Bayesian informative dropout model for longitudinal binary data with random effects using conditional and joint modeling approaches

Jennifer S. K. Chan*

Department of Statistics, The University of Sydney, NSW 2006, Australia

Received 31 March 2014; revised 15 July 2015; accepted 15 July 2015

Dropouts are common in longitudinal study. If the dropout probability depends on the missing observations at or after dropout, this type of dropout is called informative (or nonignorable) dropout (ID). Failure to accommodate such dropout mechanism into the model will bias the parameter estimates. We propose a conditional autoregressive model for longitudinal binary data with an ID model such that the probabilities of positive outcomes as well as the drop-out indicator in each occasion are logit linear in some covariates and outcomes. This model adopting a marginal model for outcomes and a conditional model for dropouts is called a selection model. To allow for the heterogeneity and clustering effects, the outcome model is extended to incorporate mixture and random effects. Lastly, the model is further extended to a novel model that models the outcome and dropout jointly such that their dependency is formulated through an odds ratio function. Parameters are estimated by a Bayesian approach implemented using the user-friendly Bayesian software WinBUGS. A methadone clinic dataset is analyzed to illustrate the proposed models. Result shows that the treatment time effect is still significant but weaker after allowing for an ID process in the data. Finally the effect of drop-out on parameter estimates is evaluated through simulation studies.

Keywords: Bayesian analysis; Conditional and joint model; Informative dropout; Longitudinal binary data; Selection model.

Additional supporting information including source code to reproduce the results may be found in the online version of this article at the publisher's web-site

1 Introduction

In longitudinal studies, attrition is a common problem. Subjects often drop out prematurely, particularly when the observation period is lengthy. Failure to obtain a full set of observations on a given unit may result in incomplete and unbalanced data and possibly loss of efficiency when the units with missing data are discarded. This research is motivated by a dataset of urine drug screens results, positive or negative to the presence of morphine, a biological marker of heroin, collected from patients in a methadone maintenance treatment (MMT) program at a clinic in Sydney in 1986. The data have substantial population heterogeneity and intracluster correlation as well as salient dropouts. Chan et al. (1998) and (1997) study the treatment effect for a single drug use and a couple of drug uses respectively allowing for population heterogeneity and clustering but they do not consider the dropout process. The aim of the study is to derive efficient dropout models to measure treatment effectiveness, after allowing for methadone dosage effect and dropouts.

*Corresponding author: e-mail: jchan@maths.usyd.edu.au, Phone: +612-93514873, Fax: +612-93514534

Dropout is a kind of monotone missings. While Laird (1988) and Troxel (l998) consider the intermittent missing, many analyses specifically look at monotone missings or dropouts (Fitzmaurice et al., 1995; Little, 1995). Rubin (1976) introduces different dropout processes that are summarized in Little and Rubin (1987) and Laird (1988). Completely random drop-out (CRD) refers to the dropout mechanism that does not depend upon the observed outcomes. If the dropout mechanism depends on the observed responses and possibly some covariates, but not the unobserved outcomes during or after the time of dropout, we have a random dropout (RD). CRD and RD are said to be *ignorable* because it is not necessary to specify a dropout model to obtain valid parameter inference in the outcome model. Likelihood-based methods using standard algorithms give consistent parameter estimates because the joint densities of observed data can be separated into densities involving parameters in the outcome model and parameters in the dropout model, respectively. Hence, the estimation of parameters in the outcome model can be done independently. On the contrary, if the probabilities of dropout depend on the unobserved outcomes, we have informative dropout (ID). In this case, a valid likelihood-based inference can only be obtained by specifying a dropout model. Diggle and Kenward (1994) demonstrate that there are biases in the parameter estimates if such dropout mechanism is not accommodated in the model.

Different linear-mixed models that incorporate dropout models have been proposed for normally distributed data (Crouchley and Ganjali, 2002; Park et al., 2002; Stubbendick and Ibrahim, 2003). However, there is less agreement on models for nonnormal data such as binary data or counts (Bradlow and Zaslavsky, 1999; Fitzmaurice et al., 2001). For binary data, Wu and Carroll (1988) use a probit censoring model, Cowles et al. (1996) assign informative priors directly to the missing data, Bradlow and Zaslavsky (1999) consider dropouts as a category of outcome variable, Alfo and Altkin (2000) model the total number of observations before dropout for each patient as one of the covariates in the outcome model (see Section 3.4.1), Albert and Follmann (2007) include shared random effects between the outcome and dropout models and Albert and Follmann (2003) adopt a multinomial logit transition model for the three events of observed, intermittent missing, and dropout. Instead, Chan et al. (2009) set up an explicit dropout model such that the probability of dropout for each outcome (observed or dropped out) is logit linear in some covariates and outcomes. This approach has an advantage: the explicit dropout model can be applied to data with nonmonotone missing and can incorporate covariates that affect the missingness process. To model an ID process using a selection model approach (Rubin, 1976), the conditional probability of dropout will depend on the unobserved outcome at the time of dropout. Chan and Wan (2011) adopt this selection model to the bivariate outcome model of Chan et al. (1997) for two drug uses. Adopting the same bivariate model of Chan et al. (1997), this paper advances the selection model to a joint model for both the probabilities of outcome and dropout so that the dependency between the two components that captures an ID process is now modeled by an odds ratio model. This innovative modeling approach has an advantage that marginal instead of conditional probability of dropout can be obtained and the odds ratio model provides greater flexibility in modeling the dependency.

To implement the dropout models, various methodologies have been proposed and they include the weighted generalized estimating equations approach (Robins et al., 1995; Preisser et al., 2002), the Monte Carlo EM approach (Verzilli and Carpenter, 2002), the classical maximum likelihood (ML) approach (Ten Have et al., 2002) and the semiparametric approach (Robins et al., 1995; Rotnitzky et al., 1998). However, the ML approach for random effects model with an ID modeling is not straightforward since inclusion of random effects complicates the likelihood function considerably. Chib (1995) suggests using the Gibbs output (Geman and Geman, 1984; Gelfand and Smith, 1990) in calculating the marginal likelihood function. Chib and Jeliazkov (2001) further investigate the use of Metropolis Hastings output (Metropolis, 1953; Hastings, 1970) when the full conditional densities are nonstandard. Chan et al. (2005) propose the method of Monte Carlo (MC) approximation using Gibbs output in the likelihood approach and Chan et al. (2009) apply the method to implement the selection model. Chan and Wan (2011) consider the Bayesian approach for the selection model with bivariate outcomes.

Comparing to classical likelihood approach, the Bayesian approach using Markov chain Monte Carlo (MCMC) algorithm converts an optimization problem into a sampling problem, thus avoiding the numerical difficulties associated with the maximization of complicated high-dimensional likelihood functions, such as those with random effects. This is done by iterative simulation of model parameters conditional on other parameters and the data. In case of nonstandard posterior distributions, MCMC method (Smith and Roberts, 1993; Gilks et al., 1996) with Gibbs sampling and Metropolis Hastings algorithm produces samples from the intractable posterior distributions of all unknown parameters. Moreover, the emergence of `WinBUGS`, a user-friendly software for Bayesian analysis using MCMC techniques, makes the parameter estimation much easier for nonexpert (Spiegelhalter et al., 2000). Taking into account the computational efficiency, convenience of implementation, popularity in the literature and the flexibility for model extension, we adopt the Bayesian MCMC approach.

Along the line of model development, we first develop the conditional autoregressive (AR) model with an explicit dropout model for longitudinal binary data and extend it to incorporate mixture and random effects to capture population heterogeneity and intercluster correlation. Bayesian MCMC approach avoids the complication in approximating the likelihood functions with random effects (Chan et al., 2009). We also adopt the mixture approach of Chan (2000) to deal with the initial stage problem in all AR models. Moreover the random effects are set to be end-(drug)use dependent because, as discussed in Section 3, there is substantial evidence suggesting that the heterogeneity of heroin use and dropout process are both end-use dependent. Lastly, we extend the conditional dropout model with a marginal outcome model to a joint model using the bivariate model of Chan et al. (1997). As our proposed models should provide robust inference on parameters even under moderate departures from modeling assumptions on the dropout mechanism, this is investigated through a simulation experiment.

The paper is presented as follows. Section 2 reviews different drop-out models. Section 3 introduces a methadone clinic data that is analyzed to evaluate the treatment effects using all proposed models. Section 4 describes the two types of AR models with an ID modeling: the selection model and joint model on univariate and bivariate settings, respectively. Then the methodology of Bayesian inference is explained in Section 5. Section 6 reports the numerical results with discussion. A simulation study is performed in Section 7 to assess the sensitivity of alternative assumptions of the dropout process on parameter inference. Finally, Section 8 gives concluding remarks.

2 The dropout model

There are different ways to model the dropout process, for example, Alfo and Aitkin (2000) modeled the total number of observed events for each subject. We model the probability of dropout in each occasion and hence the model allows any sequences of missing data, say 1,1,0,0,1 where "1" represents missing and "0" represents observed. Consequently, the model can be applied to monotone missing as well as nonmonotone missing. To model both the outcome and dropout process, Rubin (1976) outlined the selection model and pattern mixture model. Distinction between the two models lies on the way to factorize the joint probabilities of outcomes and dropouts. A selection model specifies a marginal model for outcomes $\boldsymbol{y}$ and a conditional model for dropouts $\boldsymbol{w}$ given outcomes using the factorization:

$$f(\boldsymbol{y}, \boldsymbol{w}|\boldsymbol{\theta}) = f(\boldsymbol{y}|\boldsymbol{\theta}) f(\boldsymbol{w}|\boldsymbol{y}, \boldsymbol{\theta})$$

whereas a pattern mixture model uses an alternative factorization:

$$f(\boldsymbol{y}, \boldsymbol{w}|\boldsymbol{\theta}) = f(\boldsymbol{w}|\boldsymbol{\theta}) f(\boldsymbol{y}|\boldsymbol{w}, \boldsymbol{\theta}),$$

containing a marginal model for dropouts $\boldsymbol{w}$ combined with a conditional model for outcomes $\boldsymbol{y}$ given dropouts (Glynn et al., 1986; Little, 1994; Michiels et al., 2002). Measurements for dropouts $\boldsymbol{w}$ can

be the times of dropout or indicators of dropout in each occasion. The shared parameter approach adopts some common parameters such as random effects into the marginal models for both outcomes and dropouts, $f(\boldsymbol{y}|\boldsymbol{\theta})$ and $f(\boldsymbol{w}|\boldsymbol{\theta})$ respectively, to characterize the relationship between measurement and missingness and induce their conditional dependence (Dunson and Perreault, 2001; Ten Have et al., 2002).

