THE ANALYSIS OF METHADONE CLINIC DATA USING MARGINAL AND CONDITIONAL LOGISTIC MODELS WITH MIXTURE OR RANDOM EFFECTS

JENNIFER S.K. CHAN[1], ANTHONY Y.C. KUK[2], JAMES BELL[3] AND CHARLES MCGILCHRIST [4]

University of Hong Kong,   University of New South Wales, Prince of Wales Hospital   and   Australian National University

Summary

The paper develops methods for the statistical analysis of outcomes of methadone maintenance treatment (MMT). Subjects for this study were a cohort of patients entering MMT in Sydney in 1986. Urine drug tests on these subjects were performed weekly during MMT, and were reported as either positive or negative for morphine, the marker of recent heroin use. To allow correlation between the repeated binary measurements, a marginal logistic model was fitted using the generalized estimating equation (GEE) approach and the alternating logistic regression approach. Conditional logistic models are also considered. Results of separate fitting to each patient and score tests suggest that there is substantial between-patient variation in response to MMT. To account for the population heterogeneity and to facilitate subject-specific inference, the conditional logistic model is extended by introducing random intercepts. The two, three and four group mixture models are also investigated. The model of best fit is a three group mixture model, in which about a quarter of the subjects have a poor response to MMT, with continued heroin use independent of daily dose of methadone; about a quarter of the subjects have a very good response, with little or no heroin use, again independent of dose; and about half the subjects responded in a dose-dependent fashion, with reduced heroin use while receiving higher doses of methadone. These findings are consistent with clinical experience. There is also an association between reduced drug use and increased duration in treatment. The mixture model is recommended since it is quite tractable in terms of estimation and model selection as well as being supported by clinical experience.

Key words: Correlated binary data; serial dependence; GEE approach; EM algorithm; random effects model; mixture model with latent groups.

1. Introduction

In recent years, there has been a resurgence in the support of a methadone maintenance treatment (MMT) in many countries as studies have revealed its contribution in reducing the

risk of HIV infection among injecting heroin users (Stimson, 1990; Novick, Joseph & Croxson, 1990). However, there is wide variation in treatment approaches, and limited consensus over the factors which make treatment effective. The existing consensus is that higher doses of methadone are associated with less heroin use, and that longer treatment is associated with better outcomes (Ward, Mattick & Hall, 1992; Hargreaves, 1983). However, even on these issues the research literature is somewhat inconsistent, and clinical practice is often out of line with research evidence (D'Aunno & Vaughn, 1992).

We have tried to develop statistical procedures which would help clarify the factors associated with better outcomes from MMT. Retaining people in treatment, and reduced heroin use by those remaining in treatment, are the major objectives of methadone treatment. However, research into factors associated with less heroin use is difficult. Treatment needs to be flexible and interactive to be optimally effective, and many patients seek higher or lower doses at different times of treatment, often fluctuating markedly during their time in treatment. The freedom to request such changes is an important part of treatment, as it gives patients a sense of control over their treatment. Because of this, it is not straightforward to randomize people to receive a certain daily dose so as to compare the outcomes of different dose levels.

The current study aimed to develop methods for investigating the relationship between (fluctuating) daily dose of methadone, time in treatment, and heroin use as detected by urine drug testing. It is based on a data set derived from the treatment records of a group of heroin users who entered MMT at a single clinic in western Sydney in 1986. Outcome measure is heroin use as measured by urine testing performed once a week, on a day determined at random. Screens were recorded as positive or negative for morphine, the biological marker for heroin use. The clinic required attendance for dosing seven days per week, with take-home doses of medication provided only in exceptional circumstances.

The analysis was performed using a restricted data set in which subjects who completed less than 4 weeks of treatment were excluded. The dosing records for some patients were missing and such patients were also excluded from analysis. Finally, experience showed that beyond the first 6 months of treatment non-random drop-out began to occur, with patients who continued to use heroin being more likely to leave treatment. In part, this reflects a clinic policy of withdrawing patients who continued regular heroin use after a period of time. Such selective attrition from treatment can lead to a false impression of a progressive reduction in heroin use with increasing duration of treatment, because what is really happening is that regular heroin users are being removed from treatment. Therefore, this study looked only at results of urine screens collected in the first 26 weeks of treatment.

