Appl. Statist. (1994)
43, No. 2, pp. 371–384

Subject-specific and Population-averaged Continuation Ratio Logit Models for Multiple Discrete Time Survival Profiles

By THOMAS R. TEN HAVE†

Pennsylvania State University, Hershey, USA

and DAVID H. UTTAL

University of Illinois, Urbana, USA

[Received November 1992. Final revision June 1993]

SUMMARY

Subject-specific and population-averaged continuation ratio logit models are presented for multivariate discrete time survival data. The models characterize data from a psychological experiment by using a quadratic polynomial relationship across time that depends on a time-independent condition. A multivariate normal random effects distribution is imposed on intercept, linear and quadratic terms in the subject-specific model, which is fitted by using a combination of Gibbs sampling and buffered stochastic substitution. Variance components that tend towards 0 are addressed in this context. In addition, generalized estimating equations estimates of the parameters in the population-averaged model are compared with analogous estimates for the mixed effects model.

Keywords: Bayesian models; Generalized estimating equations; Gibbs sampling; Hyperpriors; Mixed effects; Variance components

1. Introduction

Subject-specific and population-averaged continuation ratio logit (CRL) models are presented for correlated discrete time survival data. Given a set of ordinal multinomial response probabilities summing to 1, the continuation ratio is defined to be the ratio of a multinomial probability over the partial sum of the remaining multinomial probabilities (see, for example, Agresti (1990), pages 319–321, for details).

The data of interest (listed in Appendix E) consist of multiple discrete time survival profiles for each subject of a psychological study that attempted to determine whether children could compensate for map rotation if they were given repeated opportunities to discover that the map was rotated. The sample comprised 89 children, of ages 35–67 months, each of whom attempted to find a toy hidden under one of 20 buckets scattered throughout a room. The toy was hidden at 10 different locations, the order of which was the same for each child. For each of

†Address for correspondence: Center for Biostatistics and Epidemiology, Pennsylvania State University, Hershey, PA 17033, USA.

© 1994 Royal Statistical Society 0035–9254194143371

Copyright © 2001. All Rights Reserved.

these 10 trials, each subject was allowed three attempts to find the toy after seeing a map indicating the location of the toy. Each subject was randomly assigned to one of two groups: one in which the map was rotated when presented by an investigator and one in which the map was presented correctly. The primary question of interest is: did the rotated map group recover and eventually find the toy as successfully as the non-rotated map group? Hence the focus of the analysis is on modelling and comparing the trends in correctness across the 10 locations in the two groups, where location represents a discrete time variable.

The CRL model, which was introduced by Cox (1972), has been employed by several researchers in the context of discrete time data. Fienberg and Mason (1978) estimated simultaneously age, period and cohort effects with the CRL model. Thompson (1977) showed that for grouped time survival data the CRL model converges to the Cox proportional hazards model when the number of intervals increases and the interval lengths go to 0. Efron (1988) used semiparametric smoothing to fit the CRL model to head and neck cancer data. Gillespie *et al.* (1993) fitted similar CRL models to evaluate the risk of lung cancer in a sample of ex-smokers. In addition, Ryan (1992) discussed fitting CRL models with an overdispersion parameter to correlated developmental toxicity data that involved a discrete time variable (see also D'Agostino et *al.* (1990) and Agresti (1990), pages 319–321 and 337).

The subject-specific and population-averaged CRL models considered here are extensions of the logistic regression models discussed by Zeger *et al.* (1988). The subject-specific CRL model is a mixed effects model where the expectation of a response is conditional on a subject-specific or cluster-specific random effect. In the population-averaged CRL model, the expectation of a response is obtained by integrating out the subject-specific random effects and hence is interpreted as an average for the population of interest as opposed to an individual subject.

The mixed effects CRL model presented here is distinguishable from the subject-specific model of Zeger *et al.* (1988) in several respects:

(a) the CRL model represents an extension of the logistic random effects models;
(b) the model here includes a multivariate random effects distribution for the intercept and linear and quadratic effects;
(c) a proper hyperprior distribution is imposed on the random effects variance–covariance components;
(d) several previously reported approaches are integrated to yield a modified Gibbs sampling routine.

These points will be elucidated later.

For fitting the population-averaged logistic regression model to clustered data, Liang and Zeger (1986) proposed generalized estimating equations (GEEs), which are score-like functions that are derived by incorporating a working correlation matrix (i.e. the correlation is treated as a nuisance parameter) into the likelihood equations for logistic regression. The solution of this function yields GEE estimators, which are consistent regardless of which working correlation structure is assumed. Results based on the GEE method are presented for both the independence and the exchangeable working correlation matrices. Henceforth, when the independence or exchangeable working correlation matrix is assumed, the GEE

Copyright © 2001. All Rights Reserved.

estimates will be referred to as the independence or exchangeability GEE estimates respectively.

The subject-specific CRL model is developed in Sections 2 and 3. The corresponding population-averaged model is presented in Section 4. The results of fitting the mixed effects and population-averaged models are presented in Section 5. Section 6 concludes the paper with a discussion.

2. Subject-specific Continuation Ratio Logit Model

Let $Y_{ijkl}=1$ when the kth child, who is in the lth group (for $l=1$ (rotated), 2 (non-rotated)), finds the toy on the ith attempt (for $i=1, 2, 3$) at the jth hidden location (for $j=1, \ldots, 10$), and let $Y_{ijkl}=0$ otherwise for all i, j, k, l. Also let $Y_{4jkl}=1$ when the child fails to find the toy after three attempts, and let $Y_{4jkl}=0$ otherwise for all j, k, l. For given j, k, l, $\Sigma_{i=1}^{4} Y_{ijkl}=1$.

Conditional on subject k in the lth group at the jth location, assume that the random vector $\mathbf{Y}_{jkl}=(Y_{1jkl} \ldots Y_{4jkl})$ has a multinomial distribution with mean parameters $\{\pi_{ijkl}=E(Y_{ijkl}|\tau_k), i=1, \ldots, 4: \Sigma_{i=1}^{4}\pi_{ijkl}=1)$, where τ_k is the vector of subject k's random effects parameters in the subject-specific CRL model defined below.

