters that did not significantly (i.e. $p \geqslant 0.05$) differ from zero. we also tested whether the intercept parameters satisfy constraint (6) in the proportional odds model. Table IV details the process used to arrive at the final model, which is shown in Table V. We used robust estimates of the variance–covariance matrix; however, the corrected standard errors were similar to the uncorrected estimates. In general, the covariate effects were similar for the various transition types, allowing us to reduce the number of parameters by forcing equality of these parameters. With the exception of chemotherapy group, none of the covariates analysed was significantly associated with the observation probabilities.

To investigate how well the large-sample properties of the estimators hold in the context of the IBCSG data, we simulated 2000 samples of 974 chains under the estimated model shown in Table V. We generated values for the initial QOL state, the covariates, and the chain lengths from their joint empirical distribution by drawing bootstrap samples from the IBCSG data. Using the simulation results, we then estimated the 95 per cent coverage probability for each model parameter. It should be noted that this simulation study is limited by the fact that the true parameter values are unknown. Overall, coverage probabilities ranged from 93.4 to 95.9 per cent, indicating generally adequate coverage.

Results from the reduced model suggest that older women tend to have lower probability of making transitions into ‘poor’ or ‘medium’. This may be due to the better prognosis generally seen in older women with breast cancer as compared to younger patients. Chemotherapy was associated with a higher probability of transitions into ‘poor’ or ‘medium’. This effect is likely due to toxicity experienced with chemotherapy. As expected, higher values of lagged QOL scores were associated with lower probability of transitions into ‘poor’ or ‘medium’.

The analysis also suggested the presence of some non-ignorable missingness in the data. The predicted observation probability was highest for the ‘medium’ state and lowest for the ‘poor’ state. Chemotherapy was associated with an increased observation probability for the ‘poor’ state but did not significantly impact the observation probabilities for the other states. There are a number of possible explanations for this effect. First, patients undergoing chemotherapy had to make a clinic visit, and therefore, may have had a greater opportunity to fill out questionnaires (e.g. while waiting for chemotherapy infusions). Second, anticipation of adverse

Table III. Parameter estimates for the full proportional odds model applied to the IBCSG clinical trial with three QOL states: 1 = 'poor', 2 = 'medium' and 3 = 'good'.

Parameter	Variable	Estimate	(SE)	p-value
Transitions from 'poor'				
α_{11}	Intercept for 'poor'	2.2212	(0.5537)	0.0001
α_{12}	Intercept for 'medium'	3.8063	(0.5588)	<0.0001
β_{11}	CMF	0.1372	(0.1108)	0.2
β_{12}	4+nodes	0.0967	(0.1165)	0.4
β_{13}	Age	-0.0168	(0.0090)	0.06
β_{14}	1-lag QOL	-0.4392	(0.0968)	<0.0001
β_{15}	2-lag QOL	-0.0206	(0.0969)	0.8
Transitions from 'medium'				
α_{21}	Intercept for 'poor'	-0.1237	(0.6002)	0.8
α_{22}	Intercept for 'medium'	2.0642	(0.5868)	0.0004
β_{21}	CMF	0.2254	(0.1222)	0.07
β_{22}	4+nodes	0.0772	(0.1148)	0.5
β_{23}	Age	-0.0069	(0.0091)	0.4
β_{24}	1-lag QOL	-0.3233	(0.0837)	0.0001
β_{25}	2-lag QOL	-0.1376	(0.0712)	0.05
Transitions from 'good'				
α_{31}	Intercept for 'poor'	-0.2609	(0.7123)	0.7
α_{32}	Intercept for 'medium'	1.1890	(0.7179)	0.1
β_{31}	CMF	0.2582	(0.1466)	0.08
β_{32}	4+nodes	0.3727	(0.1560)	0.02
β_{33}	Age	-0.0237	(0.0112)	0.03
β_{34}	1-lag QOL	-0.3105	(0.0778)	0.0001
β_{35}	2-lag QOL	-0.1907	(0.0613)	0.002
Logits of observation probabilities for 'poor'				
η_{10}	Intercept	0.7756	(0.6680)	0.2
η_{11}	CMF	0.3649	(0.1319)	0.006
η_{12}	4+nodes	-0.1878	(0.1626)	0.2
η_{13}	Age	-0.0021	(0.0107)	0.8
Logits of observation probabilities for 'medium'				
η_{20}	Intercept	1.1954	(3.5495)	0.7
η_{21}	CMF	0.9408	(1.2667)	0.5
η_{22}	4+nodes	-2.1020	(2.3775)	0.4
η_{23}	Age	0.0434	(0.0348)	0.2
Logits of observation probabilities for 'good'				
η_{30}	Intercept	3.0957	(1.3360)	0.02
η_{31}	CMF	-0.2267	(0.1918)	0.3
η_{32}	4+nodes	1.6013	(0.5932)	0.007
η_{33}	Age	-0.0241	(0.0206)	0.2