There were 136 heroin users, submitting a total of 2872 urine screens with 16% of them being positive for heroin. The dosage averaged over the 2872 incidents is 64 mg. For all analyses, each urine screen result rather than each patient served as the unit of analysis. Methadone dose at the time the urine test was collected was included as a predictor variable, as was duration of treatment in weeks. Initially, we also included the interaction effect between dose and time, but we found it to be not significant and therefore dropped it subsequently.

Since our data consists of repeated measurements, it is important to take serial correlation into account in our analysis. To achieve this, we consider both marginal and conditional logistic models. In Section 2.1, we fit a marginal logistic model using the first order GEE approach (Liang & Zeger, 1986) with a working correlation matrix of the autoregressive or exchangeable type. As the second order GEE approach is computationally infeasible, we use instead the alternating logistic regression approach proposed by Carey, Zeger & Diggle (1993), assuming equal pairwise odds ratio. In Section 2.3, we fit a conditional logistic model

proposed by Bonney (1987) where the logit of $\Pr\{Y_t = 1 \mid Y_{t-1}\}$ depends linearly on Y_{t-1} and other covariates.

Results of separate fitting to each patient and score tests suggest that there is substantial between-patient variation. To account for population heterogeneity and to facilitate subject-specific inference, we extend the conditional logistic model by introducing random intercepts. This is done in Section 3, using the estimation procedure proposed by McGilchrist (1994). Our main conclusion is that reduced drug use is associated with increased methadone dose and increased duration in treatment. The interaction between dose and time effects is not significant.

Finally, in Section 4, we consider mixture models which postulate that there are different groups of patients who react differently to methadone treatment. These models are more tractable than the random intercept model as the likelihood function can be computed easily without the need of integration. Estimation can be carried out using the EM algorithm and model selection can be based on the Akaike Information Criterion (AIC).

The model selected is a 3 group mixture model. Under this model, methadone treatment is associated with cessation of heroin use for about 30% of the subjects regardless of the methadone doses used. Another 46% of the subjects appear to respond to treatment in a dose-dependent manner with reduced heroin use at high doses of methadone. The remaining 24% of the patients failed to respond to treatment in this study. These findings are consistent with clinical experience.

2. Marginal and Conditional Logistic Models

Let Y_{it} $(i = 1, \ldots, m;\ t = 1, \ldots, n_i)$ denote the observed outcome of the ith patient at time t and $n = n_1 + \cdots + n_m$ the total number of observations. We consider models where the marginal probabilities $P_{it} = \Pr\{Y_{it} = 1\}$ or the conditional probabilities $\Pr\{Y_{it} = 1 \mid Y_{i1}, \ldots, Y_{i,t-1}\}$ are logit-linear in some covariates.

2.1. Marginal Logistic Model

The main focus of this model is on the marginal distribution of the Y_{it}. It is assumed that

$$\mathrm{logit}(P_{it}) = \eta_{it} = \beta_0 + \beta_1 d_{it} + \beta_2 \ln t, \tag{1}$$

where d_{it} is the dosage administered to patient i at time t. The association across time of the repeated outcomes for a subject is treated as nuisance and is entered only in a working correlation matrix that appears in the estimating equation. An advantage of this approach is that the resulting estimates are robust to misspecification of the correlation structure provided that the mean model is correctly specified. A working correlation matrix for patient i is

$$R_i(\alpha) = \begin{pmatrix} 1 & \alpha & \alpha^2 & \cdots & \alpha^{n_i-1} \\ \alpha & 1 & \alpha & \cdots & \alpha^{n_i-2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \alpha^{n_i-1} & \alpha^{n_i-2} & \alpha^{n_i-3} & \cdots & 1 \end{pmatrix}, \tag{2}$$

corresponding to an autoregressive process of order 1. Following the suggestion of a referee, we also consider the equicorrelated case,

$$R_i(\alpha) = (1 - \alpha)I + \alpha J,$$

where I is the identity matrix and J the matrix of all 1s. By Liang & Zeger (1986), the first order generalized estimating equation (GEE1) is

$$U(\boldsymbol{\beta}) = \sum_{i=1}^{m} \frac{\partial \mathbf{P}_i^T}{\partial \boldsymbol{\beta}} [\text{cov}^*(\mathbf{Y})_i]^{-1}(\mathbf{Y}_i - \mathbf{P}_i) = \mathbf{0}, \tag{3}$$

where

$$\frac{\partial \mathbf{P}_i^T}{\partial \boldsymbol{\beta}} = \mathbf{X}_i^T \, \text{diag}(S_{i1}^2, \dots, S_{in_i}^2), \tag{4}$$

