A Method-of-Moments Estimation Procedure for Categorical Quality-of-Life Data With Nonignorable Missingness

Marco BONETTI, Bernard F. COLE, and Richard D. GELBER

Quality-of-life outcomes collected during clinical trials often have considerable amounts of missing data, which, if not appropriately accounted for, may lead to bias in inferences. We introduce a method-of-moments (MM) estimating procedure for a model designed to handle nonignorable missingness arising in categorical data measured on independent populations. The missingness mechanism is assumed to be the same across the populations. We derive necessary and sufficient conditions for the identifiability of the model and fit the model to quality-of-life data collected as part of a breast cancer clinical trial. We compare the MM estimator to the maximum likelihood estimator in a simulation study.

KEY WORDS: Identifiability; Multinomial model; Nonignorable missingness; Quality of life.

1. INTRODUCTION

The importance of studying the quality of life of patients participating in clinical trials is fully recognized, and the collection of quality-of-life (QL) data directly from patients is becoming common in clinical research (Berzon 1998). Usually, a QL instrument is administered to the patients in a study at a number of prespecified time points during both the anticipated treatment period and the follow-up period. Some studies attempt to collect QL data after a patient experiences disease recurrence or exacerbation. A common problem in the collection of QL data is that not all patients fill out all of the questionnaires, and the missingness of an observation may be related to the patient's QL at the time of the assessment (see Bernhard et al. 1998).

It is well known that nonresponse, or *missingness*, can lead to serious bias when it is not appropriately taken into account in the analysis. In particular, the probability that a measurement is missing may depend on the actual, unobserved value. Following the terminology of Little and Rubin (1987), we call this phenomenon *nonignorable* missingness. When the missingness probability is a function only of other, observed quantities, then the outcome is said to be missing at random (MAR). Otherwise, if the missingness probability is not a function of the observed or unobserved quantities, then the outcome is said to be missing completely at random (MCAR).

In this article we study the impact of nonignorable missingness on QL data from a clinical trial conducted by the International Breast Cancer Study Group (IBCSG). IBCSG Trial VII compared four chemoendocrine regimens for the treatment of postmenopausal breast cancer. As part of the trial, data were collected longitudinally on various QL domains. These data were first reported by Hürny et al. (1996) using an MAR assumption in the analysis. More recently, Fairclough, Peterson, Cella, and Bonomi (1998) evaluated data from the same trial using five models that ranged in assumptions regarding missingness (MCAR, MAR, and nonignorable). The two nonignorable models considered by Fairclough et al. were a joint mixed-effects model for the longitudinal QL measurements and the time to a censoring event such as disease progression or death, and a pattern mixture model that assumes that the distribution of observations is a mixture of distributions with different patterns of missing data. They found that results were similar for most of the models used, with the exception of the complete-case analysis, which included only the 33% of patients who completed all longitudinal assessments. In this article we expand on this evaluation using a categorical data approach. The advantage of this approach is that we can directly model the nonignorable missingness mechanism by assigning a missingness probability parameter to each of the categorical outcomes. Our model assumes that outcome probabilities for the QL categories differ among a number of independent samples but that the missingness mechanism is the same for each sample. In particular, we define three categories for QL using the IBCSG data (good, medium, and poor) and estimate outcome probabilities according to treatment group.

Several approaches to model categorical outcomes subject to nonignorable missingness have appeared in the literature. Nordheim (1984) discussed estimation of the preva-

Marco Bonetti is Research Associate, Department of Biostatistics, Harvard School of Public Health and Dana-Farber Cancer Institute, Boston, MA 02115 (E-mail: *bonetti@jimmy.harvard.edu*). Bernard F. Cole is Assistant Professor, Section of Biostatistics and Epidemiology, Dartmouth Medical School, Lebanon, NH 03756. Richard D. Gelber is Professor, Department of Biostatistical Science, Dana-Farber Cancer Institute, Harvard School of Public Health, and Harvard Medical School, Boston, MA 02115. This work was partially supported by grants from the National Cancer Institute (CA-06516, CA-23108 and CA-75362), the American Cancer Society (RPG-90-013-08-PBP), the International Breast Cancer Study Group [IBCSG; funding provided by The Swiss Cancer League, The Cancer League of Ticino, The Swedish Cancer League, The Australia-New Zealand Breast Cancer Trials Group (grants 880513, 910420, and 940892), The Australian Cancer Society, The Frontier Science and Technology Research Foundation, and The Swiss Group for Clinical Cancer Research], and the American-Italian Cancer Foundation (AICF 101-98). The authors thank the IBCSG for providing the data for this application, and in particular the IBCSG Quality-of-Life committee members J. Bernhard, A. Coates, C. Hürny, and H. Peterson for their encouragement of this research. They also thank Donald Rubin and Nan Laird for early discussions on this topic, as well as three anonymous referees and an associate editor for their constructive suggestions that helped significantly improve the article.

Journal of the American Statistical Association
December 1999, Vol. 94, No. 448, Applications and Case Studies

lence of a genetic abnormality when the probability of recording its presence or absence is a function of its existence. He assumed that the missingness probabilities are known and derived explicit estimators for the underlying distribution. He suggested using sensitivity analysis to assess the impact of an incorrect specification of the missingness probabilities. Fay (1986) and Baker and Laird (1988) introduced a class of log-linear models that allow for nonignorable missingness and used the EM algorithm to obtain the maximum likelihood estimators (MLEs) of the models' parameters. They discussed the possibility that the estimators lie on the boundary of the parameter space. Their framework allows for the fitting of different missingness models, so that a wider sensitivity analysis than the one discussed by Nordheim (1984) is possible. Park and Brown (1994) extended the log-linear modeling within a Bayesian framework to avoid boundary solutions during the estimation of the parameters, because MLEs on the boundary may not be stable.

The description of conditions for the identifiability of models for nonignorable missingness is still an open problem. In general, a categorical model may be inestimable even when it contains fewer parameters than the number of degrees of freedom in the data (see Baker 1995 for one such case). Fitzmaurice, Laird, and Zahner (1996) concentrated on this issue, describing maximum likelihood (ML) estimation for logistic models. They proposed testing for local identifiability through verification of the nonsingularity of the Fisher information matrix. Global identifiability may then be explored through an empirical procedure.

In Section 2 we introduce the IBCSG Trial VII and the QL component of the study. In Section 3 we describe the categorical model with nonignorable missingness, discuss its identifiability, and propose a new method-of-moments (MM) estimating procedure as an alternative to ML estimation. We apply these techniques to the IBCSG data and present the results in Section 4. Finally, in Section 5 we compare the MM procedure to ML estimation in a simulation study.

2. THE INTERNATIONAL BREAST CANCER STUDY GROUP CLINICAL TRIAL

Between July 1986 and April 1993, the IBCSG enrolled 1212 postmenopausal patients with node-positive breast cancer into a 2 × 2 factorial clinical trial evaluating chemoendocrine treatment versus endocrine therapy alone for postmenopausal breast cancer patients. The patients were randomized to receive one of four adjuvant regimens: tamoxifen (tam) alone for 5 years; tam plus three early single cycles of cyclophosphamide, methotrexate, and fluorouracil (CMF) on months 1, 2, and 3; tam plus delayed single courses of CMF on months 9, 12, and 15; and tam plus early and delayed CMF on months 1, 2, 3, 9, 12, and 15. Treatment groups were balanced according to age, race, type of primary surgery, number of positive nodes, and tumor size. (See International Breast Cancer Study Group 1997 for an extensive description of the trial and its findings.)

Patients were asked to complete a QL questionnaire at the beginning of treatment, 2 months after the start of treatment, then every 3 months for 2 years, and at 1 and 6 months after disease relapse. The QL questionnaire was administered in the clinic prior to chemotherapy treatment. Following Hürny et al. (1996), we analyzed the answers to the perceived adjustment/coping question: "How much effort does it cost you to cope with your illness?" (PACIS), which was assessed with a single-item linear analog self-assessment scale ranging from 0 ("no effort at all") to 100 ("a great deal of effort"). We focused on the QL data collected during the first 18 months following randomization.

Date of relapse was defined as the time when recurrent disease was diagnosed or, if confirmed later, when it was first suspected. Disease-free survival (DFS) was defined as the length of the time from the date of randomization to any relapse (including ipsilateral breast recurrence), the appearance of a second primary cancer (including contralateral breast cancer), or death, whichever occurred first. Overall survival of patients within the first 18 months was very high (around 95%) and roughly the same across treatment arms. DFS, however, did have a more substantial absolute and differential impact, and the treatment of patients was usually modified after relapse. Because the intent of the analysis was to assess QL prior to relapse and within the first 18 months, for the 175 patients (15% of the total) who relapsed prior to the 18-month assessment, we only considered observations collected before relapse. We also excluded assessments made within 3 days of the relapse, because the patient might have known about the relapse when filling out the form.