Log-likelihood = -6045.36.

side effects may have also played a part in motivating patients to complete questionnaires. These results differ somewhat from a previous analysis of the same data set [10] which found higher levels of missing data for the 'good' QOL state. This difference is likely due to differing

Table IV. Summary of sequential Wald tests used to arrive at a final parsimonious model for the proportional odds analysis of the IBCSG clinical trial.

Model	No. pars.	Description	χ^2	d.f.	p-value
0	33	Full model	—	—	—
1	30	$\beta_{15}=\beta_{25}=\beta_{35}=0$ in Model 0	13.22	3	0.004
2	31	$\beta_{15}=\beta_{25}=\beta_{35}$ in Model 0	2.22	2	0.3
3	28	$\beta_{14}=\beta_{24}=\beta_{34}=0$ in Model 2	45.61	3	<0.0001
4	29	$\beta_{14}=\beta_{24}=\beta_{34}$ in Model 2	0.21	2	0.9
5	26	$\eta_{11}=\eta_{21}=\eta_{31}=0$ in Model 4	11.14	3	0.01
6	27	$\eta_{21}=\eta_{31}=0$ in Model 4	1.73	2	0.4
7	24	$\eta_{12}=\eta_{22}=\eta_{32}=0$ in Model 6	11.95	3	0.008
8	25	$\eta_{12}=\eta_{22}=0$ in Model 6	4.82	2	0.09
9	22	$\eta_{13}=\eta_{23}=\eta_{33}=0$ in Model 8	1.82	3	0.6
10	19	$\beta_{11}=\beta_{21}=\beta_{31}=0$ in Model 9	10.26	3	0.02
11	20	$\beta_{11}=\beta_{21}=\beta_{31}$ in Model 9	0.46	2	0.8
12	17	$\beta_{12}=\beta_{22}=\beta_{32}=0$ in Model 11	5.52	3	0.1
13	14	$\beta_{13}=\beta_{23}=\beta_{33}=0$ in Model 12	8.94	3	0.03
14	15	$\beta_{13}=\beta_{23}=\beta_{33}$ in Model 12	1.15	2	0.6
15	14	$\eta_{23}=0$ in Model 14	2.74	1	0.1
16	13	$\alpha_{12}-\alpha_{11}=\alpha_{22}-\alpha_{21}=\alpha_{32}-\alpha_{31}$ in Model 15	30.98	2	<0.0001

Table V. Parameter estimates for the reduced proportional odds model applied to the IBCSG clinical trial with three QOL states: 1 = 'poor', 2 = 'medium' and 3 = 'good'.