$\mathbf{X}_i$ is the $n_i \times 3$ design matrix for patient i with row vectors $(1, d_{it}, \ln t)$, $t = 1, \dots, n_i$, $\mathbf{P}_i = (P_{i1}, \dots, P_{in_i})^T$ and $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in_i})^T$. We set the working covariance of $\mathbf{Y}_i$ to be

$$\text{cov}^*(\mathbf{Y}_i) = \mathbf{V}_i = \text{diag}(S_{i1}, \dots, S_{in_i}) \, R_i(\alpha) \, \text{diag}(S_{i1}, \dots, S_{in_i}),$$

where $S_{it}^2 = P_{it}(1 - P_{it}) = e^{\eta_{it}}/(1 + e^{\eta_{it}})^2$. For the case of AR(1) working correlation, we estimate α by the lag-1 autocorrelation

$$\widehat{\alpha} = \frac{1}{n-m} \sum_{i=1}^{m} \sum_{t=1}^{n_i} \frac{(Y_{it} - P_{it})(Y_{i,t+1} - P_{i,t+1})}{[P_{it}(1 - P_{it}) \, P_{i,t+1}(1 - P_{i,t+1})]^{\frac{1}{2}}}. \tag{5}$$

For the equicorrelated case, we estimate α by

$$\widehat{\alpha} = \frac{1}{\sum_{i=1}^{m} n_i(n_i - 1)/2} \sum_{i=1}^{m} \sum_{t=2}^{n_i} \sum_{t'=1}^{t-1} \frac{(Y_{it} - P_{it})(Y_{it'} - P_{it'})}{[P_{it}(1 - P_{it}) \, P_{it'}(1 - P_{it'})]^{\frac{1}{2}}}. \tag{6}$$

With estimated $\alpha^{(k)}$, we can update $\boldsymbol{\beta}$ to $\boldsymbol{\beta}^{(k+1)}$ by solving (3) using the Newton–Raphson method. Then α is subsequently updated using $\boldsymbol{\beta}^{(k+1)}$ in (5) or (6) and the cycle repeats again until convergence is reached. Finally, cov($\boldsymbol{\beta}$), the variance-covariance matrix of $\boldsymbol{\beta}$ (Liang & Zeger, 1986), equals

$$\left[\sum_{i=1}^{m} \frac{\partial \mathbf{P}_i^T}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} \frac{\partial \mathbf{P}_i}{\partial \boldsymbol{\beta}^T} \right]^{-1} \left[\sum_{i=1}^{m} \left(\frac{\partial \mathbf{P}_i^T}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} \text{cov}(\mathbf{Y}_i) \mathbf{V}_i^{-1} \frac{\partial \mathbf{P}_i}{\partial \boldsymbol{\beta}^T} \right) \right] \left[\sum_{i=1}^{m} \frac{\partial \mathbf{P}_i^T}{\partial \boldsymbol{\beta}} \mathbf{V}_i^{-1} \frac{\partial \mathbf{P}_i}{\partial \boldsymbol{\beta}^T} \right]^{-1}, \tag{7}$$

where cov($\mathbf{Y}_i$) is estimated by $(\mathbf{Y}_i - \widehat{\mathbf{P}}_i)(\mathbf{Y}_i - \widehat{\mathbf{P}}_i)^T$.

The GEE1 approach is not likelihood-based and does not require the complete specification of the joint distribution of the binary responses. Fitzmaurice & Laird (1993) and Fitzmaurice, Laird & Rotnitzky (1993) discuss a likelihood-based method for the case of equal n_i using a log-linear representation of the joint probabilities of the binary responses. They parametrize in terms of the marginal means and the canonical association parameters of the log-linear model and show that the resulting likelihood equations for the regression parameters are of exactly the same form as GEE1. However, the form of their model is not reserved to subsets of observations and the interpretation of their model parameters is dependent on the cluster size. Thus their method is not appropriate for the case of unequal n_i which is a feature of the methadone clinic data.