Next define λ_{ijkl} as follows:

$$\lambda_{ijkl} = \pi_{ijkl} \Big/ \sum_{i'=i}^{4} \pi_{i'jkl}$$

$$= \Pr(Y_{ijkl}=1 \mid Y_{i'jkl}=0, \ i'<i; \ \tau_k) \tag{1}$$

for $i=1, 2, 3$ and all j, k and l, and where $Y_{0jkl}\equiv 0$. Given the random effect of the kth subject in the lth group, and given that he or she has failed to find the toy at the jth hidden location on the previous $i-1$ attempts, λ_{ijkl} is the conditional probability that the child finds the toy on the ith attempt, i.e. λ_{ijkl} is a discrete hazard rate conditional on the kth subject effect. (See, for example, Agresti (1990), p. 337, for a hazard rate interpretation of a ratio of probabilities similar to λ_{ijkl}.)

Conditional on the effect of the kth subject in the lth group, the log-likelihood for the jth hidden location can be expressed in terms of a sum of Bernoulli log-likelihoods involving the λ_{ijkl} parameters as follows:

$$\begin{aligned} L_{jkl} &= y_{1jkl}\log\pi_{1jkl} + y_{2jkl}\log\{(1-\pi_{1jkl})\lambda_{2jkl}\} + y_{3jkl}\log\{(1-\pi_{1jkl})(1-\lambda_{2jkl})\lambda_{3jkl}\} \\ &\quad + y_{4jkl}\log\{(1-\pi_{1jkl})(1-\lambda_{2jkl})(1-\lambda_{3jkl})\} \\ &= \sum_{i=1}^{3}\left\{ y_{ijkl}\log\psi_{ijkl} + \log(1-\lambda_{ijkl})\sum_{i'=i}^{4} y_{i'jkl}\right\} \end{aligned} \tag{2}$$

for $i=1, 2, 3$ and all j, k and l, and where $\psi_{ijkl}=\lambda_{ijkl}(1-\lambda_{ijkl})^{-1}$. The parameter ψ_{ijkl} is a continuation ratio and can be interpreted as the conditional odds that subject k in group l at location j finds the toy on the ith attempt given that the subject has not found the toy on the previous i attempts.

Now consider a mixed effects logistic regression model for ψ_{ijl}, $i=1, 2, 3$, which accounts for the correlation among repeated observations within individuals. The variable indicating the location at which the toy is hidden is ordinal, in that the children improve by learning as they proceed across the 10 hidden locations but then

Copyright © 2001. All Rights Reserved.

become fatigued. Consequently, the location effect can be characterized by a polynomial growth curve. A priori information and a preliminary analysis of the data suggest that learning and fatigue in this context correspond to a quadratic relationship between the conditional log-odds of a correct choice and location. Orthogonal contrasts are employed here to help to maintain numerical stability and to improve the precision of the location estimates. A preliminary analysis of the data also reveals that the effects of location appear to be independent of attempt, although the effects of location and attempt do appear to interact separately with rotation of the map. The resulting mixed effects model is

$$\log \psi_{ijkl} = \beta^0 + \beta^{A}_{il} + \beta^{L1}X_{1j} + \beta^{L2}X_{1j} + \beta^{R}_{l} + \beta^{L1R}_{l}X_{1j} + \beta^{L2R}_{l}X_{2j} + \tau^0_k + \tau^{L1}_k X_{1j} + \tau^{L2}_k X_{2j}, \tag{3}$$

where $\beta^{A}_{1l} = \beta^{R}_{1} = \beta^{L1R}_{1} = \beta^{L2R}_{1} = 0$ and X_{1j} and X_{2j} are orthogonal linear and quadratic coefficients for the lth location. The τ-parameters in equation (3) are random effects parameters and will be collectively referred to as the 3 x 1 vector τ_k for subject k. The remaining parameters, denoted by the 10 x 1 vector β, are fixed effects parameters and are interpreted as follows:

(a) β^0, the intercept, represents the log-odds that a subject finds the toy on the first attempt when all covariates are 0;
(b) β^{A}_{il} is the change in the conditional log-odds of a correct choice when a subject in the lth group finds the toy on the ith attempt instead of on the first attempt at a give hidden location;
(c) β^{L1} and β^{L2} represent linear and quadratic location effects respectively for a subject in the rotated map group;
(d) β^{R}_{2} is the change in the intercept if a subject in the rotated map group were to receive a non-rotated map instead;
(e) β^{L1R}_{2} and β^{L2R}_{2} are interactions between the rotation effect and the linear and quadratic location effects, i.e. these parameters represent the changes in the linear and quadratic effects, when a subject changes rotation status.

The elements of β are assumed to have a multivariate non-informative prior distribution, which by definition does not provide information towards estimating the fixed effects parameters since the prior variances are assumed to go to infinity. Equivalently,

$$\beta \sim N_{10}(\eta, \Sigma), \qquad \Sigma^{-1} \to 0, \tag{4}$$

where η is a 10 x 1 vector of unspecified population mean parameters and Σ is a 10 x 10 matrix, the elements of the inverse of which go to $\mathbf{0}$.

The random effects parameters τ^0_k, τ^{L1}_k and τ^{L2}_k in equation (3) represent subject-specific intercept, linear and quadratic random effects terms respectively, which are independent of the β-parameters. The distribution of these parameters is multivariate normal:

$$\tau_k \sim N_3(\mathbf{0}, \Omega), \tag{5}$$

where $\mathbf{0}$ is a 3 x 1 vector of 0s and Ω is a 3 x 3 variance–covariance matrix.

Following Zeger and Karim (1991), a non-informative hyperprior (i.e. a distribution for the parameters of a prior distribution) can be assumed for Ω:

Copyright © 2001 All Rights Reserved

$$\Pr(\Omega) \propto |\Omega|^{-(p+1)/2}, \tag{6}$$

where p (= 3) is the number of random effects variates. Alternatively, we can specify an informative proper hyperprior for Ω^{-1}, such as a Wishart distribution,

$$\Omega^{-1} \sim W_3\{(\rho R)^{-1}, \rho\}, \tag{7}$$

where W_3 denotes the trivariate Wishart distribution, and R and ρ are an *a* priori specified 3 x 3 non-singular symmetric matrix and a positive scalar respectively.