3. A MISSINGNESS MODEL FOR CATEGORICAL DATA

3.1 The Model and Identifiability

Let Y_i denote the outcome of an individual in the ith group, and let $p_{ij} = \Pr(Y_i = j)$, where $i = 1, \ldots, q$ and $j = 1, \ldots, k$; q denotes the number of independent populations; and k denotes the number of categories for the outcome variable. We call $\mathbf{P} = \{p_{ij}\}$ the matrix containing the probabilities p_{ij}. We also let $\mathbf{r} = \{r_j\}$ be the vector of probabilities,

$$r_j = \Pr(Y_i \text{ obs} | Y_i = j), \qquad j = 1, \ldots, k.$$

For each j, r_j denotes the conditional probability of observing the outcome variable given that the outcome variable has the value j. Note that r_j is independent of i, so that the missingness mechanism is the same across the independent samples. We assume throughout that $r_j > 0$ for all j.

We call the collection $\mathcal{M}$ of the q probability mass functions and the missingness mechanism $\mathbf{r}$ a $q \times k$ *model*. We say that a model $\mathcal{M}$ is *nonidentifiable* if there exists another model $\mathcal{M}^* \neq \mathcal{M}$ such that $\Pr(Y_i = j|\mathcal{M}^*)\Pr(Y_i \text{ obs}|Y_i = j, \mathcal{M}^*) = \Pr(Y_i = j|\mathcal{M})\Pr(Y_i \text{ obs}|Y_i = j, \mathcal{M})$ for all i and j (i.e., $P^*_{ij} r^*_j = P_{ij} r_j$). When $\mathcal{M}$ is nonidentifiable, $\mathcal{M}$ and $\mathcal{M}^*$ are sometimes called "observationally equivalent."

A 2×2 model is nonidentifiable if and only if the two underlying "true" probability mass functions are identical (see Elashoff and Elashoff 1974 for a proof). This result suggests that we may be able to reveal the structure of the nonignorable missingness acting on two dichotomous outcomes if the missingness mechanism is the same for both outcomes, and if the two underlying distributions are *not* identical. In the Appendix we prove the following more general result: A $q \times k$ model is nonidentifiable if and only if the column rank of the matrix $\mathbf{P}$ is not full. An immediate corollary to this result is that for $q < k$, the model is never identifiable. It also follows immediately from the proof in the Appendix that a $q \times k$ model with $q \geq k + 1$ is nonidentifiable if $(q - k + 1)$ or more of the distributions are identical. (When $q = k$, the model is nonidentifiable if two or more distributions are identical.) After derivation of this result, we became aware of the work of Glonek (1999), which contains essentially the same result for the case $k = 2$.

Observe how this last condition describes just one of the many possible cases that would cause $\mathbf{P}$ to not have full column rank, because this generally will happen whenever its number of linearly independent rows (and therefore columns) is less than q. As a simple example, consider the following 4×4 model:

$$\mathbf{P} = \begin{bmatrix} .25 & .25 & .25 & .25 \\ .40 & .10 & .20 & .30 \\ .2 & .05 & .6 & .15 \\ .10 & .40 & .30 & .20 \end{bmatrix}; \qquad \mathbf{r} = \begin{bmatrix} .20 \\ .10 \\ .30 \\ .60 \end{bmatrix}.$$

It is easy to verify that this model is observationally equivalent to the model defined by

$$\mathbf{P}^* = \begin{bmatrix} .22 & .23 & .25 & .30 \\ .35 & .09 & .20 & .36 \\ .17 & .05 & .60 & .18 \\ .09 & .37 & .30 & .24 \end{bmatrix}; \qquad \mathbf{r}^* = \begin{bmatrix} .23 \\ .11 \\ .30 \\ .50 \end{bmatrix}.$$

In fact, the two observed distributions are identical.

It should be noted that randomly dividing an outcome category into two categories will not "generate" identifiability, as the resulting matrix of probabilities would be less than full rank. In particular, for finite sample size, this procedure could be dangerous, because the observed matrix of counts might actually allow estimation of the parameters of such a meaningless model.

Also, because the inference is based on observed finite counts, one may not be able to identify a model that actually would be identifiable, simply because of the particular sample produced by random variation. Observe how in general one cannot empirically check identifiability of the model, because formal verification of the conditions described earlier requires knowledge of the true distributions. The issues of model identifiability and estimate stability are closely related, and we elaborate on this in the Appendix.

3.2 Method-of-Moments Estimation

Let n_{ij} denote the number of times that outcome j is observed from sample i, where $i = 1, \ldots, q$ and $j = 1, \ldots, k$. Letting n_i denote the total number of observed and missing observations in sample i, we have that the number of missing observations for sample i is $n_i - \sum_{j=1}^{k} n_{ij}$.

The MM estimation procedure proceeds by first estimating the quantities $p^o_{ij} = \Pr(\{Y_i = j\} \cap \{Y_i \text{ obs}\}) = \Pr(Y_i \text{ obs} | Y_i = j)\Pr(Y_i = j) = r_j\, p_{ij}$. Therefore, $p_{ij} = p^o_{ij}/r_j$. The outcome frequencies $(n_{i1}, \ldots, n_{ik}, (n_i - \sum_{j=1}^{k} n_{ij}))$ for each $i = 1, \ldots, q$ are realizations of a multinomial random variable with the $(k + 1)$ probabilities $(p^o_{i1}, p^o_{i2}, \ldots, p^o_{ik}, (1 - \sum_{j=1}^{k} p^o_{ij}))$. This is the observed model, and we can estimate each value p^o_{ij} with the corresponding MLE $\hat{p}^o_{ij} = n_{ij}/n_i$. Substituting, p^o_{ij}/r_j for p_{ij} in the formula $\sum_{j=1}^{k} p_{ij} = 1$ yields the following system of equations:

$$\begin{cases} \hat{p}^o_{11}/r_1 + \cdots + \hat{p}^o_{1k}/r_k = 1 \\ \hat{p}^o_{21}/r_1 + \cdots + \hat{p}^o_{2k}/r_k = 1 \\ \qquad\qquad \vdots \\ \hat{p}^o_{q1}/r_1 + \cdots + \hat{p}^o_{qk}/r_k = 1. \end{cases} \tag{1}$$

Let us rewrite (1) in matrix form as

$$\begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} = \mathbf{D} \begin{bmatrix} 1/r_1 \\ 1/r_2 \\ \vdots \\ 1/r_k \end{bmatrix}, \quad \text{where} \quad \mathbf{D} = [\hat{p}^o_{ij}].$$

We obtain the values $r_1, \ldots, r_k$ that minimize the objective function $||\mathbf{D}(r_1^{-1}, \ldots, r_k^{-1})' - \mathbf{1}'_q||^2$, which is the squared distance of the left side of (1) from the vector $[1, 1, \ldots, 1]'$.

If we define $\beta_j = 1/r_j - 1$, then it is easy to show that this is equivalent to solving the optimization problem $\min_{\boldsymbol{\beta}}[\mathbf{z} - \mathbf{D}\boldsymbol{\beta}]'[\mathbf{z} - \mathbf{D}\boldsymbol{\beta}]$, where $\mathbf{z} = (\mathbf{I} - \mathbf{D})\mathbf{1}$ and $\boldsymbol{\beta} \geq 0$. Standard techniques can be used, and once the solution $\tilde{\beta}$ has been found, all of the quantities r_j and p_{ij} can be obtained immediately.

Inference can be conducted by bootstrapping from the original sample. A $(1 - \alpha)100\%$ confidence interval for the parameters can be obtained by extracting the $(\alpha/2)100$th percentile and the $(1 - \alpha/2)100$th percentile from the collection of estimates obtained from the bootstrap samples. Using bootstrap confidence intervals has the advantage of guaranteeing the validity of the intervals (i.e., they have asymptotically correct coverage probabilities) even when the point estimates lie on the boundary (see Hahn 1996).

Observe that the minimization is constrained. If the model is correct and the sample size is sufficiently large, then consistency of the $\hat{p}^o_{ij}$ implies that the foregoing minimization would have the same solution if it were done without constraints; that is, it would be the least squares solution $(\mathbf{D}'\mathbf{D})^{-1}\mathbf{D}'[1, \ldots, 1]'$. Some measure of the distance between the two solutions thus might be a measure of the model's goodness of fit. Further work needs to be done in this area, because boundary solutions commonly arise, particularly in small sample sizes. (This fact was also observed in Baker and Laird 1988 when fitting log-linear models.)

As we have seen, for $q = k$, we can invert the matrix $\mathbf{D}$ and obtain a solution to the system in (1). In practice, however, inverting the matrix $\mathbf{D}$ may yield a solution outside the parameter space. To avoid this problem, the constrained minimization procedure is recommended even when $q = k$. Moreover, for $n \to \infty$, the matrix $\mathbf{D}^{-1}$ exists with probability 1, but for finite n, it might not, due to the particular sample counts observed, even if the underlying model is identifiable.