2.2. Alternating Logistic Regression

The GEE1 approach is primarily concerned with estimating the regression parameters of the marginal model. Zhao & Prentice (1990) and Liang, Zeger & Qaqish (1992) proposed second order generalized estimating equations (GEE2) which simultaneously model the regression and association parameters. This approach is appropriate if the association structure is also of scientific interest and provided that the association structure is correctly specified, GEE2 will give more efficient estimates than GEE1. A drawback of GEE2 is that, unlike GEE1, it is not robust to misspecification of the association structure. Moreover, GEE2 is extremely computationally intensive as it involves the inversion of matrices of dimension $O(n_i^2) \times O(n_i^2)$. Using the marginal odds ratios to model the association, Carey *et al.* (1993) suggested an alternating logistic regression (ALR) approach which alternates between the use of GEE1 to estimate the regression coefficients, and a logistic regression of each response on others from the same individual with an appropriate offset to update the odds ratio parameters. This approach is computationally less demanding than GEE2 as it only requires the inversion of matrices of dimension $n_i \times n_i$. Using a few examples, Carey *et al.* (1993) demonstrated that the ALR approach is reasonably efficient relative to GEE2. We apply the ALR approach to analyse the methadone clinic data under the assumption of equal odds ratio α between each pair of binary responses from the same patient.

2.3. Conditional Logistic Model

In this model, the serial correlation within a subject is accounted for by including the 'previous outcome' as a covariate. To be precise, it is assumed that $\Pr\{Y_{it} = 1 \mid Y_{i1}, \ldots, Y_{i,t-1}\} = \Pr\{Y_{it} = 1 \mid Y_{i,t-1}\}$ and

$$\text{logit}\left[\Pr\left\{Y_{it} = 1 \mid Y_{i,t-1}\right\}\right] = \eta_{it} = \beta_0 + \beta_1 d_{it} + \beta_2 \ln t + \beta_3 Y_{i,t-1}. \tag{8}$$

The likelihood function is

$$\prod_{i=1}^{m} \Pr\{Y_{i1}, \ldots, Y_{in_i}\} = \prod_{i=1}^{m}\prod_{t=1}^{n_i} \Pr\{Y_{it} \mid Y_{i,t-1}\} = \prod_{i=1}^{m}\prod_{t=1}^{n_i} \left(\frac{e^{Y_{it}\eta_{it}}}{1 + e^{\eta_{it}}}\right), \tag{9}$$

and the log-likelihood function is

$$\ell(\boldsymbol{\beta}; \mathbf{Y}) = \sum_{i=1}^{m}\sum_{t=1}^{n_i} Y_{it}\eta_{it} - \sum_{i=1}^{m}\sum_{t=1}^{n_i} \ln(1 + e^{\eta_{it}}). \tag{10}$$

Bonney (1987) showed that we can obtain the maximum likelihood estimate for $\boldsymbol{\beta}$ using a standard logistic regression procedure with a suitably augmented design matrix. The variance-covariance matrix of $\boldsymbol{\beta}$ can be obtained by inverting $-\ell''(\boldsymbol{\beta})$.

2.4. Numerical Results

The results of fitting the marginal and conditional logistic model to the methadone clinic data are given in Table 1. For the marginal model, we estimate the regression coefficients using the GEE1 with both the AR1 and exchangeable working correlation matrix as well as the ALR assuming equal pairwise odds ratio. All three sets of estimates lead to similar conclusions. We see that both the dose and time effect are significant. Our conclusion is that reduced drug use is associated with increased methadone dose and increased duration in treatment. For the

TABLE 1

*Parameter estimates and standard errors (in italic) for marginal,
conditional and random intercept models.*

model	intercept	dose	time	AR
marginal model:				
AR1 correlation	−0.2197	−0.0108	−0.3434	n.a.
($\alpha = 0.4310$)	*0.3768*	*0.00569*	*0.0736*	n.a.
equal correlation	−0.1892	−0.0118	−0.2911	n.a.
($\alpha = 0.2360$)	*0.3795*	*0.00517*	*0.0771*	n.a.
equal odds ratio	−0.1717	−0.0109	−0.3580	n.a.
($\alpha = 1.0830$)	*0.4264*	*0.00623*	*0.0820*	n.a.
conditional model	−0.8423	−0.00884	−0.4049	2.3960
	0.2189	*0.00282*	*0.0628*	*0.1196*
random intercept model	−0.5859	−0.0132	−0.4100	1.4038
	0.3410	*0.00490*	*0.0682*	*0.1334*