3. Estimation of Subject-specific Model

Parameter estimation for the mixed effects CRL model can be viewed in terms of estimating univariate posterior densities (see, for example, Stiratelli et *al.* (1984) and Zeger and Karim (1991)) of the β- and Ω-parameters in models (3) and (5). The approach used to estimate these posterior densities given the priors defined above and the data involves a modification of the Gibbs sampling procedure performed by Zeger and Karim (1991). The Gibbs sampler is an iterative algorithm that simulates approximations to univariate or joint posterior densities given the data and priors. A buffered stochastic substitution procedure was incorporated into the Gibbs sampling algorithm used for this paper to reduce the correlation between iterations (see Appendix A). At each iteration of the Gibbs sampler, we sampled each of the elements of β, τ_k and Ω from their respective proper conditional posterior densities given the data and previously simulated values of all other parameters in the model (see Appendix B).

Convergence of the Gibbs sampling algorithm was not achieved with the non-informative hyperprior density (6) for Ω, because the simulated variance components of Ω became trapped at 0. However, convergence appeared to be achieved (although after *1500* iterations) with the Wishart hyperprior (7) given $\rho = 1$ and $R = 0.001 I_{3\times 3}$.

After convergence, the Gibbs sampler was run for 2000 more iterations, yielding estimated univariate posterior distributions and corresponding moments and percentiles for the fixed effects and variance components of the random effects. In addition, the Gibbs sampling output was used to compute posterior estimates $\hat{E}(\lambda_{ijkl})$ of the population average of the subject-specific discrete hazard rate,

$$E(\lambda_{ijkl}) = \int \lambda_{ijkl}\, \mathrm{d}F(\tau_k),$$

where $F(\ )$ denotes the multivariate normal distribution of τ_k (see Appendix C).

4. Population-averaged Continuation Ratio Logit Model

Now consider averaging the subject-specific hazard rate λ_{ijkl} over the 'survivors' in group l at attempt i of location j (i.e. the subjects in group l who have not found the toy by the ith attempt at the jth location), i.e. given the subject-specific model defined in equations (1) and (3), let

$$\lambda^*_{ijl} = E[E(Y_{ijkl} \mid Y_{i'jkl} = 0,\ i' < i;\ \tau_k)]. \tag{8}$$

We fitted λ^*_{ijl} with a model that is analogous to the mixed effects model in equation (3) without the random effects:

Copyright © 2001. All Rights Reserved

$$\log \psi^*_{ijl} = \beta^{0*} + \beta^{A*}_{il} + \beta^{L1*} X_{1j} + \beta^{L2*} X_{2j} + \beta^{R*}_{l} + \beta^{L1R*}_{l} X_{1j} + \beta^{L2R*}_{l} X_{2j} \qquad (9)$$

where $\psi^*_{ijl} = \lambda^*_{ijl}(1-\lambda^*_{ijl})^{-1}$ for $i = 1$, 2, 3 and all j and l. The parameters in model (9) will be collectively referred to as the 10 x 1 vector β^* and are interpreted similarly to the corresponding subject-specific @-parameters in model (3). However, instead of being interpreted with respect to subject-specific conditional probabilities, the @*-parameters in model (9) are interpreted with respect to the expected proportions of subjects who find the toy. For example, testing $\beta^{A*}_{il} = 0$ is the test of the hypothesis that, for a given location, the expected proportion of all subjects in the lth group who find the toy on the first attempt does not differ from the expected proportion that find the toy on the ith attempt among those subjects in the lth group who have not found the toy on the previous $i-1$ attempts.

Given the Bernoulli log-likelihood (2) with λ_{ijkl} and ψ_{ijkl} replaced by λ^*_{ijl} and ψ^*_{ijl} respectively, we define the corresponding GEEs, the solutions of which yield GEE estimates of β^*:

$$\sum_{l=1}^{2} \sum_{k=1}^{N_l} \mathbf{D}^{\mathrm{T}}_{kl} \mathbf{V}_{kl}^{-1} (\mathbf{y}_{kl} - \boldsymbol{\lambda}^*_{kl}) = 0, \qquad (10)$$

where $\mathbf{y}_{kl}$ is the vector of observed conditional responses (i.e. $\{y_{ijkl};\ y_{i'jkl} = 0,\ i' < i\ \forall j\}$) for the kth of N_l individuals in group l, $\boldsymbol{\lambda}^*_{kl}$ is the vector of corresponding population-averaged discrete hazard rates, $\mathbf{D}_{kl}$ is the matrix of derivatives $(\partial \boldsymbol{\lambda}^*_{kl} / \partial \beta^*)$ and $\mathbf{V}_{kl}$ is the working variance–covariance matrix for the elements of $\mathbf{y}_{kl}$.

GEE estimates of the parameters in model (9) that are based on independence and exchangeable working correlation matrices are reported below with naive and sandwich variance estimates. The sandwich estimator is a consistent estimator of the variance–covariance structure of the GEE estimates regardless of the working correlation structure used to obtain the GEE estimates. Let $\hat{\boldsymbol{\lambda}}^*_{kl}$ be the vector of GEE-based estimates of the population-averaged discrete hazard rates (see Appendix D) for subject k in group l and $\hat{\mathbf{V}}_{kl}$ and $\hat{\mathbf{D}}_{kl}$ be estimates of $\mathbf{V}_{kl}$ and $\mathbf{D}_{kl}$ respectively. The sandwich estimator is defined to be

$$\mathbf{H}^{-1} \left(\sum_{l=1}^{2} \sum_{k=1}^{N_l} \hat{\mathbf{S}}_{kl} \hat{\mathbf{S}}^{\mathrm{T}}_{kl} \right) \mathbf{H}^{-1}, \qquad (11)$$

where

$$\begin{aligned} \hat{\mathbf{S}}_{kl} &= \hat{\mathbf{D}}^{\mathrm{T}}_{kl} \hat{\mathbf{V}}_{kl}^{-1} (\mathbf{y}_{kl} - \hat{\boldsymbol{\lambda}}^*_{kl}), \\ \mathbf{H} &= \sum_{l=1}^{2} \sum_{k=1}^{N_l} \hat{\mathbf{D}}^{\mathrm{T}}_{kl} \hat{\mathbf{V}}_{kl}^{-1} \hat{\mathbf{D}}_{kl} \end{aligned} \qquad (12)$$

(see, for example, Liang and Zeger (1986) for more details).