In the $q = k$ situation, the estimators obtained are a differentiable function of the MLEs for the observed multinomial counts, and it is easy to show with the aid of the delta method that they are asymptotically unbiased and normally distributed (see Agresti 1990, sec. 12.2.1). For the general case in which $q > k$, it is not clear how the method-of-moments estimates (MMEs) behave, and in Section 5 we explore their characteristics.

As a last remark, the MMEs are very easy to obtain from the table of counts. We used the function `nnls` in the package MATLAB (The MathWorks, Inc. 1998) to solve the constrained optimization; similar functions exist in many scientific packages.

3.3 Maximum Likelihood Estimation

We used a modified Newton–Raphson algorithm to obtain the ML estimates. In the Appendix we report the expressions for the log-likelihood and its first and second derivatives. The modification consists of alternately maximizing the likelihood, first with respect to the p_{ij}'s jointly, and then with respect to the r_j's jointly. We repeated these steps until the global maximum was determined. We found that this "hill-climbing" procedure was more successful than standard Newton–Raphson in converging to a solution.

ML estimation with the (modified) Newton–Raphson algorithm has the advantage that an estimate of the asymptotic covariance matrix of the parameter estimates can be obtained as a byproduct of the maximization of the likelihood. This covariance estimation, however, is not valid in the case where any of the parameter estimates lies on the boundary of the parameter space, a situation that seems to be rather common with these models. In such cases it follows that the corresponding asymptotic confidence intervals of the parameters are also not valid.

Moreover, our experience has been that ML estimation is difficult in large problems. For example, the modified Newton–Raphson procedure failed to converge in a 7×6 model (simulated data), and even in some of the 4×2 models discussed in Section 5. It may be possible to modify the algorithm to obtain convergence; for example, by using the EM algorithm or a combination of the two algorithms (as suggested in Baker 1994). The MMEs can be used to provide a sensible set of starting values for the parameters, thus making it less likely for the algorithm to converge to the wrong boundary, or to a local maximum. For the model that we consider here, the MM procedure with bootstrapped confidence intervals seems to be the simplest solution.

4. RESULTS

Following Hürny et al. (1996), we analyzed the PACIS scores collected within the first 18 months. We discretized the PACIS measure by defining three possible values: "good" (if PACIS < 13), "medium" (if ≥ 13 and < 40), and "poor" (if ≥ 40). This recoding of PACIS scores, in addition to allowing the use of a model that does not require strong distributional assumptions, has the advantage of facilitating interpretation by clinicians, because results in terms of categories of outcome are clinically meaningful.

The three categories of QL were defined using the distribution of all the observed scores, so that each would contain approximately one-third of the available data. We fit a 4×3 model for each of the seven time points. The underlying assumption is that perceived health (and thus the measured PACIS scores) may differ across the treatment groups, but the missingness mechanism must be the same for each group. We allow for differences in the missingness mechanism over time by fitting a separate model for each time point.

Table 1. Discretized PACIS Scores in the Four Treatment Arms, for Each Time Point, for 1,212 Patients

Month	PACIS	Tam only	Early CMF	Delayed CMF	Early + delayed CMF	Total (%)
		306	302	308	296	1,212 (100)
1	Poor	96	112	103	108	
	Medium	75	89	103	82	
	Good	69	46	44	57	
	Missing	64	54	57	49	224 (18.5)
	Relapsed	2	1	1	0	4 (.3)
3	Poor	43	81	67	83	
	Medium	68	79	76	79	
	Good	82	54	62	49	
	Missing	107	84	95	85	371 (30.7)
	Relapsed	6	4	8	0	18 (1.5)
6	Poor	46	61	51	78	
	Medium	74	67	89	72	
	Good	86	76	68	68	
	Missing	86	92	81	75	334 (27.7)
	Relapsed	14	6	19	3	42 (3.5)
9	Poor	40	56	70	71	
	Medium	74	77	65	80	
	Good	84	72	65	59	
	Missing	79	82	84	75	320 (26.5)
	Relapsed	29	15	24	11	79 (6.5)
12	Poor	34	56	55	69	
	Medium	57	74	92	80	
	Good	96	75	71	58	
	Missing	78	72	61	70	281 (23.3)
	Relapsed	41	25	29	19	114 (9.4)
15	Poor	26	41	52	60	
	Medium	56	70	82	71	
	Good	104	76	69	78	
	Missing	73	81	69	61	284 (23.5)
	Relapsed	47	34	36	26	143 (11.8)
18	Poor	26	38	40	42	
	Medium	59	67	76	69	
	Good	92	84	75	73	
	Missing	70	74	75	77	296 (24.5)
	Relapsed	59	39	42	35	175 (14.4)

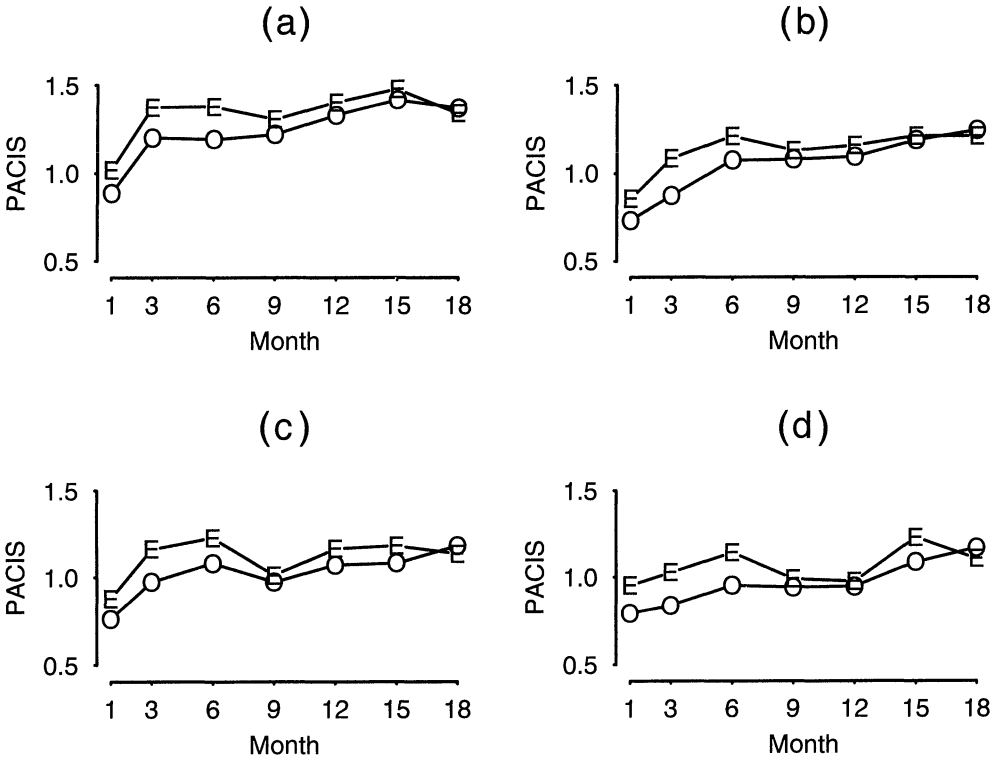

Figure 1. Means of Discretized PACIS Scores Based on Observed Data (O) and Estimated From the Model (E) for the Four Treatment Arms. Tamoxifen (Tam) Alone (a), Tam + Early CMF (b), Tam + Delayed CMF (c), and Tam + Early + Delayed CMF (d).

Based on the discussion in Section 3.2, the model is identifiable if the corresponding 4×3 matrix of probabilities has rank at least three. In fact, even though the identifiability of such models cannot be checked, we know that the number of the groups must be at least as large as the number of categories of the outcome variable. As we discuss in the Appendix, having four groups makes it possible to obtain more stable estimates for the parameters.

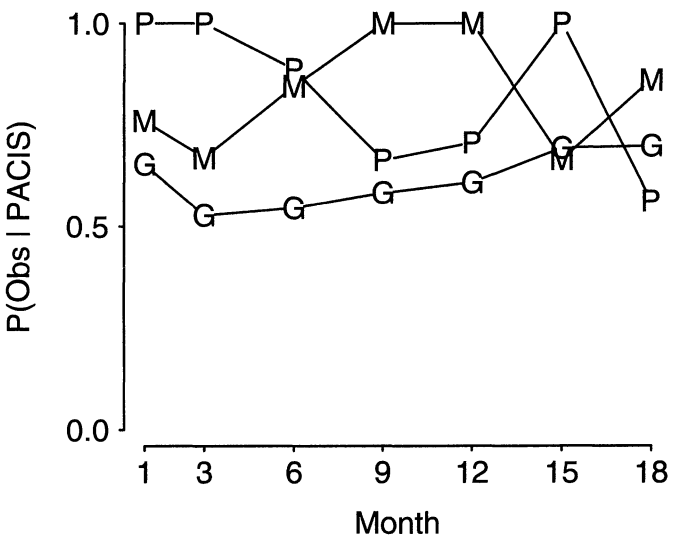

Figure 2. Probabilities of Observing the QL Conditional on the QL Score, Estimated for Each of the Seven Time Points and for the Three QL Scores Poor (P), Medium (M), and Good (G).