TABLE 2

*Standardized score statistics and P-value (in italic) for testing
homogeneity of each coefficient.*

intercept		dose		time		AR	
6.2391	*0.0000*	4.9716	*0.0000*	5.4184	*0.0000*	2.3915	*0.0084*

conditional model, there is a strong positive association between the present outcome and the previous outcome as suggested by the highly significant autoregressive coefficient. We also considered the interaction between dose and time by including the product term $d_{it} \ln t$ as a covariate, but the interaction effect is not significant (p-value $= 0.59$ under marginal model; p- value $= 0.46$ under conditional model), so we have not included this interaction in our subsequent analyses.

The results of fitting a separate conditional logistic model to each patient reveal substantial variation between patients. For formal testing, we extend the score test of homogeneity (Commenges *et al.*, 1994) to the case of the conditional logistic model. More precisely, we use the score test to test model (8) as the null hypothesis against the alternative that one of the βs in (8) is random. We conducted the score test for each covariate separately; the results are given in Table 2. It can be seen that all tests are significant with the test for a random intercept being most significant.

3. Random Intercept Model

As there is strong evidence that the regression coefficients are patient-specific, we incorporate a random intercept in the conditional logistic model. The extended model is

$$\text{logit}[\text{Pr}\{Y_{it} = 1 \mid Y_{i,t-1}\}] = \eta_{it} = \beta_0 + u_i + \beta_1 d_{it} + \beta_2 \ln t + \beta_3 Y_{i,t-1}, \qquad (11)$$

where u_i is independent and identically distributed as $\mathcal{N}(0, \sigma^2)$. This model can be written in matrix form as

$$\boldsymbol{\eta} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}\mathbf{u}, \qquad (12)$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$, $\mathbf{X}$ the corresponding $n \times 4$ design matrix, $\mathbf{u} = (u_1, \ldots, u_m)^T$ the patient effects and $\mathbf{Z}$ the corresponding $n \times m$ design matrix for $\mathbf{u}$. The log probability

density function for $\mathbf{u}$ is

$$\ell_u(\mathbf{u}; \sigma^2) = -\tfrac{1}{2} \sum_{i=1}^{m} [\ln(2\pi\sigma^2) + \sigma^{-2} u_i^2]. \tag{13}$$

McGilchrist (1994) suggested maximizing the sum of (10) and (13) with respect to $\boldsymbol{\beta}$ and $\mathbf{u}$ to obtain $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$, the so-called best linear unbiased predictors or penalized likelihood estimators if ℓ_u is regarded as a penalty function. The sum of (10) and (13) can also be interpreted as the complete data log-likelihood function based on $\mathbf{Y}$ and $\mathbf{u}$. A Bayesian interpretation of $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$ is that they are the posterior mode under a diffuse prior for $\boldsymbol{\beta}$ and a normal prior for $\mathbf{u}$. The Newton–Raphson step for finding $\widehat{\boldsymbol{\beta}}$ and $\widehat{\mathbf{u}}$ is

$$\begin{pmatrix} \boldsymbol{\beta}^{(k+1)} \\ \mathbf{u}^{(k+1)} \end{pmatrix} = \begin{pmatrix} \boldsymbol{\beta}^{(k)} \\ \mathbf{u}^{(k)} \end{pmatrix} + \begin{pmatrix} \mathbf{X}^T \mathbf{S}^2 \mathbf{X} & \mathbf{X}^T \mathbf{S}^2 \mathbf{Z} \\ \mathbf{Z}^T \mathbf{S}^2 \mathbf{X} & \mathbf{Z}^T \mathbf{S}^2 \mathbf{Z} + \sigma^{-2} \mathbf{I} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{X}^T (\mathbf{Y} - \mathbf{P}) \\ \mathbf{Z}^T (\mathbf{Y} - \mathbf{P}) - \sigma^{-2} \mathbf{u}^{(k)} \end{pmatrix}, \tag{14}$$

where $\mathbf{S}^2 = \mathrm{diag}(S_{11}^2, \ldots, S_{1n_1}^2, S_{21}^2, \ldots, S_{mn_m}^2)$ and $S_{it}^2 = e^{\eta_{it}} / (1 + e^{\eta_{it}})^2$. By writing