Parameter	Variable	Estimate	(SE)	p-value
Intercept parameters for transition probabilities				
α_{11}	'Poor'–'poor'	2.0268	(0.3080)	<0.0001
α_{12}	'Poor'–'medium'	3.5877	(0.3174)	<0.0001
α_{21}	'Medium'–'poor'	0.3419	(0.3321)	0.3
α_{22}	'Medium'–'medium'	2.5098	(0.3220)	<0.0001
α_{31}	'Good'–'poor'	−0.7611	(0.3476)	0.03
α_{32}	'Good'–'medium'	0.6729	(0.3234)	0.04
Shared coefficients for the transition probabilities				
$\beta_{11}=\beta_{21}=\beta_{31}$	CMF	0.2062	(0.0653)	0.002
$\beta_{13}=\beta_{23}=\beta_{33}$	Age	−0.0133	(0.0050)	0.007
$\beta_{14}=\beta_{24}=\beta_{34}$	1-lag QOL	−0.3545	(0.0527)	<0.0001
$\beta_{15}=\beta_{25}=\beta_{35}$	2-lag QOL	−0.1261	(0.0420)	0.003
Logits of observation probabilities for 'poor'				
η_{10}	Intercept	0.5476	(0.1067)	<0.0001
η_{11}	CMF	0.4127	(0.1187)	0.0005
Logits of observation probabilities for 'medium'				
η_{20}	Intercept	3.0287	(0.8300)	0.0003
Logits of observation probabilities for 'good'				
η_{30}	Intercept	1.7973	(0.1552)	<0.0001

Log-likelihood = − 6064.33.

assumptions used in the modelling as well as a lack of covariate adjustment in the earlier analysis.

Results for the generalized logit model were similar, and we omit them for brevity. The choice of the 'better' model is best based on interpretability, and here, the proportional odds model is more useful given the ordinal nature of the data.

To summarize the results, we computed observed and predicted mean QOL states by assigning numerical scores of 1, 2 and 3 to the states 'poor', 'medium' and 'good', respectively. The observed-data mean at each follow-up time point was computed using only those patients from whom a QOL assessment was available. This represents an analysis based on assuming that missingness is completely at random. Complete-data predicted means were based on observed data where available and means of imputations for the missing observations. Figure 1 shows the observed and predicted mean QOL states across the seven assessment time points for the four treatment groups. Generally, the non-ignorable missingness model resulted in predicted mean scores that were lower than observed scores. This is consistent with the fitted model for non-response which indicates that patients with better QOL are more likely to complete QOL assessments. Note that the time trend is affected by dropout since we did not impute beyond the second timepoint after the last observation.

6. DISCUSSION

We have introduced a model for longitudinal categorical QOL data subject to non-ignorable missingness and applied the model to QOL data derived from the IBCSG clinical trial. In general, our results confirmed those of previous analyses of the IBCSG data. Of note is the finding that non-ignorable missingness exists in the data and that the missingness mechanism was affected by treatment.

The proposed model has a number of general advantages. First, it accommodates longitudinal data, which are increasingly common in the context of QOL measurement. Second, non-ignorable missingness is likely since QOL is typically self-assessed and the probability a response is observed may depend on current QOL. Third, the model accommodates covariate effects (including time-varying covariates), both for the transition probabilities and for the missingness mechanism. Accommodating time-varying covariates enhances the usefulness of the model, especially when transition probabilities may depend on an intermittently-administered treatment. Of course, in some cases, covariate values may be missing along with a missing QOL assessment (e.g. when a patient misses a clinic visit), and our model does not accommodate this kind of missingness. Nevertheless, in many clinical research settings, some covariates are collected longitudinally by health practitioners (e.g. performance status, laboratory outcomes) even when QOL data are missing.

Time-varying covariates can also be used to relax the Markovian assumption with higher-order dependencies, or to model non-stationarity. For example, lagged QOL values or the cumulative amount of time spent in each state are potential time-varying covariates. In addition, the time since last assessment could be included as a time-varying covariate if spacings between assessments are not equal.