Table 1 summarizes, for each time point and each treatment arm, the observed number of patients falling in each of the three categories of the PACIS scores, as well as the number of missing observations and the number of assessments that were undefined due to prior relapse ("relapsed"). The amount of missing data for the different time points ranged between 18.5% and 30.7% of the total number of patients (1,212). Note that the number of undefined assessments (relapsed group) increases with time by definition, and that there is a greater proportion of such assessments in the tamoxifen alone arm compared to the chemoendocrine therapy arms.

The observed and estimated (by MM) means of the discretized PACIS scores for the patients are shown separately for each of the four treatment groups in Figure 1. The means were computed by assigning the three values 0, 1, and 2 to the three PACIS categories "poor," "medium," and "good." In this way, higher values of the discretized PACIS reflect better quality of life. The estimates obtained by ML were similar to those obtained by MM and are omitted for brevity.

We note from Figure 1 that the estimated means (E) are greater than the observed means (O). We computed the area under the curve (AUC) measurement to summarize and compare the quality of life of patients in the treatment groups. The AUC consists of the integral of the estimated outcomes for each treatment over time and is used as a summary of longitudinal outcomes (see Fairclough 1997). In the current application, the AUC can range from 0 if every assessment is in the "poor" category over all 18 months to 36 if every assessment is in the "good" category over

all 18 months. The respective values of the AUC for the four treatments a, b, c, and d were 21.5, 17.6, 18.1, and 16.4 for the "ignorable" model (O) and 23.2, 19.3, 19.5, and 18.2 for the "nonignorable" model (E). Thus consideration of the possibly nonignorable missingness adjusts all of the estimates upward.

The model provides estimates for the probabilities $\Pr(\text{obs}|\text{PACIS})$ at each time point. These probabilities, shown in Figure 2, they indicate that patients with "good" QL scores were more likely to have missing observations. This observation runs counter to the prevailing hypothesis in oncology that "poor" QL scores are associated with a greater likelihood of missing data. This, however, is based mostly on attempts to measure QL in the advanced disease setting, where patients can be "too ill" to complete the forms (see Simes et al. 1998). Troxel et al. (1998) suggested that even in the adjuvant setting, lower QL is associated with higher probability of missing assessments. Their analysis considered premenopausal breast cancer patients receiving chemotherapy and included QL measurements only for months 1, 3, and 6. The present study was on postmenopausal women who received endocrine therapy or chemoendocrine therapy and included QL assessments to 18 months. It is possible in the adjuvant setting that patients might feel "too well" to come in for clinic visits where the form would be administered.

Figure 2 also shows that some of the estimated missingness probabilities were equal to 1. This boundary-solution phenomenon is common to both estimating procedures (MM and ML), and we believe that such boundary estimates should be taken as an indication of the pattern of the missingness mechanism rather than as the exact values of its parameters.

Figure 3 shows the mean discretized PACIS scores (estimated by MM) according to treatment group. For treatment comparison, we considered the contrast between the AUC in the treatment arm involving only tamoxifen (A) and the mean of the AUCs of the three chemoendocrine therapy arms (B, C, and D). A 95% bootstrap confidence interval was constructed by resampling based on each subject's cluster of repeated measurements. The point estimate for the contrast was 4.0, and the 95% confidence interval (based on 10,000 samples) was $[2.1, 5.8]$, showing a contrast significantly different from 0 in favor of the tamoxifen alone group (with $p < .05$). In comparison, the same contrast computed using the ignorable missingness model (i.e., the observed frequencies) yielded a point estimate of 4.1 and the confidence interval $[2.6, 5.6]$. Thus, although the separate treatment AUCs differed between the "nonignorable" and the "ignorable" models as shown previously, the contrast of interest was similar between the two models.

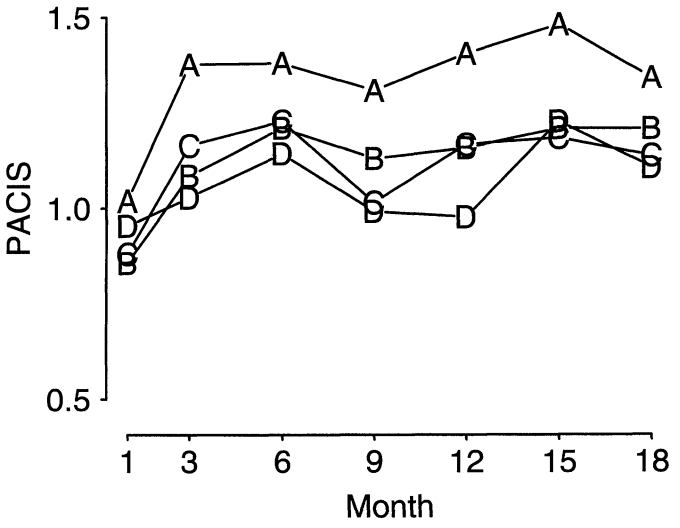

Figure 3. Means of Discretized PACIS Scores Estimated From the Model for the Four Treatment Arms Tamoxifen (Tam) Alone (A), Tam + Early CMF (B), Tam + Delayed CMF (C), and Tam + Early + Delayed CMF (D).

The analysis of the AUC for discretized PACIS scores showed that QL was significantly higher for the tamoxifen alone arm compared to the chemoendocrine therapy groups. But this analysis did not take into consideration the fact that a greater proportion of patients in the tamoxifen alone arm relapsed within 18 months compared to the other treatment groups. Quality-adjusted survival analysis is an alternative approach that simultaneously considers both QL and time to event differences between treatments. The Q-TWiST (Quality-adjusted Time Without Symptoms of disease or Toxicity of treatment) method (Gelber, Cole, Gelber, and Goldhirsch 1995) was applied to the overall survival, disease-free survival, and toxicity duration data from the

Table 2. Parameter Combinations Used in the Simulations

Parameter	p_{11}	p_{21}	p_{31}	p_{41}	r_1	r_2
1	.20	.40	.60	.80	.80	.80
2	.20	.40	.60	.80	.50	.50
3	.10	.20	.30	.40	.80	.80
4	.10	.20	.30	.40	.50	.50
5	.10	.25	.75	.90	.80	.80
6	.10	.25	.75	.90	.50	.50
7	.10	.20	.30	.90	.80	.80
8	.10	.20	.30	.90	.50	.50
9	.20	.40	.60	.80	.95	.80
10	.20	.40	.60	.80	.80	.95
11	.20	.40	.60	.80	.80	.50
12	.20	.40	.60	.80	.50	.80
13	.20	.40	.60	.80	.70	.30
14	.20	.40	.60	.80	.30	.70
15	.10	.20	.30	.40	.95	.80
16	.10	.20	.30	.40	.80	.95
17	.10	.20	.30	.40	.80	.50
18	.10	.20	.30	.40	.50	.80
19	.10	.20	.30	.40	.70	.30
20	.10	.20	.30	.40	.30	.70
21	.10	.25	.75	.90	.95	.80
22	.10	.25	.75	.90	.80	.95
23	.10	.25	.75	.90	.80	.50
24	.10	.25	.75	.90	.50	.80
25	.10	.25	.75	.90	.70	.30
26	.10	.25	.75	.90	.30	.70
27	.10	.20	.30	.90	.95	.80
28	.10	.20	.30	.90	.80	.95
29	.10	.20	.30	.90	.80	.50
30	.10	.20	.30	.90	.50	.80
31	.10	.20	.30	.90	.70	.30
32	.10	.20	.30	.90	.30	.70