Selection model is intuitively more appealing because marginal model is adopted for the parameters of interest in the outcome model. It enables studies of marginal treatment effects and facilitates treatments comparison. On the other hand, the pattern mixture model measures treatment effects condition on dropouts (Little, 1993; Demirtas and Schafer, 2002; Pauler et al., 2003). To evaluate marginal treatment effects, one may need to average the conditional likelihoods for the observed data across probabilities of dropouts. Since we are interested in studying marginal treatment effect, we adopt a selection model approach. As the model implicitly fills in some missing outcomes based on the dropout model, validity of the analyses depends on the use of correct dropout model. In fact, both selection model and pattern mixture model rely on some *untestable* assumptions upon the dropout mechanism and/or its influence on the observed outcomes. The sensitivity of our proposed model to different dropout assumptions is investigated in Section 7 through simulation experiments.

3 The methadone clinic data

This research is motivated by a methadone clinic data of the results of urine drug screens from patients in a methadone maintenance treatment (MMT) program at a clinic in Sydney in 1986 (Chan et al., 1997, 1998). Outcomes are the weekly urine test results that are positive or negative for morphine, a biological marker for heroin use. The analysis is performed using a restricted dataset in which patients who completed less than four weeks of treatment are excluded because they may not have received adequate methadone treatment for measuring the treatment effects on early dropouts. Patients with missing dose records are also excluded from our analysis. Finally, past experience showed that the treatment was most effective in the first half year of maintenance and consequently, our study includes only results of urine screens collected in the first 26 weeks of treatment so as to avoid the distorting effect from patients being on a withdrawal regimen, something that usually begins after the first half year of maintenance.

There are 136 heroin users, submitting a total of 2872 urine screens ($y_{ij}, i = 1, \ldots, 136; j = 1, \ldots, n_i, n_i \leq 26$) with 16% of them being positive for heroin. Table 1 reports the summary statistics of those dropout ($n_i < 26$) and nondropout patients cross-classified with (1) initial ($y_{i1} = 1$) and noninitial user groups and (2) end ($y_{i,n_i} = 1$) and nonend user groups. From Table 1, the overall average dosage of methadone is 64 mg and the average number of treatment weeks per patient is 21.1 (each patient stayed 4–26 weeks). Fifty-one of them dropped out before the end of 26 weeks according for 37.5% of all patients. For all analyses, each urine screen result rather than each patient serves as the unit of analysis.

As previous research revealed that methadone dosage and duration of treatment were significant treatment factors, methadone dose d in mg at the time of urine test and log of duration of treatment $\ln t$ in weeks are included as predictor variables in all outcome models described in Section 4. As outcomes are longitudinal measurements over time, a first-order AR term is also added to the outcome models. Interaction between dose and time effects and higher order AR terms are found to be insignificant and are excluded from the models subsequently. Results of separate fitting to each patient and score tests suggested that there is substantial between-patient variation (Chan et al., 1998). To account for the population heterogeneity and to facilitate subject-specific inference, we incorporate mixture and random effects into the models to allow for different drug taking behaviors across patients. Table 1 suggests, in general, higher heroin use, lower dosage, shorter treatment duration, and higher dropout rate among dropout, initial-use, and end-use patient groups. Figure 1A and B display their variabilities across time with respect to the proportions of heroin use and dropout rates, respectively. Figure 1A

Table 1 Summary statistics for the methadone clinic dataset.

	Heroin use	Initial $Y_{i1}=1$	Not $Y_{i1}=0$	End $Y_{in_i}=1$	Not $Y_{in_i}=0$	Overall
Dropout	Number of patients	12	39	15	36	51
($n_i < 26$)	Average dose	56.7	67.8	59.2	67.7	65.3
	Weeks in treatment	12.5	13.1	12.2	13.3	13.0
	Percentage of positive test	46.7	19.7	38.8	20.9	25.8
Nondropout	Number of patients	25	60	6	79	85
($n_i = 26$)	Average dose	59.4	66.0	49.8	65.1	64.1
	Weeks in treatment	26	26	26	26	26
	Percentage of positive test	25.8	8.3	34.0	11.9	13.5
Total	Number of patients	37	99	21	115	136
	Average dose	58.9	66.5	54.9	65.6	64.4
	Weeks in treatment	21.6	20.9	16.1	22.0	21.1
	Percentage of positive test	29.8	11.1	36.6	13.6	16.3
	Percentage of dropout	32.4	39.4	71.4	31.3	37.5

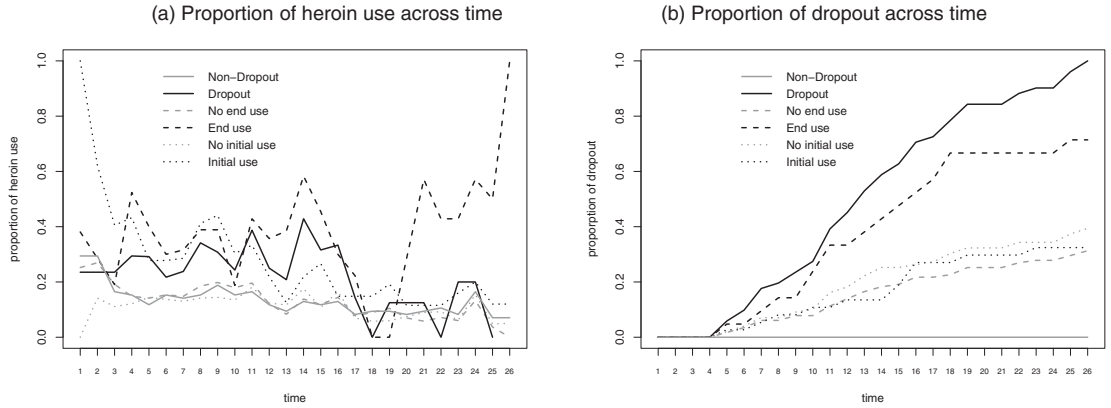

Figure 1 (A) Proportion of heroin use across time. (B) Proportion of dropout across time.

shows heroin uses in the dropout and nondropout groups converge approximately after week 17 when the dropout rate is high. As such selective attitude toward dropout among the heavy heroin users may lead to a false treatment effect of reduced heroin use over time if the reduced use is primarily due to the dropout of heavy users, the dropout process should be properly accounted for in the model. Moreover, there is substantial variation in heroin uses and dropout rates across time between the end-use and non end-use groups. To account for the heterogeneity, we will adopt an end-use specific mixture and random effects models as well as an ID model in the next section.

4 The models

Let $y_{it}(i = 1, \ldots, m; t = 1, \ldots, n_i)$ denote the observed outcome of urine test for the i-th patient at week t and let $N = n_1 + \cdots + n_m$ denote the total number of observations. The vector of all $n = 26$

possible outcomes for patient i can be separated into

$$\boldsymbol{y}_i = (y_{it})^T = (\underbrace{y_{i1}, \ldots, y_{i,n_i}}_{\text{Observed } \boldsymbol{y}_{o,i}^T}, \underbrace{y_{i,n_i+1}, \ldots y_{i,n}}_{\text{Unobserved } \boldsymbol{y}_{m,i}^T})^T$$

where n_i denotes the number of observed y_{it} and the vector of observed outcomes is denoted by $\boldsymbol{y}^T = (\boldsymbol{y}_{o,1}^T, \boldsymbol{y}_{o,2}^T, \ldots, \boldsymbol{y}_{o,m}^T)$.

Similarly, let W_{it} denote the drop-out indicator for patient i in week t such that $W_{it} = 1$ if Y_{it} is unobserved ($t > n_i$) and zero otherwise. Then the vector of drop-out indicators is $\boldsymbol{w}^T = (\boldsymbol{w}_1^T, \ldots, \boldsymbol{w}_m^T)$ where $\boldsymbol{w}_i^T = (0, \ldots, 0, I_{w,i})$ for patient i is a series of $(n_i - 1)$ "0" for $t = 2, \ldots, n_i$ followed by $I_{w,i} = I(n_i < 26)$, an indicator of whether patient i drops out from treatment before week 26.

4.1 Conditional AR model

We model both outcome and dropout process simultaneously using the selection model approach. For the outcome model, we model the serial correlation using an AR model such that the conditional probabilities of heroin use are logit linear in some covariates including the duration of treatment in weeks t, the dosage administered d_{it} in mg and the "previous outcomes" $y_{i,t-1}$. Writing $\Pr(Y_{it} = 1 | y_{i,t-1}, \boldsymbol{\beta}) = p_{y,it}$, we have

$$\text{logit}(p_{y,it}) = \eta_{it} = \beta_0 + \beta_d d_{it} + \beta_t \ln t + \beta_{pv} y_{i,t-1}, \tag{1}$$