The proportional odds form of the model may be used when categories are ordinal. The generalized logit form can be used for unordered categorical data, or when the proportional odds assumption is not appropriate. The model may be used to evaluate covariate effects or to

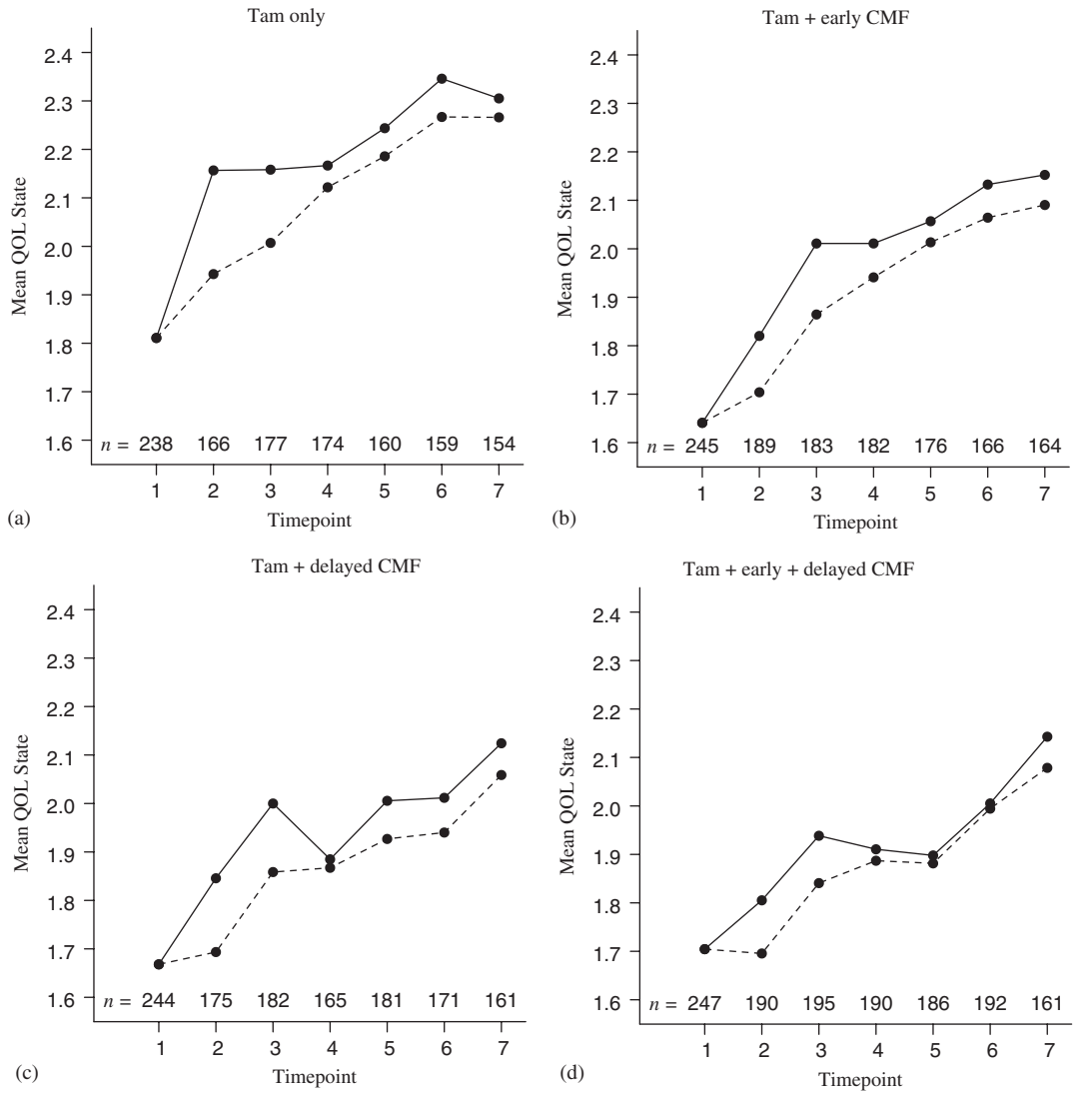

Figure 1. Observed and predicted mean quality-of-life states in the IBCSG clinical trial. The solid line shows observed means of available cases and the dashed line shows completed-data means with missing data imputed under the proportional odds model. The numbers of observations available at each time point are shown along the *x*-axis.

impute values for the missing data. The model is easily estimated using maximum likelihood. In simulation studies with moderately large samples (125 or more subjects), confidence interval coverage probabilities were adequate.