The naïve estimator of the variance–covariance matrix of the GEE estimators is $\mathbf{H}^{-1}$, which is the negative inverse of the observed information when the GEEs are likelihood equations. Hence, when independence holds and the independence working correlation is used, comparing the GEE estimates with their respective naïve standard errors in equivalent to performing Wald tests based on maximum likelihood estimates and their respective observed information-based standard

Copyright © 2001. All Rights Reserved

TABLE 1
Comparison of estimates of parameters in the subject-specific model (3) with independence and exchangeability estimates of the parameters in the population-averaged model (9)

Explanatory variable	*Mixed effects*		*Marginal independence*			*Marginal exchangeable*		
	Mean	*Standard error*	*Estimate*	*Naïve standard error*	*Robust standard error*	*Estimate*	*Naïve standard error*	*Robust standard error*
Intercept	−1.35	0.13	−1.20	0.12	0.20	−1.23	0.19	0.20
Attempt 2-rotated	0.11	0.19	−0.076	0.19	0.19	0.20	0.16	0.15
Attempt 3-rotated	−0.11	0.22	−0.40	0.22	0.22	0.05	0.18	0.16
Attempt 2-non-rotated	−0.28	0.18	−0.50	0.17	0.17	−0.19	0.16	0.16
Attempt 3-non-rotated	−0.89	0.24	−1.22	0.24	0.24	−0.67	0.20	0.19
Rotate	1.50	0.18	1.34	0.16	0.25	1.37	0.24	0.25
Linear	0.086	0.015	0.077	0.016	0.018	0.081	0.014	0.017
Quadratic	−0.095	0.025	−0.085	0.024	0.023	−0.088	0.021	0.022
Linear-rotated	−0.065	0.020	−0.062	0.020	0.020	−0.065	0.018	0.020
Quadratic-rotated	0.074	0.035	0.065	0.032	0.030	0.064	0.027	0.030

errors. Table 1 reveals that the naive and robust standard errors for the independence GEE estimates are very similar for the data of interest. Although all results reported here for the independence GEE estimates pertain to the comparison of the independence GEE point estimates with their respective sandwich standard errors, these results also hold approximately for the maximum likelihood analysis under independence.

5. Results

5.1. *Comparison of Population-averaged and Subject-specific Models*

Table 1 indicates that the mixed effects estimates of the non-attempt parameters (i.e. parameters other than β_{il}^{A}, $i=2, 3, l=1, 2$) exceed in magnitude (up to 15%) the corresponding GEE estimates of the population-averaged model under both

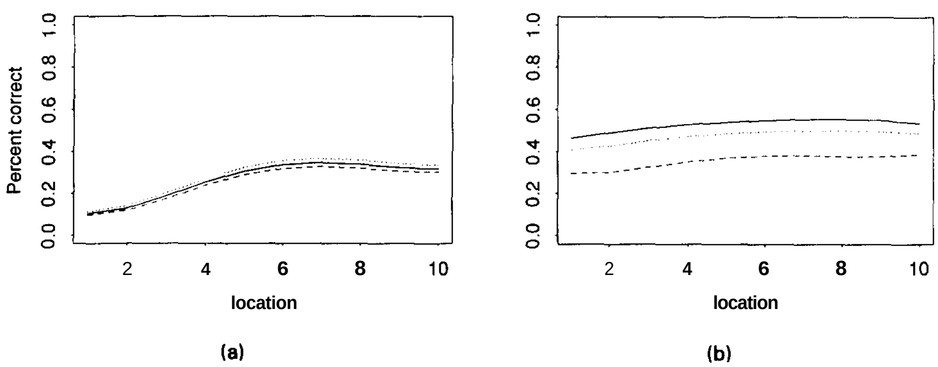

Fig. 1. Plot of mixed effects estimates of the population averages of the subject-specific hazard probabilities across location for each attempt for (a) the rotated map population and (b) the non-rotated map group: ———, attempt 1; ·······, attempt 2; -------, attempt 3

Copyright © 2001. All Rights Reserved

TABLE 2
Observed percentage correct† displayed for all combinations of attempt and rotation summed over location

Rotation status	*Percentages for the following number of attempt:* 1	2	3
Yes	24.4 (105/430)	21.8 (71/325)	16.5 (42/254)
No	53.5 (246/460)	40.7 (87/214)	25.2 (32/127)

†The numbers of correct choices over the total number of observations are given in parentheses.

working correlation matrices. Note that the exchangeability and independence GEE estimates of the non-attempt parameters differ by no more than 5%.

Inferences based on the estimates of the group–attempt interactions are affected by the choice of the model (Fig. 1). For example, the non-rotated map subjects performed significantly more poorly on attempt 2 than on attempt 1 if the independence GEE estimate is used ($Z=0.50/0.17=-2.94$), yet the exchangeability GEE ($Z=-0.19/0.16=-1.19$) and mixed effects ($Z=-0.28/0.18=-1.56$) estimates of this effect are not significant (Table 1). A similar situation exists for the attempt 3–attempt 1 contrast for the rotated map group, where the independence GEE approach shows borderline significance ($Z=-0.40/0.22=-1.82$) in contrast with the clearly non-significant estimates of the other approaches.