$$\begin{pmatrix} \mathbf{X}^T \mathbf{S}^2 \mathbf{X} & \mathbf{X}^T \mathbf{S}^2 \mathbf{Z} \\ \mathbf{Z}^T \mathbf{S}^2 \mathbf{X} & \mathbf{Z}^T \mathbf{S}^2 \mathbf{Z} + \sigma^{-2} \mathbf{I} \end{pmatrix}^{-1} = \begin{pmatrix} \cdot & \cdot \\ \cdot & \mathbf{T} \end{pmatrix} \tag{15}$$

and $\mathbf{T}^* = (\mathbf{Z}^T \mathbf{S}^2 \mathbf{Z} + \sigma^{-2} \mathbf{I})^{-1}$, McGilchrist (1994) proposed updating σ^2 by

$$\hat{\sigma}_{ML}^2 = \frac{\mathbf{u}^T \mathbf{u}}{m - \mathrm{tr}(\mathbf{T}^*)/\sigma^2} \tag{16}$$

and

$$\hat{\sigma}_{REML}^2 = \frac{\mathbf{u}^T \mathbf{u}}{m - \mathrm{tr}(\mathbf{T})/\sigma^2}, \tag{17}$$

where all the quantities on the right-hand sides of (16) and (17) are evaluated at the current estimates of $\boldsymbol{\beta}$, $\mathbf{u}$ and σ^2. Using this new estimate of σ^2, we can update our estimate of $\boldsymbol{\beta}$ and $\mathbf{u}$ by (14). McGilchrist (1994) suggested iterating (14) and (16) or (17) until convergence to obtain approximate maximum likelihood (ML) or residual maximum likelihood (REML) estimates.

The results of fitting the random intercept model can also be found in Table 1. The ML and REML estimates of the random intercept model are very similar so we report only the former. As expected, the standard errors are inflated relative to the model without random intercepts but the conclusions are qualitatively the same, namely, reduced drug use is associated with increased methadone dose and increased duration in treatment. The estimate of the variance component, σ^2, is 1.244 (S.E. = 0.214).

4. Mixture Model

4.1. Mixture Model and the EM Algorithm

Under this model, we assume that there are G groups of patients who react differently to methadone treatment. Suppose each patient has a probability π_k of coming from group $k, k = 1, \ldots, G$. If patient i belongs to group k, then

$$\mathrm{logit}[\Pr\{Y_{it} = 1 \mid Y_{i,t-1}\}] = \eta_{itk} = \beta_{0k} + \beta_{1k} d_{it} + \beta_{2k} \ln t + \beta_{3k} Y_{i,t-1}. \tag{18}$$

TABLE 3

Parameter estimates, standard errors (in italic) and AIC for mixture models with different number of groups.

model	intercept	dose	time	AR	π	L	AIC
$G = 2$ (group 1)	−1.1732	−0.0153	−0.4317	1.5610	0.6689	−977.37	1968.75
	0.5165	*0.00692*	*0.0686*	*0.1371*	*0.0571*		
(group 2)	−0.1321	0.00097	same	same	0.3311		
	0.3100	*0.00421*			*0.0571*		
$G = 3$ (group 1)	−5.2306	0.0249	−0.4262	1.4313	0.3039	−967.45	1954.89
	2.0002	*0.0242*	*0.0694*	*0.1397*	*0.0785*		
(group 2)	0.1314	−0.0245	same	same	0.4603		
	0.3922	*0.00603*			*0.0846*		
(group 3)	0.1305	0.00047	same	same	0.2358		
	0.4008	*0.00528*			*0.0500*		
$G = 4$ (group 1)	−5.2110	0.0258	−0.4233	1.3886	0.3104	−966.37	1958.75
	1.8800	*0.0232*	*0.0705*	*0.1426*	*0.0780*		
(group 2)	0.3120	−0.0275	same	same	0.4449		
	0.4670	*0.00737*			*0.0842*		
(group 3)	0.5743	−0.00707	same	same	0.1853		
	0.5381	*0.00778*			*0.0665*		
(group 4)	−1.3950	0.0270	same	same	0.0594		
	1.0416	*0.0169*			*0.0559*		
$G = 5$						−966.29	1964.58