Table 3. Ratio of Estimated MSE for MMEs Versus MLEs

	$n = 30$						$n = 100$						$n = 1{,}000$					
Parameter	p_{11}	p_{21}	p_{31}	p_{41}	r_1	r_2	p_{11}	p_{21}	p_{31}	p_{41}	r_1	r_2	p_{11}	p_{21}	p_{31}	p_{41}	r_1	r_2
1	.86	.98	.97	1.14	.81	.81	1.05	1.11	.98	1.17	.90	1.07	1.06	1.14	1.28	1.13	1.06	1.03
2	.76	1.01	1.02	1.07	.63	.69	1.04	1.01	1.14	1.48	.83	.95	1.00	1.12	1.37	1.47	.97	.97
3	.75	.89	.88	.90	.79	.70	.94	.92	.94	.91	.82	.78	.92	.99	.96	.99	.94	.98
4	.52	.56	.73	.77	.81	.42	.80	.79	.87	.92	.90	.69	.93	.94	1.03	.96	.97	.94
5	1.03	.96	1.39	1.57	.95	.91	1.05	1.04	1.31	1.65	1.07	.95	1.07	1.00	1.27	1.68	.88	1.06
6	.97	1.03	1.53	1.94	.89	.89	1.15	1.07	1.75	2.83	1.12	.97	1.13	1.01	1.71	2.51	1.06	1.12
7	.95	.99	1.06	1.16	.95	.90	1.19	1.11	1.02	.97	.95	1.09	.98	1.13	1.07	1.00	1.01	1.08
8	.79	1.11	1.12	1.22	.91	.88	1.09	1.05	1.23	1.12	1.00	1.12	1.11	1.07	1.16	1.25	1.01	1.07
9	1.03	1.04	.96	1.10	.95	.86	.96	1.06	1.02	1.09	.96	1.03	1.04	1.04	.92	1.04	1.04	.98
10	1.07	1.15	1.20	1.48	.87	.90	1.15	1.18	1.18	1.17	.90	1.00	.96	1.29	1.35	1.25	1.13	1.00
11	.79	.97	.94	1.09	.83	.92	.98	.87	.98	1.09	.89	.94	1.06	1.03	1.09	1.00	1.02	1.01
12	1.05	1.33	1.48	1.79	.82	.91	.89	1.35	1.76	1.75	1.08	.92	1.03	1.59	1.97	1.55	1.03	1.14
13	1.02	.97	.87	1.06	.74	.97	.99	.92	.91	.90	.84	.99	1.02	1.00	1.03	1.16	1.08	1.02
14	1.13	1.53	1.94	3.42	1.20	.87	.88	1.60	2.03	2.24	1.08	.83	1.14	1.94	2.72	2.15	1.05	1.22
15	.79	.86	.95	.92	.79	.73	.92	.95	1.01	1.05	1.02	.94	1.00	.93	.99	1.00	.97	.91
16	.89	.92	.99	1.03	.81	.75	1.01	.97	1.02	1.12	.85	.79	.96	.99	1.14	1.02	.93	.93
17	.65	.73	.80	.90	.83	.68	.95	1.00	.90	.99	.92	1.00	.99	.97	1.01	1.05	1.01	1.07
18	.61	.82	.89	1.03	.88	.58	.80	1.01	1.06	1.06	.93	.74	.95	1.10	1.07	1.04	.97	.90
19	.67	.92	.88	1.10	.83	.80	1.04	.93	.97	.97	.87	1.10	1.17	1.17	1.16	1.30	1.51	1.28
20	.48	.69	.94	1.21	1.23	.48	.74	.96	1.08	1.18	1.14	.65	1.02	1.21	1.35	1.14	.92	.95
21	.96	1.02	.98	1.10	.88	.83	.90	.98	.97	1.11	.96	1.02	.98	1.06	.88	1.33	1.07	1.04
22	1.14	1.07	1.52	1.96	.82	.95	1.00	1.16	1.51	1.69	.99	1.01	1.08	1.06	1.61	1.89	1.05	1.13
23	1.05	.99	.92	1.29	.93	.96	.91	1.16	1.12	1.46	.83	.89	.98	1.01	1.08	1.53	1.11	.93
24	.87	1.17	2.34	3.62	1.10	.78	1.00	1.28	2.36	3.83	.99	1.00	1.10	1.34	2.69	3.74	1.18	1.06
25	1.15	1.06	.90	1.14	.84	1.05	.88	1.06	.93	1.32	.95	1.05	1.03	.95	.99	1.58	.96	.89
26	.81	1.25	3.31	5.99	1.10	.71	1.10	1.35	3.43	6.63	.99	.93	1.05	1.47	4.35	6.72	1.03	1.05
27	.99	.99	.96	1.11	.91	.67	1.04	1.00	1.03	1.03	.95	.88	1.05	1.03	.90	1.00	1.05	1.03
28	.93	1.09	1.06	1.20	.94	.88	.87	1.07	1.26	1.02	1.01	.90	.95	1.04	1.19	.96	.95	.89
29	1.09	1.05	1.07	1.25	1.09	1.11	.95	1.01	.98	1.18	1.15	1.16	1.03	.95	.95	1.17	1.02	.94
30	.95	1.16	1.35	1.48	.89	.83	.97	1.20	1.44	1.18	1.05	.97	1.06	1.16	1.64	1.12	1.12	1.21
31	1.04	.99	.99	1.14	.97	1.32	.93	1.02	1.01	1.24	1.10	1.17	1.11	.95	1.01	1.12	1.08	.99
32	.77	1.12	1.52	3.08	.89	.67	.96	1.27	1.71	1.59	1.10	.97	1.06	1.37	1.81	1.26	1.02	1.09

IBCSG Trial VII. The results showed that despite the higher QL provided by tamoxifen alone, the superior control of disease relapse achieved by chemoendocrine therapy balanced the toxic effects of this treatment (Gelber et al. 1998).

The decision about the number of categories used to discretize the PACIS score is likely to affect the results, and we have experimented with two and four categories. (More than four categories would make the model not identifiable.) The cutoffs were chosen so that each category would contain roughly the same number of observations. To properly scale the results for comparison with the three-category model, the highest QL group was assigned a value of 2 and the lowest a value of 0, and intermediate levels were equally spaced.

After appropriately rescaling the y-axis to account for the different number of categories, the plots of the estimated means according to treatment were similar to those shown in Figure 3. The point estimates for the AUC contrast between tamoxifen alone and the other treatments were 4.4 for the two-category case and 3.4 for the four-category case. The 95% bootstrap confidence intervals were $[2.7, 6.3]$ and $[1.6, 5.0]$, confirming the statistical significance of the result obtained using three categories.

The model that we describe here could easily be modified to assume that the missingness mechanism behaves in a different way—for example, that it is constant over time but varying by group, or constant over time and group. In the context of our data, however, a constant missingness mechanism over time is not realistic, because the timing of QL assessment did not always coincide with treatment administration. Moreover, we wanted our model to allow for patients' changing attitudes toward QL assessment over time.

5. SIMULATION STUDY COMPARISON OF THE ESTIMATORS

We performed a simulation study to explore the properties of the MM estimators—in particular, to assess the relative efficiency of the MMEs compared to the MLEs. We considered the 4×2 model and explored an array of possibilities for the values of the four parameters defining the distributions ($p_{i1} = \Pr(Y_i = 1)$, $i = 1, \ldots, 4$ and $\mathbf{r}_j = \Pr(\mathrm{Y\,obs} | Y = j)$, $j = 1, 2$). Table 2 shows the 32 parameter combinations that define the simulation experiment. As the table shows, we considered missingness mechanisms with probabilities r_j equal to at least .3 and included some cases in which the missingness was actually ignorable (i.e., $r_1 = r_2$; rows 1–8).

For each set of parameter values, we simulated 1,000 samples of size $n = 30, 100, 1{,}000$ and fit them using both MM and ML. We then estimated the mean squared error (MSE) of the resulting estimators, and decomposed it into its bias and variance components. The bias component of

the MSE was negligible for both estimating procedures, with ratios between variance and MSE very close to 1. Table 3 shows the ratios of the estimated MSE for the MMEs versus MLEs of each of the six parameters. The results clearly indicate that the MSE of the two estimators are very similar, and that the MMEs do better than the MLEs in some cases, particularly when the true model has ignorable missingness.

We also estimated the coverage probabilities of the bootstrap-based 95% confidence intervals for MMEs of the parameters in each of the models. These estimates are shown as the nonitalic numbers in Table 4. The values are based on 500 samples simulated from each model, and the confidence intervals were constructed from 100 bootstrap samples. The coverage probabilities are all very close to .95 for large n, and are very satisfactory even for $n = 30$.

It is possible to define the naive estimator for each probability p_{i1}, $i = 1, \ldots, 4$, as the ratio between the number of observations in category 1 and the total number of nonmissing observations Y_i. Thus this naive estimate assumes ignorable missingness. One can show that the expected value of the naive estimator converges to $r_1 p_{i1}/(r_1 p_{i1} + r_2(1 - p_{i1}))$, whereas its variance converges to $\Pr(Y_i = 1|Y_i \text{ obs})(1 - \Pr(Y_i = 1|Y_i \text{ obs})E(N_i^{-1}|N_i > 0)$, where N_i follows the binomial$(n, (r_1 p_{i1} + r_2(1 - p_{i1}))$ distribution and represents the total number of nonmissing observations. It is easy to show that the asymptotic distribution of the naive estimator is normal, and thus theoretical coverage probabilities for, say, 95% confidence intervals for the parameters can be easily computed under different alternative hypotheses. These coverage probabilities are shown in italic type in Table 4 for the parameter combinations considered in the simulation study. Observe how the theoretical coverage probabilities for the first eight parameter combinations are exactly 95%, because these combinations correspond to ignorable missingness. As anticipated, the coverage probabilities for the naive estimators decrease (quite rapidly) as the sample size increases when missingness is informative (rows 9–32).