The frequencies (summed across location) and corresponding percentages for the cross-classification of the rotate and attempt effects are displayed in Table 2 to help to explain these differences in group-specific attempt effect estimates among the models. There is a precipitous drop in percentage correct between attempts 1 and 3 (53.5% compared with 25.2%) for the non-rotated map group, which is shown to be very significant by all the models. A less substantial drop from 53.5% to 40.7% between attempts 1 and 2 for the non-rotated map group corresponds to the conflicting inferences described above for this contrast. Fig. 2(b) suggests that the very significant independence estimate of this effect ($Z=-2.94$) is questionable since the independence GEE procedure apparently weighted the clinically unexplained spike at location 5 for the attempt 1 profile more heavily than did the exchangeability GEE and mixed effects estimation approaches. A similar examination of Table 2 and Fig. 2(a) reveals that for the rotated group the independence GEE procedure weighted the spike at location 7 for attempt 1 in Fig. 2(a) more heavily than did the other procedures.

The similarity of the parameter inferences between the mixed effects and exchangeability GEE approaches extends to the comparison of the mixed effects estimates of the population-averaged discrete hazard rate (i.e. $\hat{E}(\lambda_{ijkl})$; see Appendix C) with the corresponding fitted values from the exchangeability GEE estimates (i.e. $\hat{\lambda}^*_{ijl}$; see Appendix D). The plot of these two sets of estimates against each other for all attempt–rotation–location combinations in Fig. 3 indicates good agreement between these two sets of estimates. Fig. 3 also includes a similar plot

Copyright © 2001. All Rights Reserved

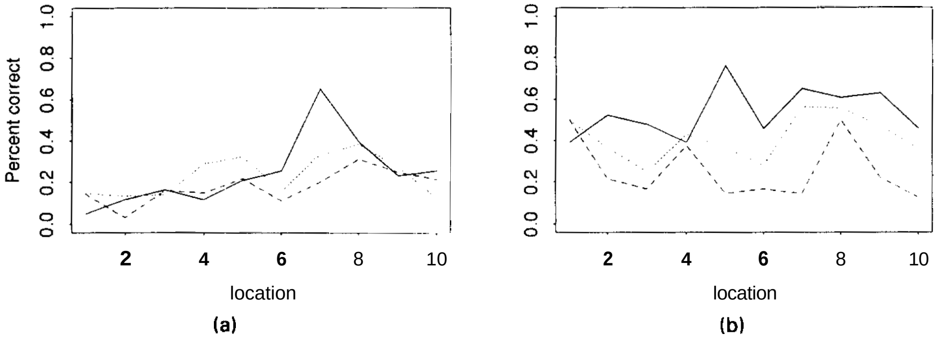

Fig. 2. Plot of observed percentage correct across location for each attempt for (a) the rotated map group and (b) the non-rotated map group: ——, attempt 1; ········, attempt 2; -------, attempt 3

of the independence GEE estimates of λ^*_{ijl} and the mixed effect estimates of λ^*_{ijl}, revealing poorer agreement between these two sets of estimates.

5.2. *Summary of Substantive Results*

Although the following results are based on the subject-specific estimates in Table 1, they are supported by all models.

(a) For a given subject in the rotated map group, the conditional probability of a correct choice improved significantly across locations (linear $Z = 0.086/0.015 = 5.73$), but this increase levelled off in the later locations (quadratic $Z = -0.095/0.025 = -3.8$) so that the conditional probability of

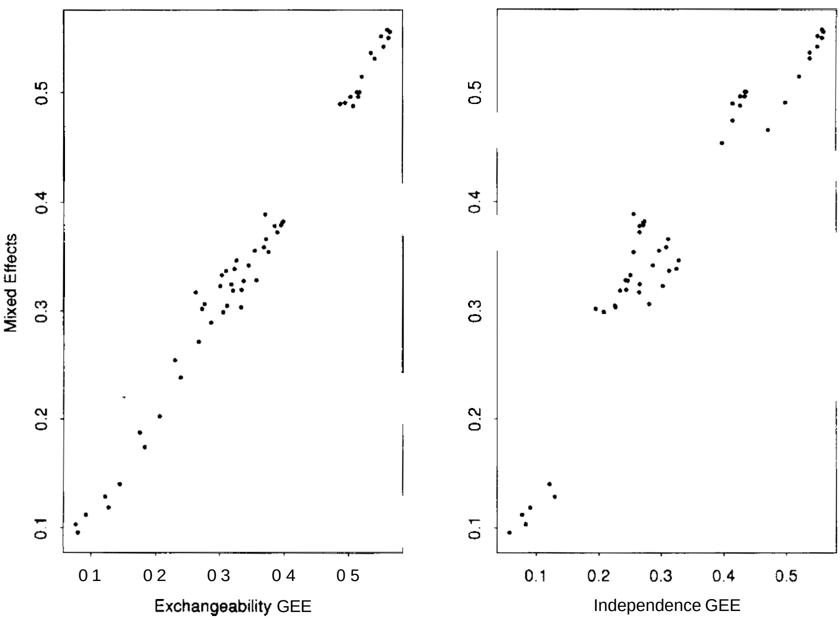

Fig. 3. Mixed effects estimates of the population-averaged discrete hazard rate (i.e. $\hat{E}(\lambda_{ijkl})$) *versus* the corresponding fitted values from the exchangeability and independence GEE estimates, separately (i.e. $\hat{\lambda}^*_{ijl}$)

Copyright © 2001. All Rights Reserved

TABLE 3
Posterior estimates of the variance–covariance components of Ω for the mixed effects model (3)†

Component of Ω	Posterior estimates				
	Mean	**Median**	**Standard error**	**2.5%**	**97.5%**
Ω_{int}	0.31	0.30	0.14	0.059	0.61
$\Omega_{int\text{-}lin}$	0.0085	0.0083	0.0034	0.0025	0.016
$\Omega_{int\text{-}quad}$	-0.0068	-0.0058	0.0012	-0.033	0.016
Ω_{lin}	0.00036	0.00035	0.00012	0.00018	0.00063
$\Omega_{lin\text{-}quad}$	0.00069	0.00065	0.00039	0.000015	0.0015
Ω_{quad}	0.011	0.011	0.0038	0.0052	0.020

†The column headings 2.5% and 97.5% designate the 2.5 and 97.5 percentiles respectively.

success did not reach the conditional success rate on attempt 1 if the subject had received a non-rotated map.