$t = 2, \ldots, n_i$ where $\boldsymbol{\beta} = (\beta_0, \beta_d, \beta_t, \beta_{pv})^T$ is a q-dimensional ($q = 4$) vector of parameters. However when $t = 1$, Y_{i0} is unobserved. Alfo and Aitkin (2000) fitted AR models for $t \geq 2$ and added an interaction term, $Y_{i1} \times d_{it}$ (Tables 2 and 5) or initial outcome specific group probabilities, π_{k0} and π_{k1} (Table 3) to the mixture models. To avoid loss of information and difficulties in parameter interpretation, we adopt the mixture approach in Chan (2000) to solve the initial stage problem in AR models. Essentially, we assume Y_{i0} follows a Bernoulli distribution, that is

$$Y_{i0} \sim B(\pi_{y,0}) \quad \text{with} \quad \pi_{y,0} = \Pr(Y_{i0} = 1) \tag{2}$$

being constant across patients since there is no data information when $t = 0$. Writing $p_{y,i1b} = \Pr(Y_{i1} = 1 | Y_{i0} = b, \boldsymbol{\beta})$, we have

$$\text{logit}(p_{y,i1b}) = \eta_{i1b} = \beta_0 + \beta_d d_{i1} + \beta_{pv} b, \tag{3}$$

where $y_{i0} = b = 0, 1$ is an estimate of Y_{i0}. For the dropout model, we model the conditional probabilities of dropout as logit linear in some covariates as well as the "present outcomes" y_{it} that signify an ID process. Writing $\Pr(W_{it} = 1 | y_{it}, \boldsymbol{\alpha}) = p_{w,it}$, $2 \leq t \leq n_i + 1$, we have

$$\text{logit}(p_{w,it}) = \zeta_{it} = \alpha_0 + \alpha_t \ln t + \alpha_{ps} y_{it} \tag{4}$$

where the dropout indicators $W_{it} = 0$ for $2 \leq t \leq n_i$ when Y_{it} are observed. Then the s-dimensional ($s = 3$) vector of parameters in the dropout model is $\boldsymbol{\alpha} = (\alpha_o, \alpha_t, \alpha_{ps})^T$. At the time of dropout when $t = n_i + 1$ and $n_i < 26$, Y_{i,n_i+1} is unobserved. Writing $p_{w,i,n_i+1,b} = \Pr(W_{i,n_i+1} = 1 | Y_{i,n_i+1} = b, \boldsymbol{\alpha})$, we have

$$\text{logit}(p_{w,i,n_i+1,b}) = \zeta_{i,n_i+1,b} = \alpha_0 + \alpha_t \ln(n_i + 1) + \alpha_{ps} b \tag{5}$$

where $y_{i,n_i+1} = b = 0, 1$ as an estimate of the missing Y_{i,n_i+1} is given by (1). Hence the vector of model parameters is $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \pi_{y,0})^T$ and the vector of missing data is $(\boldsymbol{y}_0^T, \boldsymbol{y}_{n_i+1}^T)^T$ where $\boldsymbol{y}_t =$

Table 2 Parameter estimates and s.e. (in italics) for AR, MX, RI and BRI models with and without an ID modeling.

Model		Intc	Dose	Time	Prev	Pres	σ_0	σ_1	π_{y0}	π_1	*BIC*	*DIC*
AR	β	−1.19	−0.0092	−0.24	2.44				0.24		2186	2282
with ID		*0.26*	*0.0029*	*0.08*	*0.12*				*0.10*			
	α	−6.44		0.68		2.05						
		0.98		*0.26*		*0.82*						
AR	β	−1.23	−0.0086	−0.26	2.45				0.24		2064	2080
no ID		*0.26*	*0.0030*	*0.08*	*0.12*				*0.09*			
MX	β_1	−1.63	−0.015	−0.24	1.58				0.43	0.64	2004	1936
with ID		*0.62*	*0.008*	*0.09*	*0.14*				*0.23*	*0.06*		
	β_2	−0.52	0.0009	−0.24*	1.58*				0.33			
		0.32	*0.0042*	*0.09**	*0.14**				*0.19*			
	α	−6.76		0.73		2.38						
		1.05		*0.26*		*0.85*						
MX	β_1	−1.65	−0.013	−0.33	1.60				0.31	0.65	1867	1823
no ID		*0.63*	*0.008*	*0.09*	*0.14*				*0.16*	*0.06*		
	β_2	−0.42	0.0010	−0.33*	1.60*							
		0.34	*0.0043*	*0.09**	*0.14**							
RI	β	−0.90	−0.017	−0.25	1.46		1.76	2.37	0.37		**1838**	1975
with ID		*0.42*	*0.006*	*0.09*	*0.15*		*0.44*	*1.14*	*0.19*			
	α	−6.45		0.68		2.09						
		0.94		*0.28*		*0.70*						
RI	β	−1.02	−0.014	−0.32	1.47		1.79	2.35	0.32		1714	1867
no ID		*0.39*	*0.005*	*0.08*	*0.15*		*0.45*	*1.15*	*0.17*			
JRI	β	−1.08	−0.015	−0.37	1.23		2.98		0.34		1851	**1832**
with ID		*0.38*	*0.005*	*0.07*	*0.10*		*0.58*		*0.18*			
	α	−4.88		0.38								
		0.37		*0.14*								
	γ	−16.1	−1.62		50.0							
		0.98	*0.53*		*0.99*							

*Set to be constant across groups.
Models with *BIC* and/or *DIC* in bold are the best models among all with an ID modeling.

$(y_{1t}, \ldots, y_{mt})^T$, $t = 0$ or $n_i + 1$. The conditional likelihood function for the complete data is given by

$$f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}|\boldsymbol{\theta}) = \prod_{i=1}^{m} \left\{ \left[\prod_{b=0}^{1} \left(\pi_{y,0b} \frac{e^{y_{i1}\eta_{i1b}}}{1+e^{\eta_{i1b}}} \right)^{y_{i0b}} \right] \left[\prod_{t=2}^{n_i} \left(\frac{e^{y_{it}\eta_{it}}}{1+e^{\eta_{it}}} \right) \left(\frac{1}{1+e^{\zeta_{it}}} \right) \right] \times \right.$$
$$\left. \times \left[\prod_{b=0}^{1} \left(\left(\frac{e^{b\,\eta_{i,n_i+1}}}{1+e^{\eta_{i,n_i+1}}} \right) \left(\frac{e^{\zeta_{i,n_i+1,b}}}{1+e^{\zeta_{i,n_i+1,b}}} \right) \right)^{y_{i,n_i+1,b}} \right]^{I_{w,i}} \right\} \tag{6}$$

where $y_{it1} = y_{it}$, $y_{it0} = 1 - y_{it}$ for $t = 0, n_i + 1$, $\pi_{y,01} = \pi_{y,0}$ and $\pi_{y,00} = 1 - \pi_{y,0}$.

Table 3 Summary statistics for heavy and light use groups.

	Heroin use	Heavy $I_{gi} \leq 0.5$	Light $I_{gi} > 0.5$	Overall
Dropout	Number of patients	29	22	51
($n_i < 26$)	Average dose	63.0	68.5	65.3
	Weeks in treatment	13.1	12.8	13.0
	Average of u_i	1.26	−0.55	0.48
	Average of $\hat{y}_{i,n_i+1}$	0.75	0.32	0.57
	Percentage of positive test	42.3	3.6	25.8
Nondropout	Number of patients	22	63	85
($n_i = 26$)	Average dose	57.2	66.4	64.1
	Weeks in treatment	26	26	26
	Average of u_i	1.18	−0.76	−0.26
	Percentage of positive test	39.5	4.4	13.5
Total	Number of patients	51	85	136
	Average dose	59.5	66.7	64.4
	Weeks in treatment	18.7	22.6	21.1
	Average of u_i	1.23	−0.71	0.02
	Percentage of positive test	40.6	4.3	16.3
	Percentage of dropout	56.9	25.9	37.5

4.1.1 Extension to mixture model

To accommodate the substantial group effects across patients in the methadone clinic data, Chan et al. (1998) considered the mixture (MX) model. They assumed that $\boldsymbol{\beta}$ follows a multinomial distribution with $\boldsymbol{\beta} = \boldsymbol{\beta}_k = (\beta_{k0}, \beta_{kd}, \beta_{kt}, \beta_{kpv})^T$ at a probability $\pi_{g,k}$ for the k-th group of patients and selected the number of groups using Akaike Information Criterion (AIC). We adopt a two-group MX model ($k = 2$) for illustration and define the group-one membership indicator $I_{g,i} = 1$ if patient i belongs to group one and $I_{g,i} = 0$ otherwise. Note that $I_{g,i}$ is unobserved since the group membership of each patient is unknown. Based on Chan et al. (1998), we include group-specific intercept and dose effect while the time and previous outcome effects are set to be the same across groups in the outcome model. If patient i belongs to group k ($k = 1, 2$), the conditional probabilities $p_{y,itk} = \Pr(Y_{it} = 1 | y_{i,t-1}, \boldsymbol{\beta} = \boldsymbol{\beta}_k)$ are

$$\text{logit}(p_{y,itk}) = \eta_{itk} = \beta_{k0} + \beta_{kd}\, d_{it} + \beta_t \ln t + \beta_{pv}\, y_{i,t-1} \tag{7}$$

where $\boldsymbol{\beta}_k^* = (\beta_{ko}, \beta_{kd})^T$, $\boldsymbol{\beta}^* = (\beta_t, \beta_{pv})^T$ and $\boldsymbol{\beta}_k = (\boldsymbol{\beta}_k^{*T}, \boldsymbol{\beta}^{*T})^T$, $k = 1, 2$. When $t = 1$, Y_{i0} is again unobserved. Writing $p_{y,i1kb} = \Pr(Y_{i1} = 1 | Y_{i0} = b; \boldsymbol{\beta} = \boldsymbol{\beta}_k)$ conditioning on group k and $Y_{i0} = b$, we have

$$\text{logit}(p_{y,i1kb}) = \eta_{i1kb} = \beta_0 + \beta_{kd}\, d_{it} + \beta_{pv}\, b \tag{8}$$

where $y_{i0} = b = 0, 1$ is an estimate of Y_{i0} and $Y_{i0} \sim B(\pi_{y,0k})$ under the mixture approach of Chan (2000) with group-specific probability $\pi_{y,0k} = \Pr(Y_{i0} = 1 | \boldsymbol{\beta} = \boldsymbol{\beta}_k)$. For the dropout model, Eqs. (4) and (5) follow. Then the vector of model parameters is $\boldsymbol{\theta} = (\boldsymbol{\beta}_1^{*T}, \boldsymbol{\beta}^{*T}, \boldsymbol{\beta}_2^{*T}, \boldsymbol{\alpha}^T, \boldsymbol{\pi}_{y,0}, \pi_{g,1})^T$ and the vector of missing data is $(\boldsymbol{y}_0^T, \boldsymbol{y}_{n_i+1}^T, \boldsymbol{I}_g^T)^T$ where $\boldsymbol{I}_g = (I_{g,1}, \ldots, I_{g,m})^T$ and $\boldsymbol{\pi}_{y,0} = (\pi_{y,01}, \pi_{y,02})^T$. The

conditional likelihood function for the complete data is given by

$$
\begin{aligned}
&f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{I}_g|\boldsymbol{\theta}) = \\
&= \prod_{i=1}^{m} \left\{ \left[\prod_{k=1}^{2} \left(\pi_{g,k} \prod_{b=0}^{1} \left(\pi_{y,0kb} \frac{e^{y_{i1}\eta_{i1kb}}}{1+e^{\eta_{i1kb}}} \right)^{y_{i0b}} \prod_{t=2}^{n_i} \frac{e^{y_{it}\eta_{itk}}}{1+e^{\eta_{itk}}} \right)^{I_{g,ik}} \right] \left[\prod_{t=2}^{n_i} \frac{1}{1+e^{\zeta_{it}}} \right] \times \right. \\
&\quad \left. \times \left[\prod_{g=0}^{1} \left(\left[\prod_{k=1}^{2} \left(\pi_{g,k} \frac{e^{b\,\eta_{i,n_i+1,k}}}{1+e^{\eta_{i,n_i+1,k}}} \right)^{I_{g,ik}} \right] \frac{e^{\zeta_{i,n_i+1,b}}}{1+e^{\zeta_{i,n_i+1,b}}} \right)^{y_{i,n_i+1,b}} \right]^{I_{w,i}} \right\}
\end{aligned} \tag{9}
$$

where $\pi_{y,0k1} = \pi_{y,0k},\ \pi_{y,0k0} = 1 - \pi_{y,0k}, k = 1, 2, I_{g,i1} = I_{g,i,}$ and $I_{g,i2} = 1 - I_{g,i}$.