This is essentially a discrete random effects model such that $\boldsymbol{\beta} = \boldsymbol{\beta}_k = (\beta_{0k}, \beta_{1k}, \beta_{2k}, \beta_{3k})^T$ with probability π_k. This model has a tractable likelihood function which does not involve integration

$$\prod_{i=1}^{m} \Pr\{Y_{i1}, \ldots, Y_{in_i}\} = \prod_{i=1}^{m} \left[\sum_{k=1}^{G} \pi_k \left(\prod_{t=1}^{n_i} \frac{e^{Y_{it}\eta_{itk}}}{1 + e^{\eta_{itk}}} \right) \right]. \tag{19}$$

To estimate the parameters, it is convenient to use the EM algorithm. Specifically, we define $W_{ik} = 1$ if patient i belongs to group k and $W_{ik} = 0$ otherwise. Note that the W_{ik} are unobserved since the group membership of each patient is unknown. The log-likelihood function based on the so-called 'complete' data $(\mathbf{Y}, \mathbf{W})$ is

$$\ell(\boldsymbol{\theta}; \mathbf{Y}, \mathbf{W}) = \sum_{k=1}^{G} \ln \pi_k \sum_{i=1}^{m} W_{ik} + \sum_{k=1}^{G} \sum_{i=1}^{m} \sum_{t=1}^{n_i} W_{ik} Y_{it} \eta_{itk} - \sum_{k=1}^{G} \sum_{i=1}^{m} \sum_{t=1}^{n_i} W_{ik} \ln(1 + e^{\eta_{itk}}), \tag{20}$$

where $\boldsymbol{\theta}$ denotes the parameter vector $(\boldsymbol{\beta}_1^T, \ldots, \boldsymbol{\beta}_G^T, \pi_1, \ldots, \pi_{G-1})^T$. At the E-step of the EM algorithm, we estimate the unknown W_{ik} by the conditional expectation,

$$\widehat{W}_{ik} = \mathrm{E}(W_{ik} \mid \mathbf{Y}) = \frac{\pi_k \prod_{t=1}^{n_i} e^{Y_{it}\eta_{itk}}/(1 + e^{\eta_{itk}})}{\sum_{k=1}^{G} \pi_k \left[\prod_{t=1}^{n_i} e^{Y_{it}\eta_{itk}}/(1 + e^{\eta_{itk}}) \right]}, \tag{21}$$

evaluated at the current parameter estimates. At the M-step, we replace W_{ik} by $\widehat{W}_{ik}$ in (20) and maximize the resulting expression to obtain an updated estimate of the parameters. The procedure is iterated until convergence.

4.2. Numerical Results

We fit the mixture model with different numbers of groups to the data and the results are given in Table 3. It can be seen that there is almost no increase in likelihood beyond 3 groups which explains why the AIC criterion picks a 3 group mixture model with group specific intercept and dose coefficient as the best model. The decision to keep the time and autoregression coefficients fixed across the different groups is also based on the AIC criterion. Since the time coefficient is negative and significant, we conclude that there is an association between reduced drug use and increased duration in treatment. The dose coefficient is not significant in either group 1 ($\pi_1 = 0.30$) or group 3 ($\pi_3 = 0.24$) suggesting that heroin use is independent of daily methadone dose for patients in these groups. To gain more knowledge about the makeup of these groups, we look at $\widehat{W}_{ik}$, the estimated probability that subject i belongs to group k. The 11 subjects who are undoubtedly in group 1 ($\widehat{W}_{i1} \geq 0.9$) returned no heroin positive results out of a total of 258 screens. Thus subjects in group 1 are those who have ceased heroin use as a result of treatment. Individuals with $\widehat{W}_{i3} \geq 0.9$ are those with a persistently large number of heroin positive screens. There are 17 such subjects and they returned 184 heroin positive results out of 357 screens or 52% compared with 16% for all patients. These subjects respond poorly to treatment, with continued heroin use regardless of the methadone dose received. It is only in group 2 ($\pi_2 = 0.46$) that methadone dose is significant. Subjects in this group respond to treatment in a dose-dependent fashion with reduced heroin use at high methadone dose.

5. Discussion

The key finding in this paper is that a 3 group mixture model fits the data well. The first group comprises about 30% of the sample, and responds well to treatment, with cessation of heroin use. Methadone dose is not predictive of heroin positive urines in this group, consistent with clinical experience that there is a proportion of patients who have an excellent response to treatment, and can be satisfactorily maintained on quite modest doses of methadone. The second group, 46% of the sample, tends to use heroin when on lower methadone doses, but when maintained on high doses does quite well. Again, this fits well with clinical experience. Finally, there is a group, 24% in this study, which tends to continue to use heroin, regardless of methadone dose. It is well recognized that all treatment programs have to confront the problem of treatment failures, although the proportion of such failures probably varies according to the quality of other aspects of treatment (Ball & Ross, 1991). The current study is useful in identifying that methadone dose is of critical importance in influencing heroin use in around half the patients in treatment.