(b) For a given subject in the non-rotated map group, the conditional probability of a correct choice did not rise as much across location as if the subject had received a rotated map (linear–rotated $Z = -0.065/0.020 = -3.25$). Moreover, the conditional probability of success on the third attempt was significantly less than that of the first attempt for a given subject in the non-rotated map group at a given location ($Z = -0.89/0.24 = -3.71$), which contrasts with the non-significant attempt effects for a given subject in the rotated map group.

Finally, the posterior estimates and densities of the variance componznts presented in Table 3 and Fig. 4 reveal that the estimated density of the variance component for the linear term is massed near 0, in contrast with the densities of the intercept and quadratic variance components. Hence, it appears that the vertical placements and shapes of the subject-specific curves vary more than the slopes, which show relatively little variation.

6. Discussion

The choice between subject-specific and population-averaged models depends on whether the focus is on hypothesis testing or estimation. In terms of testing, the subject-specific and exchangeability **GEE** approaches yielded similar inferences for all parameters and are preferred over the independence **GEE** method, which was sensitive to clinically unexplained spikes in the observed profiles as seen in Figs 2(a) and 2(b). The similarities in inference between the subject-specific and exchangeability **GEE** methods agree with the results reported by Zeger et al. (1988) for the logistic regression model.

With regard to estimation, the subject-specific model is preferred if we are interested in the effects of covariates on subject-specific discrete hazard rates. However, if we are interested in estimating the population average of the subject-specific hazard rate then either model is theoretically and empirically appropriate, although the population-averaged model is easier to fit.

Copyright © 2001. All Rights Reserved

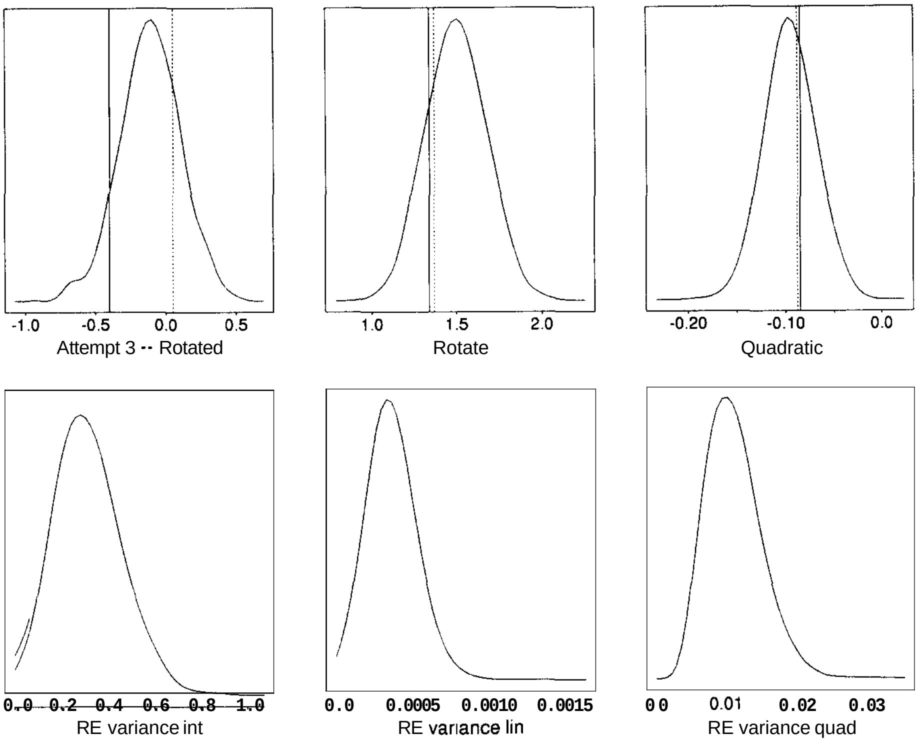

Fig. 4. Marginal posterior densities of representative fixed effects and variance component parameters of the subject-specific model (3): the full and dotted vertical lines indicate the positions of the independence and exchangeability GEE estimates respectively, corresponding to the displayed fixed effects distributions

The observed differences between the independence and exchangeability GEE estimates of the attempt effects are attributable to the relatively small numbers of subjects (42 or fewer) on which the second and third attempt effects estimates are based. Our preference for the exchangeability GEE estimate conforms with previous work (see, for example, Lipsitz et al. (1991) and Liang et al. (1992)). However, McDonald (1993) contends that the independence GEE estimate is more 'stable' than GEE estimates that account for working correlations when fitting binary logistic regression models for small samples. Determining which view holds for CRL models is a subject of future research.

Acknowledgements

The contribution of the first author was funded in part by National Institutes of Health grant CA53787 when he was Visiting Assistant Professor, Department of Biostatistics, University of Michigan. The data were collected under a grant from the Spencer Foundation. The authors would like to thank Professor Mark P. Becker, two referees and the Associate Editor for invaluable comments that substantially improved the paper.

Copyright © 2001. All Rights Reserved.

Appendix A

The extension of Clayton's buffered stochastic substitution algorithm to the CRL model with multiple prior parameters can be explained as follows. Perform the following three steps, which comprise a buffered stochastic substitution algorithm, at the ith iteration.

(a) Draw Ω and β simulated at an iteration randomly selected from the previous B_1 iterations.
(b) Given the randomly drawn simulations of Ω and β, perform the rejection sampling algorithm to simulate the components of τ_k. Repeat this step B_2 times.
(c) Given the components of τ_k sampled at the last (B_2th) repetition of step (b), simulate new Ω and β, which then enter the pool of simulations of Ω and β from which we randomly sample for the next iteration. Initially B_1 is small, but it is then increased as the equilibrium distribution is attained. For the present paper, B_1 was increased to 100 and $B_2=5$ (see Clayton (1991) for more details).

Appendix B

Consider simulating the components of Ω^{-1}. Given the Wishart hyperprior specified in distribution (7), the conditional posterior distribution for Ω^{-1} is a Wishart distribution with parameters that are functions of τ_k, ρ and R:

$$\Omega^{-1} \sim W_3\left\{\left(\sum_{k=1}^{K} \tau_k \tau_k' + \rho R\right)^{-1}, K+\rho\right\} \tag{13}$$

Setting $\rho=0$ in distribution (13) yields the conditional density which results from the non-informative prior specified in model (6). Generating random variates from distribution (13) is straightforward using algorithms discussed by Odell and Feivson (1966).