Our assumptions about the longitudinal dependence structure enable us to 'borrow strength' from adjacent observations to estimate parameters of the non-ignorable missingness

mechanism. Without any such assumptions, the data are uninformative about the non-ignorable aspect of the mechanism. In that case, we would be limited to analysis of the sensitivity of the results to various hypothesized values of the missingness parameters, as done in the semi-parametric setting [16].

Limitations of the model are that it requires a Markov assumption and that the algorithm for maximum likelihood estimation is computationally intensive. Also, the number of parameters can be large but may be reduced by appropriate restrictions. Ideally, such restrictions should be identified *a priori* based on a scientific rationale. As we have shown, Wald tests can be used to aid in model selection, but data-driven model-building approaches may suffer from a lack of statistical power given the large number of parameters involved.

The model that we have introduced is for categorical outcomes. In practice, it may be necessary to categorize a continuous QOL measure in order to fit our model. If there is concern that information is lost in this process, a finer slicing of the data can be used, as the proposed model does not limit the number of QOL states. Of course, this will result in a larger number of parameters and may make estimation and interpretation more challenging. The proposed model may suffer from a lack of identifiability in cases were missingness is either very heavy or very light and/or specific transition types are poorly represented. Additional work is needed to establish identifiability conditions along these lines.

The proposed model can be extended in a number of ways. One useful extension would allow for different types of missing data. For example, data that are missing by design, non-ignorable missing data while enrolled, and dropout could be handled in the same analysis by allowing the model for R_t to vary by reason for non-response. Other possible extensions involve including random effects or accounting explicitly for unequally spaced observations along the lines of Lee *et al.*'s [20] proposed bivariate model for markers and latent health status.

APPENDIX A: LOG-LIKELIHOOD DERIVATIVES

In this appendix, we compute the first and second derivatives of $\ell(\theta;\mathbf{y}^{\text{Obs}})$ for the purpose of applying the Newton–Raphson algorithm for maximization. To improve notation, we define the complete-data log-likelihood for a single chain by $\ell^c(\theta;\mathbf{y}_i)=\ln L^c(\theta;\mathbf{y}_i)$.

We have that

$$\frac{\partial \ell(\theta;\mathbf{y}^{\text{Obs}})}{\partial \theta_u}=\sum_{i=1}^{n}\frac{1}{L(\theta;\mathbf{y}_i^{\text{Obs}})}\left\{\sum_{\mathbf{y}_i^{\text{Miss}}}L^c(\theta;\mathbf{y}_i)\frac{\partial \ell^c(\theta;\mathbf{y}_i)}{\partial \theta_u}\right\}$$

and

$$\begin{aligned}&\frac{\partial^2 \ell(\theta;\mathbf{y}^{\text{Obs}})}{\partial \theta_u \partial \theta_v}\\&=\sum_{i=1}^{n}\frac{1}{L(\theta;\mathbf{y}_i^{\text{Obs}})}\sum_{\mathbf{y}_i^{\text{Miss}}}L^c(\theta;\mathbf{y}_i)\left[\left\{\frac{\partial \ell^c(\theta;\mathbf{y}_i)}{\partial \theta_u}\right\}\left\{\frac{\partial \ell^c(\theta;\mathbf{y}_i)}{\partial \theta_v}\right\}+\frac{\partial^2 \ell^c(\theta;\mathbf{y}_i)}{\partial \theta_u \partial \theta_v}\right]\\&\quad-\sum_{i=1}^{n}\frac{1}{L^2(\theta;\mathbf{y}_i^{\text{Obs}})}\left[\sum_{\mathbf{y}_i^{\text{Miss}}}L^c(\theta;\mathbf{y}_i)\left\{\frac{\partial \ell^c(\theta;\mathbf{y}_i)}{\partial \theta_u}\right\}\right]\left[\sum_{\mathbf{y}_i^{\text{Miss}}}L^c(\theta;\mathbf{y}_i)\left\{\frac{\partial \ell^c(\theta;\mathbf{y}_i)}{\partial \theta_v}\right\}\right]\end{aligned}$$