4.1.2 Extension to random intercept model

Since the methadone clinic data exhibits considerable individual variation, Chan et al. (1998) added a random intercept (RI) into the mean function. Then the conditional probabilities $p_{y,it} = \Pr(Y_{it} = 1|\, y_{i,t-1}, \boldsymbol{\beta}, \boldsymbol{u})$ in the outcome model for $t = 2, \ldots, n_i$ become

$$
\operatorname{logit}(p_{y,it}) = \eta_{it} = \beta_o + \beta_d\, d_{it} + \beta_t\, \ln t + \beta_{pv}\, y_{i,t-1} + \lambda_i, \tag{10}
$$

and

$$
\operatorname{logit}(p_{y,i1b}) = \eta_{i1b} = \beta_0 + \beta_d\, d_{i1} + \beta_{pv}\, b + \lambda_i \tag{11}
$$

for $t = 1$ where λ_i is a RI term and the vector of RIs is $\boldsymbol{\lambda} = (\lambda_1, \ldots, \lambda_m)^T$. The dropout models again follow Eqs. (4) and (5).

We further assume that the random effects λ_i depend on the end drug-use $Y_{i,n_i} = h$ in general, and follow a separate normal distribution $N(0, \sigma_h^2)$ with density $\phi(\cdot|\, 0, \sigma_h^2)$. A vector of parameters for the whole model is $\boldsymbol{\theta} = (\boldsymbol{\beta}^T, \boldsymbol{\alpha}^T, \pi_{y,0}, \boldsymbol{\sigma}^2)^T$ where $\boldsymbol{\sigma}^2 = (\sigma_0^2, \sigma_1^2)^T$. The conditional likelihood function $f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{\lambda}|\boldsymbol{\theta})$ for the complete data is given by

$$
\begin{aligned}
&f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{\lambda}|\boldsymbol{\theta}) = \\
&= \prod_{i=1}^{m} \left\{ \left[\prod_{b=0}^{1} \left(\pi_{y,0b} \frac{e^{y_{i1}\eta_{i1b}}}{1+e^{\eta_{i1b}}} \right)^{y_{i0b}} \right] \left[\prod_{t=2}^{n_i} \left(\frac{e^{y_{it}\eta_{it}}}{1+e^{\eta_{it}}} \right) \left(\frac{1}{1+e^{\zeta_{it}}} \right) \right] \times \right. \\
&\quad \left. \times \left[\prod_{b=0}^{1} \left(\left(\frac{e^{b\,\eta_{i,n_i+1}}}{1+e^{\eta_{i,n_i+1}}} \right) \left(\frac{e^{\zeta_{i,n_i+1,b}}}{1+e^{\zeta_{i,n_i+1,b}}} \right) \right)^{y_{i,n_i+1,b}} \right]^{I_{w,i}} \left[\prod_{b=0}^{1} \phi\left(\lambda_i|\, 0, \sigma_b^2\right)^{y_{in_ib}} \right] \right\}
\end{aligned} \tag{12}
$$

where $y_{in_i1} = y_{in_i}$ and $y_{in_i0} = 1 - y_{in_i}$.

4.2 Joint model with random intercept

Chan and Wan (2011) applied a bivariate model to describe two drug uses jointly. We adopt this approach to model jointly the heroin use and dropout process because bivariate model can better describe their association. Similar to the univariate model, this model allows easy extension to incorporate mixture and random effects to account for population heterogeneity and facilitate subject-specific inference. We consider a joint RI (JRI) model and define simultaneously the conditional probabilities $p_{y,it} = \wp_{1t}(1,\cdot) = \Pr(Y_{it} = 1 \mid y_{i,t-1}, \boldsymbol{\beta})$, $p_{w,it} = \wp_{1t}(\cdot,1) = \Pr(W_{it} = 1 \mid y_{i,t-1}, \boldsymbol{\alpha})$ as well as the odds ratio

$$\psi_{it} = \frac{\wp_{1t}(1,1)\,\wp_{1t}(0,0)}{\wp_{1t}(1,0)\,\wp_{1t}(0,1)}$$

for the bivariate outcomes Y_{it} and W_{it} using the models

$$\text{logit}(p_{y,it}) = \eta_{it} = \beta_o + \beta_d\, d_{it} + \beta_t\, \ln t + \beta_{pv}\, y_{i,t-1} + \lambda_i, \tag{13}$$

$$\text{logit}(p_{w,it}) = \zeta_{it} = \alpha_o + \alpha_t\, \ln t, \tag{14}$$

$$\ln(\psi_{it}) = \varsigma_{it} = \gamma_o + \gamma_d\, d_{it} + \gamma_c\, C_{i,t-1} \tag{15}$$

for $t = 2, \ldots, n_i + 1$ and the random intercept $\lambda_i \sim N(0, \sigma^2)$. Note that the JRI model has an odds ratio model (15) to model the dropout dynamic as related to the outcomes. To reduce the model complexity, we do not adopt an end-use specific random effect model as in the RI model. When $t = 1$, Y_{i1} is modeled marginally instead of jointly with w_{i1} since $w_{i1} = 1$ always. Then $p_{y,i1b} = \Pr(Y_{i1} = 1 |\, Y_{i0} = b, \boldsymbol{\beta})$ is given by (11) where $Y_{i0} \sim B(\pi_{y,0})$ adopting again the mixture approach of Chan (2000). Note that while the data model for $p_{y,it}$ follows Eq. (1), the dropout model for $p_{w,it}$ differs from (4) by dropping the current heroin use term $\alpha_{ps}\, y_{it}$ because the dependency of W_{it} on Y_{it} is now modeled by the odds ratio in the bivariate model. In the odds ratio model, the *concordance indicator*, $C_{i,t-1} = 1$ if $Y_{i,t-1} = W_{i,t-1}$ and 0 otherwise, is included as a covariate to model the autoregressive effect for cross-correlation. This model also has a tractable likelihood function that facilitates likelihood inference. The joint probabilities $\wp_{uv,it} = \Pr(Y_{it} = u,\ W_{it} = v |\, y_{i,t-1}, \boldsymbol{\theta}),\ u, v = 0, 1$ where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \pi_{y,0}, \sigma^2)$ can be obtained by first expressing $\wp_{11,it}$ in terms of $p_{y,it}$, $p_{w,it}$ and the odds ratio ψ_{it} (Fleiss, 1981, p.68) as follows

$$\begin{aligned}\wp_{11,it} &= f_{yw}(1,1|p_{y,it}, p_{w,it}, \psi_{it}) = \\ &= \frac{1}{2(\psi_{it}-1)}\Big\{(\psi_{it}-1)[p_{y,it}+p_{w,it}] + 1 - \{1 + (\psi_{it}-1)[\psi_{it}[p_{y,it} + \\ &\quad -p_{w,it}]^2 - [p_{y,it} + p_{w,it}(\cdot,1)]^2 + 2[p_{y,it} + p_{w,it}]]\}^{\frac{1}{2}}\Big\},\end{aligned} \tag{16}$$

$$\wp_{10,it} = f_{yw}(1,0|p_{y,it}, p_{w,it}, \psi_{it}) = p_{y,it} - f_{yw}(1,1|p_{y,it}, p_{w,it}, \psi_{it}), \tag{17}$$

$$\wp_{01,it} = f_{yw}(0,1|p_{y,it}, p_{w,it}, \psi_{it}) = p_{w,it} - f_{yw}(1,1|p_{y,it}, p_{w,it}, \psi_{it}), \tag{18}$$

$$\wp_{00,it} = f_{yw}(0,0|p_{y,it}, p_{w,it}, \psi_{it}) = 1 - p_{y,it} - p_{w,it} + f_{yw}(1,1|p_{y,it}, p_{w,it}, \psi_{it}). \tag{19}$$

For simplicity, we write $f_{yw}(u,v)$ for $f_{yw}(u,v|p_{y,it}, p_{w,it}, \psi_{it})$ in all subsequent equations. Then the complete data likelihood $f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{\lambda}|\boldsymbol{\theta})$ is given by

$$\begin{aligned}f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{\lambda}|\boldsymbol{\theta}) = \prod_{i=1}^{m}\Bigg\{\Bigg[\prod_{b=0}^{1}\left(\pi_{y,0b}\frac{e^{y_{i1}\eta_{i1b}}}{1+e^{\eta_{i1b}}}\right)^{y_{i0b}}\Bigg] \times \\ \times\left(\prod_{t=2}^{n_i} f_{yw}(y_{it},0)\right) f_{yw}(y_{i,n_i+1},1)^{I_{w,i}}\phi(\lambda_i|0,\sigma^2)\Bigg\}\end{aligned} \tag{20}$$

where $y_{i01} = y_{i0},\ y_{i00} = 1 - y_{i0}$, $\pi_{y,01} = \pi_{y,0}$ and $\pi_{y,00} = 1 - \pi_{y,0}$.

5 Bayesian methodology

Despite the theoretical appeal, classical ML estimation of model parameters is difficult as numerical methods must be used to evaluate some complicated marginal likelihood function of the RI model in (10) which involves integration over the vector of latent RIs $\boldsymbol{\lambda}$ from the joint density function. In recent years, simulation-based Bayesian MCMC methods become a routine tool for a wide range of complicated statistical models. When the sample size is large, the Bayesian estimator is asymptotically equivalent to ML estimator under appropriate regularity conditions (Ghosal et al., 1995).