6. DISCUSSION

We have introduced a simple estimation procedure that can be used to account for the effect of nonignorable missingness in QL data collected as part of clinical trials. By assuming that the missingness mechanism is the same at all time points, we can construct a model that is likely to be identifiable.

The MM estimating procedure that we propose is to be used as an alternative to the usual ML method. For the particular 2×2 model, the resulting estimators are identical,

Table 4. Coverage Probabilities (in Percent) for 95% Confidence Intervals for the Model Parameters

	n = 30										*n = 100*										*n = 1,000*									
Parameter	p_{11}		p_{21}		p_{31}		p_{41}		r_1	r_2	p_{11}		p_{21}		p_{31}		p_{41}		r_1	r_2	p_{11}		p_{21}		p_{31}		p_{41}		r_1	r_2
1	95	*95*	94	*95*	94	*95*	93	*95*	93	94	94	*95*	95	*95*	96	*95*	93	*95*	94	92	93	*95*	94	*95*	96	*95*	94	*95*	95	92
2	97	*95*	96	*95*	94	*95*	91	*95*	95	97	95	*95*	93	*95*	94	*95*	94	*95*	96	95	92	*95*	93	*95*	93	*95*	93	*95*	93	95
3	96	*95*	94	*95*	93	*95*	94	*95*	98	95	96	*95*	95	*95*	93	*95*	93	*95*	94	95	94	*95*	93	*95*	92	*95*	94	*95*	93	93
4	97	*95*	94	*95*	94	*95*	93	*95*	98	96	98	*95*	96	*95*	94	*95*	95	*95*	98	97	92	*95*	93	*95*	95	*95*	94	*95*	94	95
5	93	*95*	92	*95*	93	*95*	92	*95*	92	92	93	*95*	94	*95*	95	*95*	94	*95*	95	95	94	*95*	94	*95*	94	*95*	94	*95*	95	93
6	93	*95*	94	*95*	93	*95*	93	*95*	95	96	95	*95*	94	*95*	94	*95*	96	*95*	94	93	94	*95*	93	*95*	95	*95*	94	*95*	94	92
7	94	*95*	93	*95*	96	*95*	91	*95*	92	93	95	*95*	95	*95*	94	*95*	93	*95*	94	93	96	*95*	93	*95*	93	*95*	94	*95*	95	95
8	94	*95*	94	*95*	95	*95*	85	*95*	94	95	92	*95*	92	*95*	93	*95*	92	*95*	95	93	94	*95*	94	*95*	94	*95*	93	*95*	94	94
9	94	*94*	93	*93*	95	*93*	90	*93*	95	91	94	*90*	93	*88*	94	*88*	93	*90*	95	93	95	*49*	93	*30*	92	*29*	94	*45*	94	94
10	92	*93*	94	*93*	92	*93*	94	*94*	90	93	94	*90*	94	*88*	92	*88*	94	*90*	94	94	92	*45*	94	*29*	95	*30*	96	*49*	94	93
11	95	*88*	94	*83*	93	*82*	91	*86*	96	94	94	*71*	93	*55*	94	*52*	94	*63*	94	94	96	*1*	93	*0*	94	*0*	93	*0*	94	93
12	96	*86*	95	*82*	95	*83*	92	*88*	93	95	97	*63*	94	*52*	92	*55*	95	*71*	92	93	94	*0*	95	*0*	94	*0*	93	*1*	95	93
13	94	*79*	93	*66*	91	*60*	86	*68*	95	92	94	*43*	93	*18*	93	*12*	93	*22*	94	94	95	*0*	94	*0*	95	*0*	95	*0*	94	94
14	95	*68*	94	*60*	96	*66*	90	*79*	93	93	97	*22*	93	*12*	94	*18*	94	*43*	94	95	95	*0*	96	*0*	93	*0*	94	*0*	94	95
15	93	*94*	93	*94*	93	*93*	95	*93*	98	87	93	*92*	94	*90*	94	*89*	95	*88*	96	91	95	*69*	94	*49*	94	*37*	94	*30*	95	94
16	93	*94*	92	*93*	95	*93*	92	*93*	93	97	95	*92*	94	*90*	94	*88*	94	*88*	96	95	95	*65*	94	*45*	93	*34*	95	*29*	95	94
17	94	*91*	93	*88*	95	*85*	92	*83*	97	91	93	*82*	94	*71*	94	*62*	92	*55*	95	93	95	*9*	95	*1*	95	*0*	93	*0*	95	95
18	95	*90*	94	*86*	94	*83*	93	*82*	96	98	95	*76*	94	*63*	96	*55*	94	*52*	94	96	94	*2*	94	*0*	95	*0*	93	*0*	93	96
19	90	*87*	92	*79*	92	*72*	93	*66*	93	88	93	*67*	93	*43*	93	*27*	95	*18*	95	93	94	*0*	92	*0*	93	*0*	93	*0*	95	94
20	93	*79*	94	*68*	95	*62*	95	*60*	94	95	95	*44*	95	*22*	95	*14*	94	*12*	91	94	95	*0*	96	*0*	96	*0*	94	*0*	93	93
21	93	*94*	92	*93*	93	*93*	88	*94*	94	91	94	*92*	94	*90*	94	*89*	94	*92*	94	94	92	*69*	93	*42*	95	*39*	94	*65*	95	93
22	91	*94*	95	*93*	92	*93*	92	*94*	93	92	94	*92*	93	*89*	94	*90*	94	*92*	93	96	94	*65*	95	*39*	96	*42*	95	*69*	94	95
23	94	*91*	94	*87*	92	*84*	90	*90*	95	96	93	*82*	93	*66*	94	*58*	93	*76*	93	92	95	*9*	94	*0*	95	*0*	94	*2*	96	94
24	94	*90*	95	*84*	94	*87*	91	*91*	92	92	92	*76*	95	*58*	93	*66*	94	*82*	94	92	95	*2*	93	*0*	92	*0*	93	*9*	92	94
25	93	*87*	95	*75*	89	*65*	91	*79*	96	96	92	*67*	91	*34*	94	*17*	92	*44*	93	92	94	*0*	94	*0*	96	*0*	95	*0*	95	96
26	92	*79*	93	*65*	95	*75*	91	*87*	95	95	93	*44*	94	*17*	96	*34*	94	*67*	93	93	94	*0*	94	*0*	92	*0*	92	*0*	93	93
27	93	*94*	92	*94*	94	*93*	91	*94*	89	91	95	*92*	95	*90*	95	*89*	90	*92*	94	94	93	*69*	94	*49*	95	*37*	95	*65*	94	94
28	91	*94*	93	*93*	91	*93*	89	*94*	93	93	93	*92*	95	*90*	95	*88*	94	*92*	92	92	94	*65*	94	*45*	95	*34*	93	*69*	95	93
29	91	*91*	92	*88*	93	*85*	85	*90*	95	93	95	*82*	95	*71*	95	*62*	94	*76*	94	94	95	*9*	95	*1*	93	*0*	94	*2*	96	95
30	93	*90*	94	*86*	96	*83*	88	*91*	95	93	94	*76*	95	*63*	93	*55*	94	*82*	94	95	95	*2*	94	*0*	95	*0*	95	*9*	94	92
31	91	*87*	92	*79*	92	*72*	91	*79*	93	90	94	*67*	95	*43*	95	*27*	93	*44*	94	93	94	*0*	92	*0*	94	*0*	96	*0*	95	94
32	95	*79*	92	*68*	96	*62*	84	*87*	96	95	94	*44*	96	*22*	94	*14*	93	*67*	94	96	94	*0*	94	*0*	95	*0*	94	*0*	94	93

NOTE: Results for bootstrap confidence intervals for the MMEs appear in nonitalic type. For comparison, the theoretical coverage probabilities of the confidence intervals based on the naive estimator, which assumes ignorable missingness, are shown in italic type.

but in general they are not. Our estimators have the advantage of being very simple to obtain compared to the MLEs, and the simulation study has shown that the loss in terms of MSE tends to be small (and that there is the possibility of a gain for small sample sizes). Moreover, the bootstrap confidence intervals for the MMEs are valid even when the point estimates lie on the boundary.

The model that we have applied here does not take advantage of the longitudinal nature of the data. Extensions that make use of such information will have to make specific modeling assumptions, and in general the derivation of results about the identifiability of such models will be difficult (see also the discussion in the Appendix). This remains an open problem.