Therefore, we require the first and second derivatives of $\ell^c(\theta; \mathbf{y}_i)$. We have,

$$\frac{\partial \ell^c(\theta; \mathbf{y}_i)}{\partial \theta_u} = \sum_{t=2}^{T} \left[\frac{\partial \ln p_{y_{i(t-1)}, y_{it}}(\mathbf{x}_{it})}{\partial \theta_u} + r_{it} \frac{\partial \ln q_{y_{it}}(\mathbf{z}_{it})}{\partial \theta_u} + (1 - r_{it}) \frac{\partial \ln\{1 - q_{y_{it}}(\mathbf{z}_{it})\}}{\partial \theta_u} \right]$$

and

$$\frac{\partial^2 \ell^c(\theta; \mathbf{y}_i)}{\partial \theta_u \partial \theta_v} = \sum_{t=2}^{T} \left[\frac{\partial^2 \ln p_{y_{i(t-1)}, y_{it}}(\mathbf{x}_{it})}{\partial \theta_u \partial \theta_v} + r_{it} \frac{\partial^2 \ln q_{y_{it}}(\mathbf{z}_{it})}{\partial \theta_u \partial \theta_v} + (1 - r_{it}) \frac{\partial^2 \ln\{1 - q_{y_{it}}(\mathbf{z}_{it})\}}{\partial \theta_u \partial \theta_v} \right]$$

Further solution of these equations depends on the model being used. In the notation that follows we substitute $\mathbf{x}$ for $\mathbf{x}_{it}$, $\mathbf{z}$ for $\mathbf{z}_{it}$, ℓ for $y_{i,t-1}$, j for y_{it} and the β and η parameters for the θ's as described in the text.

For the generalized logit model, we have the following non-zero derivatives:

$$\begin{aligned}
\frac{\partial \ln p_{\ell j}(\mathbf{x})}{\partial \beta_{\ell u}} &= \mathbf{x} \left\{ 1_{\{j=u\}} - p_{\ell u}(\mathbf{x}) \right\} \\
\frac{\partial^2 \ln p_{\ell j}(\mathbf{x})}{\partial \beta_{\ell u} \partial \beta_{\ell v}} &= -\mathbf{x}\mathbf{x}' p_{\ell u}(\mathbf{x}) \left\{ 1_{\{u=v\}} - p_{\ell v}(\mathbf{x}) \right\} \\
\frac{\partial \ln q_j(\mathbf{z})}{\partial \eta_j} &= \mathbf{z}\{1 - q_j(\mathbf{z})\} \\
\frac{\partial^2 \ln q_j(\mathbf{z})}{\partial \eta_j^2} &= -\mathbf{z}\mathbf{z}' q_j(\mathbf{z})\{1 - q_j(\mathbf{z})\} \\
\frac{\partial \ln[1 - q_j(\mathbf{z})]}{\partial \eta_j} &= -\mathbf{z} q_j(\mathbf{z}) \\
\frac{\partial^2 \ln[1 - q_j(\mathbf{z})]}{\partial \eta_j^2} &= -\mathbf{z}\mathbf{z}' q_j(\mathbf{z})\{1 - q_j(\mathbf{z})\}
\end{aligned}$$