In the Bayesian approach, the posterior distribution of unknown parameters is generated, incorporating both data and prior information for parameters. Since prior information is unavailable, noninformative priors with large variance, normal for unrestricted parameters, and inverse gamma (IG) priors for positive parameters such as the variance σ^2, are adopted. The MCMC algorithm consists of constructing an irreducible and aperiodic Markov chain, whose equilibrium distribution is the desired joint posterior distribution. Gibbs sampler can be applied to generate a sequence of samples of one or more variables at a time from the set of full conditional distributions. Outputs from the simulated chain are used for posterior analysis, for example, parameters are estimated by their posterior means. If the full conditional distributions are not standard, techniques such as Metropolis Hastings may be used. The MCMC algorithm using Gibbs sampler can be easily implemented using the Bayesian software `WinBUGS`. Full sets of conditional distributions and WinBUGS command codes are available upon request.

Due to the complexity of the models, high posterior correlations exist between some parameters. These dependencies may slow down the convergence rate in the Gibbs samplers. As a result, the number of iterations I should be large enough to ensure that the sample is stationary. We set $I = 15{,}000$ and the burn-in period is at least $T = 5000$ iterations. After the burn-in period, parameters are drawn from every 10th iteration to mimic a random sample of size at least $M = 1000$ from the intractable joint posterior distribution. Trajectory plots and autocorrelation plots of the simulated values are used to check for independence and convergence of the sample. Then we list below the Bayesian hierarchies and joint posterior distributions for all models.

The Bayesian hierarchy for the AR model is

$$\text{Observed data model: } y_{it} \sim B(p_{y,it}), \quad 1 \le t \le n_i, \tag{21}$$

$$\text{Missing data model: } y_{it} \sim B(p_{y,it}), \quad t = 0, n_i + 1, \tag{22}$$

$$\text{Dropout model: } w_{it} \sim B(p_{w,it}), \quad 2 \le t \le n_i + 1, \tag{23}$$

where $p_{y,i0} = \pi_{y,0}$, $w_{it} = 0$ for $t \le n_i$, $w_{i,n_i+1} = 1$ for $n_i < 26$ and $p_{y,it}$, $p_{y,i1b}$, $p_{w,it}$ and $p_{w,i,n_i+1,b}$ are given by (1), (3), (4), and (5), respectively. The priors are

$$\beta_0, \beta_d, \beta_t, \beta_{pv}, \alpha_0, \alpha_t, \alpha_{ps} \sim N(0, a) \quad \text{and} \quad \pi_{y,0} \sim U(0, 1) \tag{24}$$

where $U(\cdot)$, $B(\cdot)$, and $N(\cdot)$ represent uniform, Bernoulli, and normal distributions, respectively, and the hyperparameter a is set to be large such as 10^6. The complete data likelihood $f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}|\boldsymbol{\theta})$ is given by Eq. (6). Together with the prior distributions, the joint posterior distribution for $\boldsymbol{\Theta} = (\boldsymbol{\theta}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1})$ where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \pi_{y,0})$ is given by

$$f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \pi_{y,0}) = f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}|\boldsymbol{\theta}) \left[\prod_{j=1}^{4} f_N(\beta_j)\right] \left[\prod_{j=1}^{3} f_N(\alpha_j)\right] f_U(\pi_{y,0}).$$

The Bayesian hierarchy for the MX model is given by (21) to (23) and

$$\text{Missing data model:} \quad I_{g,i} \sim B(\pi_{g,1})$$

$$\begin{aligned}\text{where} \quad \text{logit}(p_{y,it}) &= I_{g,i}\,(\beta_{10} + \beta_{1d}\,d_{it} + \beta_t\,\ln t + \beta_{pv}\,y_{i,t-1}) + \\ &\quad + (1 - I_{g,i})(\beta_{20} + \beta_{2d}\,d_{it} + \beta_t\,\ln t + \beta_{pv}\,y_{i,t-1}),\ 2 \le\ t \le n_i + 1 \\ \text{logit}(p_{y,i1,y_{i0}}) &= I_{g,i}\,(\beta_{10} + \beta_{1d}\,d_{i1} + \beta_{pv}\,y_{i0}) + (1 - I_{g,i})(\beta_{20} + \beta_{2d}\,d_{i1} + \beta_{pv}\,y_{i0})\end{aligned}$$

and $p_{w,it},\ t = 2, \ldots, n_i$ and $p_{w,i,n_i+1,y_{i,n_i+1}}$ are given by Eqs. (4) and (5), respectively. The priors are

$$\begin{aligned}\beta_{10}, \beta_{1d}, \beta_t, \beta_{pv}, \beta_{2d}, \alpha_0, \alpha_t, \alpha_{ps} &\sim N(0, a), \\ \pi_{y,0k},\ \pi_{g,1} &\sim U(0,1),\ k = 1, 2.\end{aligned}$$

Label switching problem for the MX model is handled by setting $\beta_{20} > \beta_{10}$ so that the prior for β_{20} is $\beta_{20} \sim N(0, a)I_{(\beta_{10},\infty)}$ (Lee et al., 2008). The complete data likelihood function $f(\boldsymbol{y}, \boldsymbol{w}|\boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{I}_g, \boldsymbol{\theta})$ is given by Eq. (9). Together with the prior distributions, the joint posterior distribution for $\boldsymbol{\Theta} = (\boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\boldsymbol{I}_g, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\pi}_{y,0}, \pi_{g,1})$ is given by

$$\begin{aligned}&f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{I}_g, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\pi}_{y,0}, \pi_{g,1}) = \\ &\quad = f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{I}_g|\boldsymbol{\theta}) \left[\prod_{j=1}^{5} f_N(\beta_j) f_{N^*}(\beta_6)\right] \left[\prod_{j=1}^{3} f_N(\alpha_j)\right] \left[\prod_{j=1}^{2} f_U(\pi_{y,0j})\right] f_U(\pi_{g,1})).\end{aligned}$$

where $f_{N^*}(\beta_6)$ denote the truncated normal for β_{20}.

The Bayesian hierarchy for the RI model is given by (21) to (23) and

$$\begin{aligned}&\text{Missing data model:} \quad \lambda_i \sim N\left(0, \sigma^2_{y_{i,n_i}}\right) \\ &\quad \text{where} \quad \text{logit}(p_{y,it}) = \beta_0 + \beta_d\,d_{it} + \beta_t\,\ln t + \beta_{pv}\,y_{i,t-1} + \lambda_i, \quad 2 \le t \le n_i + 1, \\ &\qquad \text{logit}(p_{y,i1y_{i0}}) = \beta_0 + \beta_d\,d_{i1} + \beta_{pv}\,y_{i0} + \lambda_i,\end{aligned}$$

and again, $p_{w,it},\ t = 2, \ldots, n_i$ and $p_{w,i,n_i+1,y_{i,n_i+1}}$ are given by Eqs. (4) and (5), respectively. The priors are given by Eq. (24) and

$$\sigma_b^2 \sim IG(a_1, a_2), \quad b = 0, 1$$

where the hyperparameters a_1 and a_2 are set to be small such as 10^{-3}. The complete data likelihood function $f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{\lambda}|\boldsymbol{\beta}, \boldsymbol{\alpha}, \pi_{y,0}, \boldsymbol{\sigma}^2)$ is given by Eq. (12) where $\boldsymbol{\lambda}$ are additional missing data and $\sigma_b^2,\ b = 0, 1$ are additional model parameters for the RI model as compared to the AR model.

Lastly, the Bayesian hierarchy for the data and latent variables for the JRI model is

$$\begin{aligned} y_{i1} &\sim \text{Bernoulli}(p_{y,i1}), \\ (I^*_{11,it}, I^*_{10,it}, I^*_{01,it}, I^*_{00,it}) &\sim \text{MN}(\wp_{11,it}, \wp_{10,it}, \wp_{01,it}, \wp_{00,it}, 1), \\ \lambda_i &\sim N(0, \sigma^2), \end{aligned}$$

where MN denotes the multinomial distribution, $I^*_{uv,it} = I(y_{it} = u, w_{it} = v)$ for $t = 2, \ldots, n^*_i$ based on the data (y_{it}, w_{it}), $n^*_i = n_i(1 - I_{w,i}) + (n_i + 1)I_{w,i}$, and $p_{y,it}$, $p_{w,it}$, ψ_{it} and $\wp_{it}(u, v)$ are given by Eqs. (13), (14), (15), and (16)–(19), respectively. Note that at the time of dropout $t = n_i + 1$, we have $(\hat{y}_{it}, w_{it})$ where $\hat{y}_{it}$ is the predicted outcomes. The complete data likelihood $f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{\lambda}|\boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \pi_{y,0}, \sigma^2)$ is given by Eq. (20) and the priors are

$$\beta_0, \beta_d, \beta_t, \beta_{pv}, \alpha_0, \alpha_t, \gamma_0, \gamma_d, \gamma_c \sim N(0, a), \quad \pi_{y,0} \sim U(0, 1), \quad \text{and} \quad \sigma^2 \sim IG(a_1, a_2). \tag{25}$$

Together with the prior distributions, the joint posterior distribution for $\boldsymbol{\Theta} = (\boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{\theta})$ where $\boldsymbol{\theta} = (\boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \pi_{y,0}, \sigma^2)$ is given by

$$\begin{aligned} &f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{\lambda}, \boldsymbol{\beta}, \boldsymbol{\alpha}, \boldsymbol{\gamma}, \pi_{y,0}, \sigma^2) = \\ &= f(\boldsymbol{y}, \boldsymbol{w}, \boldsymbol{y}_0, \boldsymbol{y}_{n_i+1}, \boldsymbol{\lambda}|\boldsymbol{\theta}) \times \\ &\quad \times \left[\prod_{j=1}^{4} f_N(\beta_j)\right] \left[\prod_{j=1}^{2} f_N(\alpha_j)\right] \left[\prod_{j=1}^{3} f_N(\gamma_j)\right] f_U(\pi_{y,0}) f_{IG}(\sigma^2). \end{aligned}$$