A referee raised the point that the methadone dose received by a patient is time-dependent. This is not a problem, as we can interpret the probabilities on the left hand side of (8) and (11) to be conditional on the entire history up to time t in which case (9) becomes a partial likelihood. In addition to the cutoff at 26 weeks used in the paper, we have also tried other cutoffs such as one year, two years and so on. The most notable observation is that with longer cutoff, the time–dose interaction which is positive in sign changes from nonsignificant to significant and this is not surprising. We feel that a 26-week cutoff is the most appropriate, as past experience and clinic policy suggest that nonrandom dropouts begin to occur after six months. A more satisfactory way to handle nonignorable dropouts is to model the dropout process and then use likelihood methods (see Follman & Wu, 1995; Fitzmaurice et al., 1995; Fitzmaurice et al., 1996).

References

BALL, J.C. & ROSS, A. (1991). *The Effectiveness of Methadone Maintenance Treatment: Patients, Programs, Services and Outcome.* New York: Springer-Verlag.

BONNEY, G.E. (1987). Logistic regression for dependent binary observations. *Biometrics* **43**, 951–973.

CAREY, V., ZEGER, S.L. & DIGGLE, P. (1993). Modelling multivariate binary data with alternating logistic regression. *Biometrika* **80**, 517–526.

COMMENGES, D., LETENNEUR, L., JACQMIN, H., MOREAU, T. & DARTIGUES, J. (1994). Test of homogeneity of binary data with explanatory variables. *Biometrics* **50**, 613–620.

D'AUNNO, T. & VAUGHN, T.E. (1992). Variations in methadone treatment practices: Results from a national study. *J. Amer. Med. Assoc.* **267**, 253–258.

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

—, — & ROTNITZKY, A.G. (1993). Regression models for discrete longitudinal responses (with discussion). *Statist. Sci.* **8**, 284–309.

—, MOLENBERGHS, G. & LIPSITZ, S.R. (1995). Regression models for longitudinal binary responses with informative drop-outs. *J. Roy. Statist. Soc. Ser. B* **57**, 691–704.

—, HEATH, A.F. & CLIFFORD, P. (1996). Logistic regression models for binary panel data with attrition. *J. Roy. Statist. Soc. Ser. A* **159**, 249–263.

FOLLMANN, D. & WU, M. (1995). An approximate generalized linear model with random effects for informative missing data. *Biometrics* **51**, 151–168.

HARGREAVES, W.A. (1983). Methadone dosage and duration for maintenance treatment. In *Research on the Treatment of Narcotic Addiction: State of the Art*, eds. J.R. Cooper, F. Altman, B.S. Brown, & D. Czechowicz. Maryland: NIDA, U.S. Department of Health and Human Sciences.

LIANG, K.Y. & ZEGER, S.L. (1986). Longitudinal data analysis using generalized linear models. *Biometrika* **73**, 13–22.

—, ZEGER, S.L. & QAQISH, B. (1992). Multivariate regression analyses for categorical data (with discussion). *J. Roy. Statist. Soc. Ser. B* **54**, 3–40.

MCGILCHRIST, C.A. (1994). Estimation in generalized mixed models. *J. Roy. Statist. Soc. Ser. B* **56**, 61–69.

NOVICK, D.M., JOSEPH, H., CROXSON, T.S., *et al.* (1990). Absence of antibody to human immunodeficiency virus in long-term, socially rehabilitated methadone maintenance patients. *Archives of Internal Med.* **150**, 97–99.

STIMSON, G.V. (1990). Aids and HIV: The challenge for British drug services. *British J. Addiction* **85**, 329–339.

WARD, J., MATTICK, R. & HALL, W. (1992). *Key Issues in Methadone Maintenance Treatment.* Kensington: The University of New South Wales Press.

ZHAO, L.P. & PRENTICE, R.L. (1990). Correlated binary regression using a quadratic exponential model. *Biometrika* **77**, 642–648.