For the proportional odds model, it is first convenient to define $w_{\ell 0} \equiv 0$ and let

$$\begin{aligned}
g_{\ell j}(\mathbf{x}) &= w_{\ell j}(\mathbf{x})\{1 - w_{\ell j}(\mathbf{x})\} / p_{\ell j}(\mathbf{x}) \\
g^*_{\ell j}(\mathbf{x}) &= w_{\ell, j-1}(\mathbf{x})\{1 - w_{\ell, j-1}(\mathbf{x})\} / p_{\ell j}(\mathbf{x})
\end{aligned}$$

For all j, the derivatives of $q_j(\mathbf{z})$ are the same as in the generalized logit model. The remaining non-zero derivatives for the proportional odds model are

$$\begin{aligned}
\frac{\partial \ln p_{\ell j}(\mathbf{x})}{\partial \beta_\ell} &= \mathbf{x} \left\{ g_{\ell j}(\mathbf{x}) - g^*_{\ell j}(\mathbf{x}) \right\} \\
\frac{\partial^2 \ln p_{\ell j}(\mathbf{x})}{\partial \beta_\ell^2} &= \mathbf{x}\mathbf{x}' [g_{\ell j}(\mathbf{x})\{1 - 2w_{\ell j}(\mathbf{x})\} - g^*_{\ell j}(\mathbf{x})\{1 - 2w_{\ell, j-1}(\mathbf{x})\} - \{g_{\ell j}(\mathbf{x}) - g^*_{\ell j}(\mathbf{x})\}^2]
\end{aligned}$$

For $j = 1, \ldots, J-1$,

$$\frac{\partial \ln p_{\ell j}(\mathbf{x})}{\partial \alpha_{\ell j}} = g_{\ell j}(\mathbf{x})$$
$$\frac{\partial^2 \ln p_{\ell j}(\mathbf{x})}{\partial \alpha_{\ell j}^2} = g_{\ell j}(\mathbf{x})\{1 - 2w_{\ell j}(\mathbf{x}) - g_{\ell j}(\mathbf{x})\}$$
$$\frac{\partial^2 \ln p_{\ell j}(\mathbf{x})}{\partial \alpha_{\ell j} \partial \beta_{\ell}} = \mathbf{x} g_{\ell j}(\mathbf{x})\{1 - 2w_{\ell j}(\mathbf{x}) - g_{\ell j}(\mathbf{x}) + g^*_{\ell j}(\mathbf{x})\}$$

For $j = 2, \ldots, J$,

$$\frac{\partial \ln p_{\ell j}(\mathbf{x})}{\partial \alpha_{\ell, j-1}} = -g^*_{\ell j}(\mathbf{x})$$
$$\frac{\partial^2 \ln p_{\ell j}(\mathbf{x})}{\partial \alpha_{\ell, j-1}^2} = -g^*_{\ell j}(\mathbf{x})\{1 - 2w_{\ell, j-1}(\mathbf{x}) + g^*_{\ell j}(\mathbf{x})\}$$
$$\frac{\partial^2 \ln p_{\ell j}(\mathbf{x})}{\partial \alpha_{\ell, j-1} \partial \beta_{\ell}} = -\mathbf{x} g^*_{\ell j}(\mathbf{x})\{1 - 2w_{\ell, j-1}(\mathbf{x}) - g_{\ell j}(\mathbf{x}) + g^*_{\ell j}(\mathbf{x})\}$$

For $j = 2, \ldots, J-1$,

$$\frac{\partial^2 \ln p_{\ell j}(\mathbf{x})}{\partial \alpha_{\ell j} \partial \alpha_{\ell, j-1}} = g_{\ell j}(\mathbf{x}) g^*_{\ell j}(\mathbf{x})$$

ACKNOWLEDGEMENTS

The authors thank the International Breast Cancer Study Group (IBCSG) and especially members of its Quality of Life Committee (Christoph Hürny, Jürg Bernhard, Alan Coates, Harriet Peterson) for their technical assistance with the IBCSG data. This work was partially supported by grants from the National Cancer Institute (CA-06516 and CA-75362) and the National Institute of Allergy and Infectious Diseases (1 R01 AI51164).