As an example of simulating the random and fixed effects parameters at the ith iteration, consider simulating the first random effects element for subject k, τ_k^0, from its conditional posterior density. Let h( ) equal the product of the exponentiated log-likelihood function (2) and the prior defined in distribution (5), and let $g(\ )$ be a univariate split t-density (Geweke, 1989) with mean and variance such that the split t-density has higher density than (i.e. envelopes) the likelihood prior across its support. The rejection sampling algorithm used to simulate a value for τ_k^0, say τ_k^{0*}, can then be described as follows:

(a) sample τ_k^{0*} from g( );
(b) generate a uniform(0, 1) random variate u and
(c) if $u < h(\tau_k^{0*})/c\,g(\tau_k^{0*})$ then accept τ_k^{0*}; otherwise reject τ_k^{0*} and return to step (a).

At each iteration, we may determine c by matching the mode of the corresponding trivariate split t-density to the mode of the likelihood prior, which can be estimated with a Newton–Raphson algorithm (see Zeger and Karim (1991) for details). The variance-covariance matrix of the multivariate split t-density is matched at each iteration to the inverse Hessian of the log-likelihood prior such that the number of rejections for a given c is minimized approximately (see Carlin and Gelfand (1991) for details).

Appendix C

Following Karim and Zeger (1992) and Zeger et al. (1988), the posterior estimate $\hat{E}(\lambda_{ijkl})$ of the population average of λ_{ijkl} is computed as follows. At each post-convergence iteration of the Gibbs sampler, the following approximation is computed for $E(\lambda_{ijkl})$ based on the simulated fixed effects parameters in the subject-specific probability model for the Ith group on the ith attempt at the jth hidden location:

$$E(\lambda_{ijkl}) = \int \lambda_{ijkl}\, \mathrm{d}F(\tau_k)$$

Copyright © 2001. All Rights Reserved.

$$\doteq \psi_{ijl}^{a_j(\Omega)}/(1+\psi_{ijl}^{a_j(\Omega)}) \tag{14}$$

where $F(\ )$ is the multivariate normal distribution of τ_k, $\log \psi_{ijl}=\beta^{\mathrm{T}}Z_{ijl}$, Z_{ijl} is the full set of covariates for the ijlth combination, $a_j(\Omega)=|\{(16\sqrt{3})/15\pi\}^2\Omega X_j X_j^{\mathrm{T}}+I_{3\times 3}|^{-3/2}$, $X_j^{\mathrm{T}}=(1\,X_{1j}\,X_{2j})$ and $I_{3\times 3}$ is the 3×3 identity matrix. The average of $\mathrm{E}(\lambda_{ijkl})$ is taken across the 2000 post-convergence iterations for each group–location–attempt combination, thus yielding the plots of the posterior estimates $\hat{E}(\lambda_{ijkl})$ of the population-averaged discrete hazard rates in Figs **2(a)** and **2(b)**.

Appendix D

The GEE-based estimate of λ_{ijl}^* is computed as $\hat{\lambda}_{ijl}^*=\hat{\psi}_{ijl}^*(1+\hat{\psi}_{ijl}^*)^{-1}$, where $\hat{\psi}_{ijl}^*=\exp(\hat{\beta}^{*\mathrm{T}}Z_{ijl})$ and $\hat{\beta}^*$ is the exchangeability or independence GEE estimate of the parameters in the population-averaged model (9).

Appendix E: Listing of Map Data

Each line consists of a subject's rotation status **(0 ≡ yes, 1 ≡ no)** and the number of attempts that a subject needed to find the toy at a each location. (The number 4 indicates that the subject failed to find the toy at a given location.)

Rotate	*L1*	*L2*	*L3*	*L4*	*L5*	*L6*	*L7*	*L8*	*L9*	*L10*	*Rotate*	*L1*	*L2*	*L3*	*L4*	*L5*	*L6*	*L7*	*L8*	*L9*	*L10*
0	4	4	4	4	4	4	4	2	4	4	0	4	4	2	2	3	4	1	4	4	4
0	2	4	4	4	2	1	2	4	3	4	0	4	4	4	4	4	4	4	2	4	4
0	4	4	4	2	4	4	3	2	4	4	0	4	4	4	4	3	4	1	4	2	2
0	3	4	4	4	4	4	4	1	4	4	0	4	4	1	4	2	4	1	4	3	3
0	4	4	4	4	4	4	4	4	4	4	0	4	4	3	4	3	4	1	1	3	2
0	4	4	4	3	1	1	2	2	2	1	0	4	4	4	4	4	1	1	1	2	3
0	4	4	4	3	3	3	1	2	2	4	0	4	4	1	1	1	2	1	1	4	4
0	4	4	4	2	4	1	1	1	2	3	0	1	1	3	2	1	1	1	1	1	1
0	4	4	4	4	2	4	1	1	4	4	0	2	4	4	1	4	4	4	4	4	4
0	1	1	4	4	1	4	1	4	3	4	0	4	4	4	4	2	1	1	3	1	4
0	4	4	4	2	4	4	1	3	3	4	1	3	4	2	2	2	4	1	3	1	1
0	4	4	4	3	4	4	1	2	2	1	1	3	2	2	2	1	2	1	2	4	3
0	4	4	4	2	4	4	1	1	4	4	1	2	1	4	4	2	4	2	3	1	1
0	4	4	4	4	4	4	1	1	4	4	1	4	4	4	2	1	1	2	2	2	4
0	4	3	1	2	1	1	1	1	1	2	1	3	1	4	4	1	1	4	4	1	1
0	4	4	4	3	4	2	4	1	4	3	1	3	4	4	4	4	3	2	2	4	4
0	2	2	2	2	4	1	2	3	1	3	1	4	1	2	2	1	4	1	1	2	1
0	2	1	2	1	2	4	1	1	2	1	1	2	1	4	3	3	1	1	2	1	1
0	2	2	3	1	1	1	1	3	1	1	1	1	1	4	1	4	4	3	1	2	1
0	3	4	4	4	4	4	4	4	4	4	1	2	4	1	1	4	1	1	2	1	2
0	4	4	4	4	2	2	1	2	1	3	1	1	2	1	1	1	4	1	1	2	2
0	4	2	4	4	4	4	1	3	2	2	1	2	2	2	3	1	2	1	1	4	2
0	4	4	4	4	2	2	1	2	4	1	1	4	1	1	3	1	2	1	1	1	2
0	4	4	4	4	4	2	3	2	3	1	1	2	1	4	2	1	4	2	1	1	4
0	3	4	2	4	2	4	1	1	4	1	1	2	3	3	4	1	1	2	3	1	1
0	3	2	1	2	1	1	1	1	1	1	1	1	4	4	2	4	4	1	1	1	4
0	3	1	1	4	2	4	1	1	2	4	1	1	4	4	3	1	1	1	4	1	4
0	2	1	1	1	1	1	1	1	1	1	1	3	1	3	2	1	1	2	2	1	4
0	4	2	1	4	1	4	1	4	1	4	1	3	1	1	2	1	4	1	1	1	1
0	4	4	3	4	4	3	4	4	4	4	1	1	2	1	4	1	4	1	4	4	1
0	4	4	4	4	2	4	2	1	4	4	1	2	4	4	4	1	3	4	2	4	4
0	4	4	2	2	2	3	2	2	1	1	1	4	1	1	2	1	1	4	1	3	4
0	4	4	3	2	3	4	1	4	4	4	1	1	1	1	2	1	1	1	1	1	1