When one does not know whether or not the missing-data mechanism is nonignorable, the simple MM fitting of a model such as the one that we have described here may constitute a useful sensitivity analysis to the MCAR (or, more generally, MAR) assumption. In the QL study that motivated this work, we have observed how allowing for nonignorable missingness in the data may help correct for bias and provide further insight in the study.

APPENDIX: PROOFS

Proof of Identifiability Result

Here we give the proofs of the identifiability results described in Section 3. We also provide formulas for the log-likelihood and its first and second derivatives, which are used in the modified Newton–Raphson algorithm for maximizing the likelihood. (General formulas for these expressions are described in Baker 1994.)

Call $\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_q$ the true probability mass functions of the q independent categorical outcomes $Y_1, Y_2, \ldots, Y_q$. Without loss of generality, let all Y_i, $i = 1, \ldots, q$ be defined on the same support $\{1, \ldots, k\}$. Let $\mathbf{r}$ be the missingness mechanism that describes the probability of observing each outcome Y_i; that is, $r_j = \Pr(Y_i \text{ is observed} | Y_i = j)$, $j = 1, \ldots, k$. We thus assume that $\mathbf{r}$ is the same for all q variables. Also, let all $r_j > 0$. We call a $q \times k$ model the $(q+1)$-ple $\mathcal{M} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_q, \mathbf{r}\}$. A model $\mathcal{M} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_q, \mathbf{r}\}$ is *nonidentifiable* if there exists another model $\mathcal{M}^* = \{\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_q^*, \mathbf{r}^*\} \neq \mathcal{M}$ such that $r_j^* \mathbf{f}_i^*(y) = r_j \mathbf{f}_i(y), i = 1, \ldots, q, j = 1, \ldots, k$.

Proposition A.1. Let $\mathcal{M}$ be a discrete model $\mathcal{M} = \{\mathbf{f}_1, \mathbf{f}_2, \ldots, \mathbf{f}_q, \mathbf{r}\}$ defined on $\{1, \ldots, k\}$. Call $\mathbf{f}_i = [p_{i1}, p_{i2}, \ldots, p_{ik}]'$, where $p_{ij} = \Pr(Y_i = j)$. Then $\mathcal{M}$ is identifiable if and only if the matrix $\mathbf{P} = [\mathbf{f}_1, \ldots, \mathbf{f}_q]'$ has full column rank.

We first show that if the matrix $\mathbf{P}$ has full column rank, then there can be no other model $\mathcal{M}^* = \{\mathbf{f}_1^*, \mathbf{f}_2^*, \ldots, \mathbf{f}_q^*, \mathbf{r}^*\} \neq \mathcal{M}$ such that $r_j \mathbf{f}_i = r_j^* \mathbf{f}_i^*$, $i = 1, \ldots, q, j = 1, \ldots, k$. Suppose that it does exist. Then it must be $r_j p_{ij} = r_j^* p_{ij}^* \iff p_{ij}^* = (r_j / r_j^*) p_{ij}, i = 1, \ldots, q$, $j = 1, \ldots, k$. From this, it follows that for $\mathcal{M}^*$ to be different from $\mathcal{M}$, it must be true that $\mathbf{r} \neq \mathbf{r}^*$ and that not all $\mathbf{f}_i = \mathbf{f}_i^*$. The following one-sum constraints must hold:

$$\begin{cases} p_{11}^* + p_{12}^* + \cdots + p_{1k}^* = 1 \\ \qquad\qquad \vdots \\ p_{q1}^* + p_{q2}^* + \cdots + p_{qk}^* = 1, \end{cases}$$

and we can rewrite the system as

$$\begin{bmatrix} p_{11} r_1 & \cdots & p_{1k} r_k \\ \vdots & & \vdots \\ p_{q1} r_1 & \cdots & p_{qk} r_k \end{bmatrix} \begin{bmatrix} 1/r_1^* \\ \vdots \\ 1/r_k^* \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \tag{A.1}$$

or, equivalently, as

$$\begin{bmatrix} p_{11} & \cdots & p_{1k} \\ \vdots & & \vdots \\ p_{q1} & \cdots & p_{qk} \end{bmatrix} \begin{bmatrix} r_1 & \cdots & 0 \\ \vdots & & \vdots \\ 0 & \cdots & r_k \end{bmatrix} \begin{bmatrix} 1/r_1^* \\ \vdots \\ 1/r_k^* \end{bmatrix} = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \tag{A.2}$$

or $[\mathbf{PR}][1/\mathbf{r}^*] = 1$.

The system in (A.2) has at least $\mathbf{r}^* = \mathbf{r}$ as a solution and thus is consistent. Multiplication of $\mathbf{P}$ by the full-rank matrix $\mathbf{R}$ is such that the resulting matrix $\mathbf{PR}$ still has full column rank, and thus the solution to the system is unique and $\mathbf{r}^* \neq \mathbf{r}$ is contradicted. Therefore, the necessity part of the proposition is shown to be true.

We now show that if $\mathbf{P}$ is not of full column rank, then the model is nonidentifiable. As we have seen, a model $\mathcal{M}^* \neq \mathcal{M}$ must be such that $\mathbf{r}^*$ satisfies the system (A.2). Also, the probabilities r_j^* must be strictly positive and not greater than 1. If $\mathbf{P}$ is not of full column rank, then $\mathbf{PR}$ is not as well, and the system (A.2) does not have a unique solution. In particular, if we call $\mathbf{G}$ a generalized inverse of the matrix $\mathbf{PR}$, then all possible solutions to the system can be written as $\mathbf{G1} + (\mathbf{GPR} - \mathbf{I})z$ by plugging in all possible values for z (see Searle 1982, p. 237). Because we assume that $\mathbf{r}$ is a solution to the system, there must exist a value $\tilde{z}$ such that $\mathbf{r} = \mathbf{G1} + (\mathbf{GPR} - \mathbf{I})\tilde{z}$. By continuity, we can find another value z^* such that $\mathbf{r}^* = \mathbf{G1} + (\mathbf{GPR} - \mathbf{I})z^* \neq \mathbf{r}$ is also in $(0, 1]^k$, thus showing the nonidentifiability of $\mathcal{M}$.

A way of constructing the solution $\mathbf{r}^* \neq \mathbf{r}$ described in the last part of the proof is as follows. Because P is not of full column rank, so is the matrix $\mathbf{PR}$, which means that (A.2) has at least two linearly independent (and, in particular, different) solutions. If we call one solution $\mathbf{r} \in (0, 1]^k$ and the other solution $\mathbf{r}'$ (where $\mathbf{r}'$ may or may not be in $(0, 1]^k$), then we can construct $\mathbf{r}^*$ as a convex linear combination of $\mathbf{r}$ and $\mathbf{r}'$ in such a way that all of the components $r_j^* \in (0, 1]$. The rank condition contained in Proposition A.1 is such that the two corollaries described in Section 3 are immediately seen to be true.

The connection between the nonidentifiability of a model and the estimability of the parameters of that model by ML is of interest. If a parametric model is not identifiable, then the Kullback–Leibler information $H(\boldsymbol{\theta}, \boldsymbol{\theta}_0)$ (a measure of the distance between the observed distributions corresponding to the parameter value $\boldsymbol{\theta}$ and the true value $\boldsymbol{\theta}_0$) is 0 over a subset of the parameter space Θ that contains more than one point. In the particular case considered here, we have seen that such a subset will be linear. The model corresponding to θ is identifiable if and only if the equation $H(\boldsymbol{\theta}, \boldsymbol{\theta}_0) = 0$ has a unique solution at $\boldsymbol{\theta} = \boldsymbol{\theta}_0$ (see Rao 1992, p. 123). Because $H''(\boldsymbol{\theta}_0, \boldsymbol{\theta}_0) = -\mathbf{I}(\boldsymbol{\theta}_0)$, where $\mathbf{I}(\boldsymbol{\theta}_0)$ is the Fisher information matrix, it follows that if the model is nonidentifiable, then the observed likelihood is flat over some subset of Θ, thus creating a ridge and making estimation of the parameters (which are not unique anymore) impossible. As we approach this condition, the MLEs of the parameters will necessarily become less and less stable.