REFERENCES

1. Bernhard J, Cella DF, Coates AS *et al.* Missing quality of life data in cancer clinical trials: serious problems and challenges. *Statistics in Medicine* 1998; **17**:517–532.
2. Little RJA, Rubin DB. *Statistical Analysis with Missing Data*. Wiley: New York, 1987.
3. Hürny C, Bernhard J, Coates AS *et al.* Impact of adjuvant therapy on quality of life in women with node-positive operable breast cancer. International Breast Cancer Study Group. *Lancet* 1996; **347**:1279–1284.
4. International Breast Cancer Study Group. Effectiveness of adjuvant chemotherapy in combination with tamoxifen for node-positive postmenopausal breast cancer patients. *Journal of Clinical Oncology* 1997; **15**:1385–1394.
5. Nordheim EV. Inference from nonrandomly missing categorical data: an example from a genetic study on Turner's syndrome. *Journal of the American Statistical Association* 1984; **79**:722–780.
6. Fay RE. Causal models for patterns of nonresponse. *Journal of the American Statistical Association* 1986; **81**:354–365.
7. Baker SG, Laird NM. Regression analysis for categorical variables with outcomes subject to non-ignorable nonresponse. *Journal of the American Statistical Association* 1988; **83**:62–69 (Correction: **83**:1232).
8. Park T, Brown MB. Models for categorical data with non-ignorable nonresponse. *Journal of the American Statistical Association* 1994; **89**:44–52.

9. Green PE, Park T. A Bayesian hierarchical model for categorical data with non-ignorable nonresponse. *Biometrics* 2003; **59**:886–896.
10. Bonetti M, Cole BF, Gelber RD. A method-of-moments estimation procedure for categorical quality-of-life data with non-ignorable missingness. *Journal of the American Statistical Association* 1999; **94**:1025–1034.
11. Fitzmaurice GM, Lipsitz SR, Molenberghs G, Ibrahim JG. Bias in estimating association parameters for longitudinal binary responses with drop-outs. *Biometrics* 2001; **57**:15–21.
12. Ten Have TR, Kunselman AR, Pulkstenis EP, Landis JR. Mixed effects logistic regression models for longitudinal binary response data with informative drop-out. *Biometrics* 1998; **54**:367–383.
13. Ten Have TR, Miller ME, Reboussin BA, James MK. Mixed effects logistic regression models for longitudinal ordinal functional response data with multiple-cause drop-out from the longitudinal study of aging. *Biometrics* 2000; **56**:279–287.
14. Huang S, Brown MB. A Markov chain model for longitudinal categorical data when there may be non-ignorable non-response. *Journal of Applied Statistics* 1999; **26**:5–18.
15. Troxel AB, Harrington DP, Lipsitz SR. Analysis of longitudinal data with non-ignorable non-monotone missing values. *Applied Statistics* 1998; **47**:425–438.
16. Rotnitzky A, Robins JM, Scharfstein DO. Semiparametric regression for repeated outcomes with non-ignorable nonresponse. *Journal of the American Statistical Association* 1998; **93**:1321–1339.
17. Fairclough DL. Multiple imputation for non-random missing data in longitudinal studies of health-related quality of life. In *Statistical Methods for Quality of Life Studies*; *Design*, *Measurements and Analysis*, Mesbah M, Cole BF, Lee M-LT (eds). Kluwer: Dordrecht, 2002.
18. Agresti A. *Categorical Data Analysis* (2nd edn). Wiley: New York, 2002.
19. Huber PJ. The behaviour of maximum likelihood estimates under non-standard conditions. *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, 1967; 221–233.
20. Lee M-LT, DeGruttola V, Schoenfeld D. A model for markers and latent health status. *Journal of the Royal Statistical Society*, *Part B* 2000; **62**:747–762.