6 Results

To compare the performance of each model, Bayesian Information Criterion (BIC) and Deviance Information Criterion (DIC) (Spiegelhalter et al., 2002) defined as

$$\begin{aligned} &\text{BIC} = -2 \ln \mathsf{f}(\boldsymbol{y}, \boldsymbol{w}|\bar{\boldsymbol{\theta}}) + p \ln N, \\ &\text{and} \quad \text{DIC} = \overline{D(\boldsymbol{\theta})} + p_D = -\frac{4}{M} \sum_{l=1}^{M} \ln \mathsf{f}(\boldsymbol{y}, \boldsymbol{w}|\boldsymbol{\theta}^{(l)}) + 2 \ln \mathsf{f}(\boldsymbol{y}, \boldsymbol{w}|\bar{\boldsymbol{\theta}}), \end{aligned} \tag{26}$$

are often used in Bayesian analysis. Both criteria contain two components: a measure of model fit and a penalty for model complexity where $\overline{D(\boldsymbol{\theta})} = E_{\theta|y,w}[D(\boldsymbol{\theta})]$ is the posterior expectation of deviance $D(\boldsymbol{\theta}) = -2 \ln \mathsf{f}(\boldsymbol{y}, \boldsymbol{w}|\boldsymbol{\theta})$, $\mathsf{f}(\boldsymbol{y}, \boldsymbol{w}|\boldsymbol{\theta})$ is the likelihood function, p is the number of model parameters, $N = 2 \times 2872 - 136 = 5608$ is the number of observations, $p_D = \overline{D(\boldsymbol{\theta})} - D(\bar{\boldsymbol{\theta}})$ is the effective number of parameters, $\bar{\boldsymbol{\theta}} = \frac{1}{M} \sum_{l=1}^{M} \boldsymbol{\theta}^{(l)}$, $\boldsymbol{\theta}^{(l)}$ is the l-th estimate of $\boldsymbol{\theta}$ in the posterior sample and M ($M = 1000$) is the posterior sample size. The likelihood functions $\mathsf{f}(\boldsymbol{y}, \boldsymbol{w}|\boldsymbol{\theta})$ for the four models are given by:

$$\begin{aligned} \text{AR \& RI} : \mathsf{f}(\boldsymbol{y}, \boldsymbol{w}|\boldsymbol{\theta}) = \prod_{i=1}^{m} \Bigg\{ &\left[\left(\frac{e^{y_{i1}\eta_{i11}}}{1 + e^{\eta_{i11}}} \right) y_{i0} + \left(\frac{e^{y_{i1}\eta_{i10}}}{1 + e^{\eta_{i10}}} \right) (1 - y_{i0}) \right] \times \\ &\times \prod_{t=2}^{n_i} \left[\left(\frac{e^{y_{it}\eta_{it}}}{1 + e^{\eta_{it}}} \right) \left(\frac{1}{1 + e^{\zeta_{it}}} \right) \right] \Bigg\}, \end{aligned}$$

$$\text{MX}: \quad \mathsf{f}(\boldsymbol{y}, \boldsymbol{w}|\boldsymbol{\theta}) = \prod_{i=1}^{m} \left\{ I_{g,i} \left[\left(\frac{e^{y_{i1}\eta_{i111}}}{1+e^{\eta_{i111}}} \right) y_{i0} + \left(\frac{e^{y_{i1}\eta_{i110}}}{1+e^{\eta_{i110}}} \right) (1-y_{i0}) \right] \prod_{t=2}^{n_i} \left(\frac{e^{y_{it}\eta_{it1}}}{1+e^{\eta_{it1}}} \right) + \right.$$

$$\left. + (1-I_{g,i}) \left[\left(\frac{e^{y_{i1}\eta_{i121}}}{1+e^{\eta_{i121}}} \right) y_{i0} + \left(\frac{e^{y_{i1}\eta_{i120}}}{1+e^{\eta_{i120}}} \right) (1-y_{i0}) \right] \prod_{t=2}^{n_i} \left(\frac{e^{y_{it}\eta_{it2}}}{1+e^{\eta_{it2}}} \right) \right\} \times$$

$$\times \prod_{t=2}^{n_i} \left(\frac{1}{1+e^{\xi_{it}}} \right),$$

$$\text{JRI}: \quad \mathsf{f}(\boldsymbol{y}, \boldsymbol{w}|\boldsymbol{\theta}) = \prod_{i=1}^{m} \left[\left(\frac{e^{y_{i1}\eta_{i11}}}{1+e^{\eta_{i11}}} \right) y_{i0} + \left(\frac{e^{y_{i1}\eta_{i10}}}{1+e^{\eta_{i10}}} \right) (1-y_{i0}) \right] \left[\prod_{t=2}^{n_i} f_{yw}(y_{it}, 0) \right]$$

where η_{i1b}, $b = 0, 1$, and η_{it} are given by (3) and (1) and $p = 8$ for AR model; by (11) and (10), and $p = 10$ for RI model; η_{i1kb}, $k = 1, 2$, $b = 0, 1$, and η_{itk} by (8) and (7) and $p = 12$ for MX model; η_{i1b}, $b = 0, 1$ by (11), $f_{yw}(y_{it}, 0)$ by (19), (17) and $p = 11$ for JRI model; and ζ_{it} by (4).

Table 2 reports the parameter estimates for the four models with and without an ID modeling. Results show that there is a gradual improvement in model performance from AR to MX and to RI models and the RI model with an ID modeling and JRI model are the best model according to *BIC* and *DIC*, respectively. In fact, JRI model gives the best model fit according to $\overline{D(\boldsymbol{\theta})}$ without a model complexity penalty as it allows the dependency between outcome and dropout through an explicit odds ratio function in Eq. (15) that includes dose and concordance effects rather than simply a current heroin use covariate in the dropout model as in Eq. (4). We also find that all models adopting the mixture approach of Chan (2000) using Eq. (2) to handle the initial stage problem for all AR type models provide better model fits. For example, the loglikelihood $\ell = \ln \mathsf{f}(\boldsymbol{y}|\boldsymbol{\theta})$ of the observed data $\boldsymbol{y}$ using the AR model with an ID modeling improves from -1041 to -1012 after adopting the mixture approach of Chan (2000).

6.1 Treatment effect

Table 2 shows that parameter estimates $\boldsymbol{\beta}$ in the outcome model are quantitatively similar and consistent in direction across the four model types as well as across models with or without an ID modeling. Moreover, they are similar to those reported in Chan et al. (1998): increases in methadone dosage and duration of treatment are associated with a decrease in heroin use. However, after allowing for the group effects, the dose effect in the MX model exists only marginally among light heroin users but not heavy users. Alfo and Altkin (2000) detect a substantial change in the parameter estimate for the dose effect after modeling the dropout process. However, we do not observe such effect in our models. The strong and positive association between the previous and present outcomes suggests that some patients tend to use heroin continuously (heavy users) while others (light users) do not, supporting the use of MX model to capture different treatments effects across patients with different levels of heroin uses.

6.2 Mixture and random effects

To classify patients, we assign patients to the group 1 of light heroin users if their group-1 indicators $I_{g,i}$ in the MX model are $\hat{I}_{g,i} > 0.5$. Otherwise, they will be classified to group 2. Table 3 reports the summary statistics of those dropout ($n_i < 26$) and nondropout patients cross-classified with light and heavy user groups.

Table 4 Chi-square test statistics for association in 2 × 2 contingency tables.

Groups	Dropout/nondropout ($n_i < 26/n_i = 26$)	Heavy/light (Gp. 2/Gp. 1)	Initial/noninitial ($Y_{i1} = 1/Y_{i1} = 0$)	End/nonend ($Y_{i,n_i} = 1/Y_{i,n_i} = 0$)
Dropout/nondropout	–			
Heavy/light	*12.6*	–		
Initial/non-initial	0.56	*18.3*	–	
End/non-end	*12.2*	*31.9*	1.5	–

Test values significant at 0.1 level are shown in italic.

Table 5 Simulation study for the effect of ID modeling on outcome parameters.

	With ID model				Without ID model			
	β_0	β_d	β_t	β_p	β_0	β_d	β_t	β_p
True	**−1.5**	**−0.5**	**−0.5**	**3.5**	**−1.5**	**−0.5**	**−0.5**	**3.5**
Lower dropout rate and less max. $n_i = 10$: $D_0 = 0.26$, $D_p = 0.48$								
APE	$_*$*−1.576*	$_*$*−0.466*	$_*$*−0.457*	*3.488*	*−1.520*	*−0.507*	*−0.535*	*3.491*
ASE	*0.245*	*0.179*	*0.127*	*0.164*	*0.245*	*0.179*	*0.124*	*0.164*
RMSE	0.292	0.188	0.156	0.153	0.290	0.187	0.159	0.151
Higher dropout rate and less max. $n_i = 10$: $D_0 = 0.46$, $D_p = 0.77$								
APE	†*−1.540	†*−0.492	†*−0.495	$_*$*3.558*	*−1.450*	*−0.570*	*−0.586*	*3.516*
ASE	*0.304*	*0.235*	*0.161*	*0.201*	*0.302*	*0.234*	*0.154*	*0.202*
RMSE	0.299	0.241	*0.156	*0.187	0.296	0.244	0.164	0.204
Lower dropout rate and more max. $n_i = 20$: $D_0 = 0.45$, $D_p = 0.75$								
APE	$_*$*−1.536*	* *−0.509*	* *−0.479*	*3.529*	*−1.526*	*−0.526*	*−0.533*	*3.530*
ASE	*0.193*	*0.142*	*0.084*	*0.140*	*0.194*	*0.142*	*0.083*	*0.139*
RMSE	0.231	0.162	*0.088	0.129	0.231	0.166	0.092	0.131
Higher dropout rate and more max. $n_i = 20$: $D_0 = 0.68$, $D_p = 0.96$								
APE	$_*$*−1.589*	$_*$*−0.441*	*−0.472*	*3.512*	*−1.526*	*−0.500*	*−0.550*	*3.489*
ASE	*0.258*	*0.202*	*0.123*	*0.183*	*0.260*	*0.204*	*0.119*	*0.183*
RMSE	$_*$0.285	$_*$0.228	*0.119	$_*$0.190	0.276	0.209	0.125	0.184

Notes: ASE: average standard error; RMSE: root mean squared error; Parameter estimate in italic is significant.
* Improvement, less bias or RMSE, in model with ID as compared with model without ID.
$_*$ Worsening, more bias or RMSE, in model with ID as compared with model without ID.
† Improvement of ID model, less bias or RMSE, in data with more missings as compared with those with less missings.