Likelihood Formulas

For reference, we report here the formulas necessary for application of the Newton–Raphson algorithm to obtain the MLEs. Let p_{ij} denote the probability of observing the jth outcome category from the ith group, where $i = 1, \ldots, q$; $j = 1, \ldots, k$; q denotes

the number of independent populations; and k denotes the number of categories for the outcome variable. Let r_j denote the conditional probability of observing the outcome variable given that the outcome variable has the value j (i.e., nonignorable probability of missingness). Letting n_{ij} denote the number of times that category j is observed in group i, n_i denote the total number of observations in group i, and $n_i^* = n_i - \sum_{j=1}^{k} n_{ij}$ denote the number of missing observations from the ith group, the observed likelihood for a sample of data is given by

$$L(\mathbf{p},\mathbf{r}) = \prod_{i=1}^{q}\left[\left(\prod_{j=1}^{k}(p_{ij}r_j)^{n_{ij}}\right)\left(\sum_{j=1}^{k}p_{ij}(1-r_j)\right)^{n_i^*}\right],$$

where $\mathbf{p}$ and $\mathbf{r}$ denote vectors of the $\mathbf{p}_{ij}$ and the $\mathbf{r}_j$. We note for the purpose of computing derivatives that $p_{ik} = 1 - \sum_{j=1}^{k-1} p_{ij}$ for all i. If we call $l(\mathbf{p},\mathbf{r})$ the log-likelihood, then the score vector is given by

$$\frac{\partial l(\mathbf{p},\mathbf{r})}{\partial p_{st}} = \frac{n_{st}}{p_{st}} + \frac{N_s^*(r_k - r_t)}{\sum_{j=1}^{k} p_{sj}(1-r_j)} - \frac{n_{sk}}{p_{sm}}$$

and

$$\frac{\partial l(\mathbf{p},\mathbf{r})}{\partial r_t} = \sum_{i=1}^{q}\frac{n_{it}}{r_t} - \frac{n_i^* p_{it}}{\sum_{j=1}^{k} p_{ij}(1-r_j)},$$

and the matrix of second derivatives of $l(\mathbf{p},\mathbf{r})$ is

$$\frac{\partial^2 l(\mathbf{p},\mathbf{r})}{\partial p_{st}^2} = \frac{-n_{st}}{p_{st}^2} - \frac{n_{sk}}{p_{sk}^2} - \frac{n_s^*(r_k - r_t)^2}{\left(\sum_{j=1}^{k} p_{sj}(1-r_j)\right)^2},$$

$$\frac{\partial^2 l(\mathbf{p},\mathbf{r})}{\partial p_{st}\partial p_{su}} = \frac{-n_s^*(r_k - r_t)(r_k - r_u)}{\left(\sum_{j=1}^{k} p_{sj}(1-r_j)\right)^2},$$

$$\frac{\partial^2 l(\mathbf{p},\mathbf{r})}{\partial p_{st}\partial p_{vu}} = 0,$$

$$\frac{\partial^2 l(\mathbf{p},\mathbf{r})}{\partial r_t^2} = \sum_{i=1}^{q}\frac{-n_{it}}{r_t^2} - \frac{n_i^* p_{it}^2}{\left(\sum_{j=1}^{k} p_{ij}(1-r_j)\right)^2},$$

$$\frac{\partial^2 l(\mathbf{p},\mathbf{r})}{\partial r_t \partial r_u} = \sum_{i=1}^{q}\frac{-n_i^* p_{it}p_{iu}}{\left(\sum_{j=1}^{k} p_{ij}(1-r_j)\right)^2},$$

$$\frac{\partial^2 l(\mathbf{p},\mathbf{r})}{\partial p_{st}\partial r_t} = \frac{n_s^*(r_k - r_t)p_{st} - n_s^*\sum_{j=1}^{k} p_{sj}(1-r_j)}{\left(\sum_{j=1}^{k} p_{sj}(1-r_j)\right)^2},$$

$$\frac{\partial^2 l(\mathbf{p},\mathbf{r})}{\partial p_{st}\partial r_u} = \frac{n_s^*(r_k - r_t)p_{su}}{\left(\sum_{j=1}^{k} p_{ij}(1-r_j)\right)^2},$$

$$\frac{\partial^2 l(\mathbf{p},\mathbf{r})}{\partial p_{st}\partial r_k} = \frac{n_s^*(r_k - r_t)p_{sk} + n_s^*\sum_{j=1}^{k} p_{sj}(1-r_j)}{\left(\sum_{j=1}^{k} p_{sj}(1-r_j)\right)^2}.$$

[Received January 1998. Revised March 1999.]

REFERENCES

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

Baker, S. G. (1994), "Composite Linear Models for Incomplete Multinomial Data (disc: P647-649)," *Statistics in Medicine*, 13, 609–622.

——— (1995), "Marginal Regression for Repeated Binary Data With Outcome Subject to Non-Ignorable Non-Response," *Biometrics*, 51, 1042–1052.

Baker, S. G., and Laird, N. M. (1988), "Regression Analysis for Categorical Variables With Outcome Subject to Nonignorable Nonresponse," *Journal of the American Statistical Association*, 83, 62–69; Corr: 83, 1232.

Bernhard, J., Cella, D. F., Coates, A. S., Fallowfield, L., Ganz, P. A., Moinpour, C. M., Mosconi, P., Osoba, D., Simes, J., and Hürny, C. (1998), "Missing Quality of Life Data in Cancer Clinical Trials: Serious Problems and Challenges," *Statistics in Medicine (special issue)*, 17, 517–532.

Berzon, R. (1998), "Understanding and Using Health-Related Quality of Life Instruments Within Clinical Research Studies," *Quality of Life Assessment in Clinical Trials: Methods and Practice*, eds. M. J. Staquet, R. D. Hays, and P. M. Fayers, New York: Oxford University Press.

Elashoff, J. D., and Elashoff, R. M. (1974), "Two-Sample Problems for a Dichotomous Variable With Missing Data," *Applied Statistics*, 23, 26–34.

Fairclough, D. L. (1997), "Summary Measures and Statistics for Comparison of Quality of Life in a Clinical Trial of Cancer Therapy," *Statistics in Medicine*, 16, 1197–1209.

Fairclough, D., Peterson, H., Cella, D., and Bonomi, P. (1998), "Comparison of Several Model-Based Methods for Analysing Incomplete Quality of Life Data in Cancer Clinical Trials," *Statistics in Medicine (special issue)*, 17, 781–796.

Fay, R. E. (1986), "Causal Models for Patterns of Nonresponse," *Journal of the American Statistical Association*, 81, 354–365.

Fitzmaurice, G. M., Laird, N. M., and Zahner, G. E. P. (1996), "Multivariate Logistic Models for Incomplete Binary Responses," *Journal of the American Statistical Association*, 91, 99–108.

Gelber, R., Bonetti, M., Cole, B., Gelber, S., and Goldhirsch, A. for the International Breast Cancer Study Group (1998), "Quality of Life Assessment in the Adjuvant Setting: Is It Relevant?," in *Recent Results in Cancer. Adjuvant Therapy of Primary Breast Cancer*, eds. H. Senn, A. Goldhirsch, R. Gelber, and B. Thurlimann, Berlin: Springer-Verlag.

Gelber, R., Cole, B., Gelber, S., and Goldhirsch, A. (1995), "Comparing Treatments Using Quality-Adjusted Survival: The Q-TWiST Method," *The American Statistician*, 49, 161–169.

Glonek, G. (1999), "On Identifiability in Models for Incomplete Binary Data," *Statistics and Probability Letters*, 41, 191–197.

Hahn, J. (1996), "A Note on Bootstrapping Generalized Method-of-Moments Estimators," *Econometric Theory*, 12, 197–197.

Hürny, C., Bernhard, J., Coates, A., Castiglione-Gertsch, M., Peterson, H. F., Gelber, R. D., Forbes, J. F., Rudenstam, C., Simoncini, E., Crivellasi, D., Goldhirsch, A., Senn, H., for the International Breast Cancer Study Group (1996), "Impact of Adjuvant Therapy on Quality of Life in Women With Node-Positive Operable Breast Cancer," *Lancet*, 347, 1279–1284.

International Breast Cancer Study Group (1997), "Effectiveness of Adjuvant Chemotherapy in Combination With Tamoxifen for Node-Positive Postmenopausal Breast Cancer Patients," *Journal of Clinical Oncology*, 15, 1385–1394.

Little, R. J. A., and Rubin, D. B. (1987), *Statistical Analysis With Missing Data*, New York: Wiley.

Nordheim, E. V. (1984), "Inference From Nonrandomly Missing Categorical Data: An Example From a Genetic Study on Turner's Syndrome," *Journal of the American Statistical Association*, 79, 772–780.

Park, T., and Brown, M. B. (1994), "Models for Categorical Data With Nonignorable Nonresponse," *Journal of the American Statistical Association*, 89, 44–52.

Rao, B. L. S. P. (1992), *Identifiability in Stochastic Models: Characterization of Probability Distributions*, New York: Academic Press.

Searle, S. R. (1982), *Matrix Algebra Useful for Statistics*, New York: Wiley.

Simes, R. J., Greatorex, V., and Gebski, V. J. (1998), "Practical Approaches to Minimize Problems With Missing Quality of Life Data," *Statistics in Medicine*, 17, 724–737.

The Mathworks, Inc. (1998), MATLAB.

Troxel, A. B., and Harrington, D. P. (1998), "Analysis of Longitudinal Data With Nonignorable Nonmonotone Missing Values," *Applied Statistics*, 47, 425–438.