(continued)

Copyright © 2001. All Rights Reserved

Rotate	*L1*	*L2*	*L3*	*L4*	*L5*	*L6*	*L7*	*L8*	*L9*	*LIO*	*Rotate*	*L1*	*L2*	*L3*	*L4*	*L5*	*L6*	*L7*	*L8*	*L9*	*L10*
1	1	2	1	1	1	1	1	1	1	3	1	1	1	4	2	1	1	1	1	1	1
1	4	4	4	4	4	4	4	4	4	4	1	1	1	1	1	2	1	2	1	4	2
1	2	1	1	1	1	1	1	1	1	1	1	2	2	1	3	1	4	4	1	2	4
1	2	1	3	1	1	1	1	1	1	2	1	1	4	2	1	1	4	1	1	3	4
1	2	4	1	1	1	1	1	1	1	1	1	1	1	4	1	2	1	2	1	2	1
1	2	1	1	1	1	3	1	1	1	1	1	2	3	4	4	4	4	1	1	2	4
1	4	1	1	1	1	2	1	1	1	1	1	1	4	1	4	1	1	1	2	1	1
1	3	3	1	2	1	1	1	2	1	4	1	1	1	1	1	1	4	1	3	1	2
1	1	1	1	1	1	2	1	1	1	1	1	1	1	1	3	1	4	4	1	1	2
1	1	1	1	1	1	2	2	1	1	1	1	4	1	2	1	1	1	1	1	1	1
1	1	1	1	1	1	1	1	1	1	1	1	1	2	1	1	1	1	1	2	1	2
1	2	2	4	4	1	2	1	1	2	4											

References

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

Carlin, B. P. and Gelfand, A. E. (1991) An iterative Monte Carlo method for nonconjugate Bayesian analysis. ***Statist. Comput.*, 1,** 129–134.

Clayton, D. G. (1991) A Monte Carlo method for Bayesian inference in frailty models. ***Biometrics*, 47,** 467–486.

Cox, D. R. (1972) Regression models and life-tables (with discussion). **J.** R. ***Statist. Soc.*** B, **34,** 187–220.

D'Agostino, R. B., Lee, M.-L., Belanger, A. J., Cupples, L. A., Anderson K. and Kannel, W. B. (1990) Relation of pooled logistic regression to time dependent Cox regression analysis: the Framingham Heart Study. ***Statist. Med.*,** 9, 1501–1515.

Efron, B. (1988) Logistic regression, survival analysis, and the Kaplan–Meier Curve. ***J. Am. Statist. Ass.*, 83,** 414–425.

Fienberg, S. E. and Mason, W. M. (1978) Identification and estimation of age-period-cohort effects in the analysis of discrete archival data. ***Sociol. Methodol.*,** 1–67.

Geweke, J. (1989) Bayesian inference in econometric models using Monte Carlo integration. ***Econometrica*, 57,** 1317–1339.

Gillespie, B., Halpern, M. T. and Warner, K. E. (1993) Patterns of lung cancer risk in ex-smokers. In ***Case Studies in Biometry*** (eds N. Lange and L. Ryan). New York: Wiley.

Karim, M. R. and Zeger, S. L. (1992) Generalized linear models with random effects: salamander mating revisited. ***Biometrics*, 48, 631–644.**

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

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

Lipsitz, S. R., Laird, N. M. and Harrington, D. P. (1991) Generalized estimating equations for correlated binary data: using the odds ratio as a measure of association. ***Biometrika*, 78,** 153–160.

McDonald, B. W. (1993) Estimating logistic regression parameters for bivariate binary data. **J.** R. ***Statist. Soc. B, 55,*** 391–397.

Odell, P. L. and Feiveson, A. H. (1966) A numerical procedure to generate a sample covariance structure. ***J. Am. Statist. Ass., 61,*** 198–203.

Ryan, L. (1992) Quantitative risk assessment for developmental toxicity. ***Biometrics*, 48,** 163–174.

Stiratelli, R., Laird, N. and Ware, J. H. (1984) Random-effects model for serial observations with binary response. ***Biometrics*, 40,** 961–971.

Thompson, Jr, W. A. (1977) On the treatment of grouped observations in life studies. ***Biornetrics*, 33,** 436–470.

Zeger, S. L. and Karim, M. R. (1991) Generalized linear models with random effects. **J. *Am. Statist. Ass.*, 86,** 75–90.

Zeger, S. L., Liang, K.-Y. and Albert, **P.** S. (1988) Models for longitudinal data: a generalized estimating equations approach. ***Biornetrics*, 44,** 1049–1060.

Copyright © 2001 All Rights Reserved