Group 1 contains 85 (62.5%) light users returning only 82 positive screens out of 1919 screens (4.3%). They respond to treatment in a dose-dependent fashion with lower heroin use at higher level of methadone dose. Patients in group 2 are heavy users returning 387 positive screens out of 953 screens (40.6%). This group has insignificant dose effect indicating that patients continue to use heroin regardless of the level of methadone dose received. Moreover, heavy users have a higher dropout rate (56.9% versus 25.9%), lower level of methadone dose (59.5 versus 66.7) and stay shorter in treatment (18.7 weeks versus 22.6 weeks). The Chi-squared tests for association in the 2 by 2 contingency tables, as reported in Table 4, show that *heavy use* is significantly associated with *end use* and *dropout*.

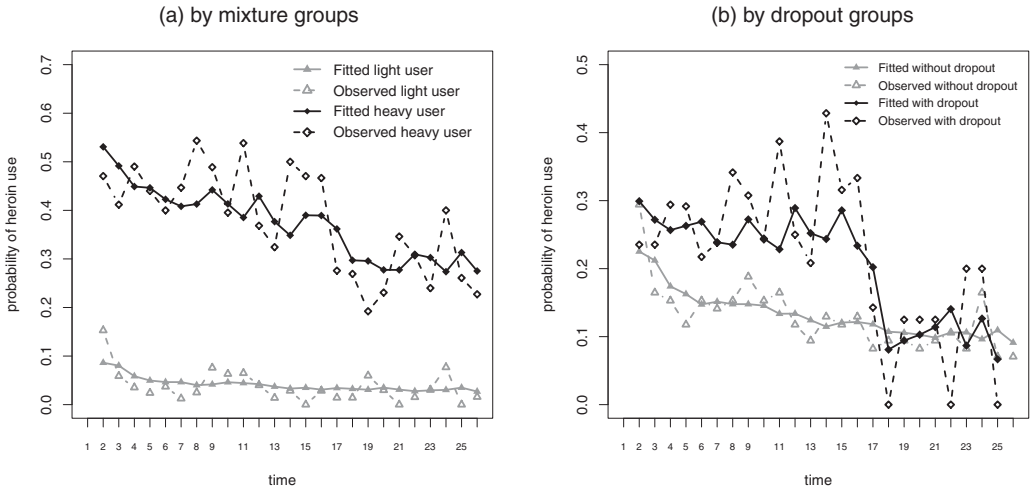

Figure 2 Observed and fitted probabilities of heroin use across time using the JRI model

The significance of the variances σ_b^2 for the random intercepts λ_i in both RI and JRI models suggests the presence of patient heterogeneity. We obtain the same classification of light and heavy-users as the MX model except for four (two) patients if we use the JRI (RI) model and classify patients with $\hat{\lambda}_i \leq 0.57$ ($\hat{\lambda}_i \leq 0.34$) to light-user group. The estimates $\hat{\lambda}_i$ using JRI (RI) model are averaged to -0.94 (-0.71) and 1.63 (1.23) respectively for the light and heavy users groups of the MX model showing heterogeneity of heroin use across the two groups. Based on classification using either $\hat{I}_{g,i}$ in the MX model or $\hat{\lambda}_i$ in the JRI and RI models, health care officers can derive a more patient-oriented MMT program allowing for the differences in responses to treatments (methadone dosage and duration) between these two groups of patients.

6.3 Dropout effect

In all dropout models, the treatment time effect α_t is significant, showing that the probability of dropout increases with duration in treatment. Moreover, the present heroin use effect α_{ps} is also significant in all conditional models, indicating the presence of an ID process in the data and hence the necessity of incorporating an ID model. It also suggests that patients currently taking heroin are more likely to drop out. The low values of intercepts α_0 in the dropout models demonstrate the low probabilities of dropout in general because only 51 (38%) patients drop out of the program before the 26-th week. Furthermore the significances of dose and concordance effects (γ_d and γ_c) in the odds ratio equation using the JRI model suggest that the association between heroin use and dropout deceases with dose and increases with previous association.

Patients who drop out have a higher percentage of heroin use than those nondropout patients (25.8% versus 13.5%) and hence are more likely to be classified as heavy users (56.9% versus 25.9%). Figure 2A displays the good agreement of the observed and fitted probabilities of heroin uses for both heavy and light-user groups based on $\wp_{10,it}$ in the JRI model. It shows a sharp decline in the probability of heroin use among the heavy users and a moderate decline among the light users across time demonstrating a clear treatment effect. Figure 3 displays the gradual increasing trend of dropout and provides another good agreement of the observed and fitted probabilities of dropout. In summary, the JRI model can describe the dynamic of both heroin use and dropout across time well.

Figure 2B shows that dropout patients (black dotted line) have heavier heroin use up to week 14 and thereafter, their level of heroin use drops rapidly to those nondropout patients possibly due to

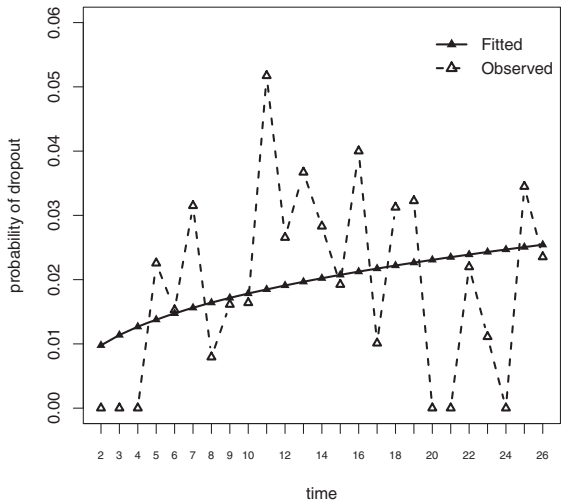

Figure 3 Observed and fitted probabilities of dropout across time using the JRI model.

the dropouts of heavy users. From Table 3, heavy users stay in the program for 18.7 weeks on average. Comparing the AR, MX, and RI models with an ID modeling to those without an ID modeling, the dose effects have consistently increased (by 13, 1, and 5%, respectively) but the time effects have consistently decreased (by 13, 24, and 23%, respectively) though they are still significant. This finding shows that the dose effect is more prominent after accounting for the ID process. Moreover, the gradual dropout of heavy users may lead to a false treatment duration effect of reduced heroin use associated with longer duration in treatment. However, results show that the treatment duration effect is still significant though weaker after allowing for an ID process in the data. This treatment duration effect is also significant in the JRI model. As the treatment effect is confirmed, health care administrators should set up policies to encourage patients to stay longer in the MMT program.

7 Simulation

Sensitivity analyses are often suggested to assess the effect of alternative assumptions about the dropout process on parameter inference (Glynn et al., 1986; Verbeke et al., 2001; Michiels et al., 2002). Hence a simulation study is carried out to evaluate the sensitivity and robustness of parameter estimates in the outcome model to misspecification of the dropout model.

7.1 Procedure

We simulate 200 datasets, each consisting of $I = 300$ heroin users using the AR model with an ID modeling and set the maximum number of outcomes n per patient to be 10 or 20 with two levels of dropout rates in each case. The total number of outcomes without dropouts should be $N = 300 \times 10 = 3000$ and 6000 for $n = 10$ and 20, respectively.

True parameters are set to be $\boldsymbol{\beta} = (-1.5, -0.5, -0.5, 3.5)$ for the outcome model and $\boldsymbol{\alpha} = (-3.0, 0.1, 1)$ or $(-2.0, 0.1, 0.5)$ for the dropout model with moderate and high dropout rates, respectively. Levels for dose are set to be four "1", four "0" and then repeat with four "1" and so on until $n = 10$ or $n = 20$ values are obtained. The values of true parameters are set to achieve a desirable percentage of dropout patients and missing outcomes, (48%, 26%) and (77%, 46%), respectively, for the moderate and high rates when $n = 10$ and (75%, 45%) and (96%, 68%), respectively, when $n = 20$,

while ensuring convergence of parameter estimates. The moderate rates are much lower than the high rates for both n to examine the effects of adopting an ID model on the outcome parameters when the dropout rate varies. It is obvious that the percentages of dropout patients and missing outcomes when $n = 20$ are higher than those when $n = 10$ because patients are more likely to drop out when they stay longer in treatment.

The simulated data with an ID process are then fitted to models with or without an ID modeling. The first 5000 samples of the Gibbs output in the burn-in period are discarded and thereafter with a skip of 50, the posterior sample of size 1000 is taken from each simulated data set for inference.

7.2 Results

Table 4 reports, for each model with and without an ID modeling, the average of parameter estimates (APE) in the outcome model and their standard error estimates (ASE) across simulated datasets, and the root mean squared errors (RMSE) defined as

$$RMSE = \left[\frac{1}{m-1} \sum_{j=1}^{m} (\widehat{\beta}_{kj} - \beta_k)^2 \right]^{\frac{1}{2}},$$

where $\widehat{\beta}_{kj}$ is the parameter estimate for β_k in the j-th simulated dataset and m is the number of simulated datasets.

Results show that all parameter estimates $\boldsymbol{\beta}$ in the outcome model are significant and the ASE are similar across the two types of models, with and without an ID modeling. Moreover, comparing outcome parameters in models with and without an ID modeling, those in models with an ID modeling show less biases and hence higher accuracy when $n = 10$ and the dropout rate is higher. This confirms the necessity to apply ID models to data with higher dropout rates. Furthermore, β_t always shows less bias and lower $RMSE$ except for $n = 10$ and dropout rate is lower. For other parameters, there are no obvious trends of improvement in terms of both biases and $RMSE$. This shows that the treatment duration effect can be estimated more accurately in the model with an ID modeling when an ID process is actually present, supporting the analysis of treatment effect in the methadone clinic data using ID models. Improvement of outcome parameters in the MX and RI models with ID modelings are similar but less in scale. One possible reason is that the ID process in the data may be partially accounted for in the outcome models with mixture or random effects. Results from the MX and RI models are omitted from reporting. Moreover, the JRI model is not considered because the ID model cannot be separated and hence the comparison of outcome parameters from models with and without an ID modeling is not possible.

8 Conclusion

Dropout is common in longitudinal studies. If the dropout process is nonrandom, it may bias certain parameter estimates in the study of treatment effect over time. Through a methadone treatment data, we demonstrate that a separate ID model is necessary. We propose two types of models, the conditional and joint models, to model the ID process. To account for autocorrelation in repeated measurements, an AR term is added to the outcome model and the mixture approach of Chan (2000) is adopted to tackle the initial state problem in AR models. Moreover, the heterogeneity and clustering effects in the data are catered for by incorporating mixture and random effects models of which the random effects are set to be end-use specific to allow for the end-use effect as demonstrated in Fig. 1A. Four proposed models, namely the AR, MX, RI, and JRI models, are estimated by Bayesian approach using MCMC sampling method and implemented using WinBUGS. Sets of full conditional distributions for

all parameters are also derived for the proposed models and are available upon request. According to *DIC*, the JRI model provides the best model performance possibly because it adopts a separate odds ratio function to model the dynamic of the ID process more precisely.

The four proposed models are demonstrated using a methadone clinic data with the aim of evaluating treatment effectiveness. Previous analyses did not allow for an ID process in the data (Chan et al., 1997, 1998). This analysis extends the models to incorporate an ID process and new results indicate the significance of duration time effect on treatment outcome and the more prominence dose effect after allowing for an ID process. Moreover, results also show that longer duration in treatment and present heroin use are associated with higher probability of dropout. Classification from the MX model shows that heavy heroin users have higher dropout rates and the dose effect is only significant among the light users. Hence to improve the treatment effectiveness in a MMT program, one important consideration is to retain patients in treatment. Lastly, a simulation study is conducted to investigate the performance of parameters in the outcome model with or without an ID model when the ID process is actually present and when the length of time series and the rate of dropout vary. Results show improvements in the accuracy of parameters in the outcome model with an ID modeling when the size n is small and the dropout rate is high.

Acknowledgments The author gratefully acknowledge Dr. James Bell of the Drug and Alcohol Unit, Prince of Wales Hospital, Sydney, Australia, for kindly providing the dataset on Methadone Maintenance Treatment Program analyzed in this paper.

Conflict of interest
The authors have declared no conflict of interest.

References

Albert, P. S. and Follmann, D. A. (2003). A random effects transition model for longitudinal binary data with informative missingness. *Statistica Neerlandica* **57**, 100–111.

Albert, P. S. and Follmann, D. A. (2007). Random effects and latent processes approaches for analyzing binary longitudinal data with missingness: a comparison of approaches using opiate clinical trial data. *Statistical Methods in Medical Research* **16**, 417–439.

Alfo, M. and Aitkin, M. (2000). Random coefficient models for binary longitudinal responses with attrition. *Statistics and Computing* **10**, 279–287.

Bradlow, E. T. and Zaslavsky, A. M. (1999). A hierarchical latent variable model for ordinal data from a customer satisfaction survey with 'no answer' responses. *Journal of American Statistical Association* **94**, 43–52.

Chan, J. S. K., Kuk, A. Y. C. and Bell, J. (1997). A Likelihood approach to analysing longitudinal bivariate binary data. *Biometrical Journal* **39**, 409–421.

Chan, J. S. K., Kuk, A. Y. C., Bell, J. and McGilchrist, C. (1998). The analysis of methadone clinic data using marginal and conditional logistic models with mixture or random effects. *Australian and New Zealand Journal of Statistics* **40**, 1–10.

Chan, J. S. K. (2000). Initial stage problem in autoregressive binary regression. *The Statistician* **49**, 495–502.

Chan, J. S. K., Kuk, A. Y. C. and Yam, H. K. (2005). Monte Carlo approximation through Gibbs output in generalized linear mixed models. *Journal of Multivariate Analysis* **94**, 300–312.

Chan, J. S. K., Leung, D. Y. P., Choy, S. T. B. and Wan, W. Y. (2009). Nonignorable dropout models for longitudinal binary data with random effects. *Computational Statistics and Data Analysis* **53**, 4530–4545.

Chan, J. S. K. and Wan, W. Y. (2011). Bayesian approach to analysing longitudinal bivariate binary data with informative dropout. *Computational Statistics* **26**, 121–144.

Chib, S. (1995). Marginal likelihood from the Gibbs Output. *Journal of American Statistical Association* **90**, 1313–1321.

Chib, S. and Jeliazkov, I. (2001). Marginal Likelihood from the Metropolis-Hasting Output. *Journal of American Statistical Association* **96**, 270–281.

Cowles, M. K., Carlin, B. P. and Connett, J. E. (1996). Bayesian tobit modeling of longitudinal ordinal clinical trial compliance data with nonignorable missingness. *Journal of American Statistical Association* **91**, 86–98.
Crouchley, R. and Ganjali, M. (2002). The common structure of several models for non-ignorable dropout. *Statistical Modeling* **2**, 39–62.
Demirtas, H. and Schafer, J. L. (2002). On the performance of random-coefficient pattern-mixture models for non-ignorable drop-out. *Statistics in Medicine* **22**, 2553–2575.
Diggle, P. J. and Kenward, M. G. (1994). Informative drop-out in longitudinal analysis (with discussion). *Applied Statistics* **43**, 49–93.
Dunson, D. B. and Perreault, S. D. (2001). Factor analytic models of clustered multivariate data with informative censoring. *Biometrics* **57**, 302–308.
Fleiss, J. L. (1981). *Statistical Methods for Rates and Proportions* Wiley: New York, NY.
Fitzmaurice, G. M., Molenberghs, G. and Lipsitz, S. M. (1995). Regression models for longitudinal binary responses with informative drop-outs. *Journal of Royal Statistical Society* **B57**, 691–704.
Fitzmaurice, G. M., Lipsitz, S. R., Molenberghs, G. and Ibrahim, J. G. (2001). Bias in estimating association parameters for longitudinal binary responses with drop-outs. *Biometrics* **57**, 15–21.
Gelfand, A. E. and Smith, A. F. M. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* **85**, 398–409.
Geman, S. and Geman, D. (1984). Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images. *IEEE Transactions on Pattern Analysis and Machine Intelligence* **6**, 721–741.
Ghosal, S., Ghosh, J. and Samanta, T. (1995). On convergence of posterior distributions. *The annuals of Statistics* **23**, 2145–2152.
Glynn, J. J., Laird, N. M. and Rubin, D. B. (1986). Selection modeling versus mixture modeling with nonignorable nonresponse. In: Wainer, H. (Ed.), *Drawing Inferences from Self Selected Samples*, Springer, New York, NY, pp. 115–142.
Hastings, W. K. (1970). Monte Carlo sampling methods using Markov Chains and their applications. *Biometrika* **57**, 97–109.
Laird, N. M. (1988). Missing data in longitudinal studies. *Statistics in Medicine* **7**, 305–315.
Lee, K., Marin, J., Mengersen, K. and Robert, C. P. (2008). Bayesian Inference on Mixtures of Distributions, Working paper.
Little, R. J. A. and Rubin, D. B. (1987). *Statistical Analysis with Missing Data*. Wiley, New York, NY.
Little, R. J. A. (1993). Pattern-mixture models for multivariate incomplete data. *Journal of American Statistical Association* **88**, 125–134.
Little, R. J. A. (1994). A class of pattern-mixture models for normal missing data. *Biometrika* **81**, 471–483.
Little, R. J. A. (1995). Modeling the drop-out mechanism in repeated-measures studies. *Journal of American Statistical Association* **90**, 1112–1121.
Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N. and Teller, A. H. (1953). Equations of State calculations by fast computing machines. *Journal of Chemical Physics* **21**, 1087–1091.
Michiels, B., Molenberghs, G., Bijnens, L., Vangeneugden, T. and Thijs, H. (2002). Selection models and pattern-mixture models to analyze longitudinal quality of life data subject to drop-out. *Statistics in Medicine* **21**, 1023–1041.
Park, S., Palta, M., Shao, J. and Shen, L. (2002). Bias adjustment in analyzing longitudinal data with informative missingness. *Statistics in Medicine* **21**, 277–291.
Pauler, D. K., McCoy, S. and Moinpour, C. (2003). Pattern mixture models for longitudinal quality of life studies in advanced stage disease. *Statistics in Medicine* **22**, 795–809.
Rotnitzky, A., Robins, J. M. and Scharfstein, D. O. (1998). Semiparametric regression for repeated outcomes with nonignorable nonresponse. *Journal of American Statistical Association* **93**, 1321–1339.
Preisser, J. S., Lohman, K. K. and Rathouz, P. (2002). Performance of weighted estimating equations for longitudinal binary data with drop-outs missing at random. *Statistics in Medicine* **21**, 3035–3054.
Robins, J. M., Rotnitzky, A. and Zhao, L. P. (1995). Analysis of semiparametric regression models for repeated outcomes in the presence of missing data. *Journal of American Statistical Association* **90**, 106–121.
Rubin, D. B. (1976). Inference and missing data. *Biometrika* **63**, 581–592.
Smith, A. F. M. and Roberts, G. O. (1993). Bayesian computation via the Gibbs sampler and related Markov Chain Monte Carlo methods. *Journal of the Royal Statistical Society, Series B* **55**, 3–23.
Spiegelhalter, D., Best, N. G., Carlin, B. P. and van der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society, Series B* **64**, 583–639.

Spiegelhalter, D., Thomas, A. and Best, N. (2003). Bayesian inference using Gibbs sampling for Window Version (WinBUGS), MRC Biostatistics Unit, Institute of Public health, Cambridge, UK (www.mrc-bsu.cam.ac.uk/bugs/welcome.shtml).

Stubbendich, A. L. and Ibrahim, J. G. (2003). Maximum likelihood methods for nonignorable missing responses and covariates in random effects models. *Biometrics* **59**, 1140–1150.

Ten Have, T. R., Reboussin, B. A., Miller, M. E. and Kunselman, A. (2002). Mixed effects logistic regression models for multiple longitudinal binary functional limitation responses with informative drop-out and confounding by baseline outcomes. *Biometrics* **58**, 137–144.

Troxel, A. B. (1998). Analysis of longitudinal data with non-ignorable non-monotone missing values. *Applied Statistics* **47**, 425–438.

Verbeke, G., Molenberghs, G., Thijs, H., Lesaffre, E. and Kenward, M. G. (2001). Sensitivity analysis for non-random dropout: a local influence approach. *Biometrics* **57**, 7–14.

Verzilli, C. J. and Carpenter, J. R. (2002). A Monte Carlo EM algorithm for random-coefficient-based dropout models. *Journal of Applied Statistics* **29**, 1011–1021.

Wu, M. C. and Carroll, R. J. (1988). Estimation and comparison of changes in the presence of informative right censoring by modeling the censoring process. *Biometrics* **44**, 175–188.

