

A Dirichlet Process Mixture Model for Non-Ignorable Dropout*

Camille M. Moore^{†,‡}, Nichole E. Carlson[†], Samantha MaWhinney[§],
and Sarah Kreidler[¶]

Abstract. Longitudinal cohorts are a valuable resource for studying HIV disease progression; however, dropout is common in these studies. Subjects often fail to return for visits due to disease progression, loss to follow-up, or death. When dropout depends on unobserved outcomes, data are missing not at random, and results from standard longitudinal data analyses can be biased. Several methods have been proposed to adjust for non-ignorable dropout; however, many of these approaches rely on parametric assumptions about the distribution of dropout times and the functional form of the relationship between the outcome and dropout time. More flexible approaches may be needed when the distribution of dropout times does not follow a known distribution or violates proportional hazards assumptions, or when the relationship between the outcome and dropout times does not have a simple polynomial form. We propose a Bayesian semi-parametric Dirichlet process mixture model to flexibly model the relationship between dropout time and the outcome and show that more accurate inference can be obtained by non-parametrically modeling the distribution of subject-specific effects as well as the distribution of dropout times. Results from simulation studies as well as an application to a longitudinal HIV cohort study database illustrate the strengths of our Bayesian semi-parametric approach.

Keywords: Dirichlet process, mixture model, missing data, dropout, MCMC.

1 Introduction

Longitudinal studies are critical to understanding disease progression over time; however, obtaining complete data on all subjects can be challenging. This is particularly true in longitudinal HIV cohorts. It is well documented that many subjects in these studies have missing observations due to death or disease progression, leading to concerns of non-ignorable dropout (Lanoya et al., 2006). In this scenario, standard longitudinal data analyses can produce biased results. For example, this work was motivated by the challenges associated with comparing laboratory markers of HIV disease progression between hard drug users and other subjects in the Women's Interagency HIV Study (WIHS).

*This work was supported by the National Institutes of Health/National Institute on Drug Abuse [DA037778, DA026743, and DA030495].

†National Jewish Health, Center for Genes, Environment, and Health, 1400 Jackson Street, Denver, CO 80206, moorec@njhealth.org

‡Colorado School of Public Health, University of Colorado Denver, Department of Biostatistics and Informatics, 13001 E. 17th Place, B119, Aurora, CO 80045, nichole.carlson@ucdenver.edu

§Colorado School of Public Health, University of Colorado Denver, Department of Biostatistics and Informatics, 13001 E. 17th Place, B119, Aurora, CO 80045, sam.mawhinney@ucdenver.edu

¶Sunrun Inc., 1515 Arapahoe St., Tower II, Suite 600, Denver, CO 80202, sarah5280@gmail.com

Since subjects tend to drop out due to declining health, results from traditional longitudinal analysis methods may be overly optimistic, as they are biased towards study completers, who tend to have more favorable outcomes. In addition, since dropout and missing data are more common among drug users, differences between drug users and other subjects may also be underestimated. As similar dropout related problems have been identified in quality of life data from clinical trials of cancer therapies (Fairclough et al., 1998), anti-depressant clinical trials (Molenberghs et al., 2004), and studies of smoking cessation programs (Hogan et al., 2004b), addressing the analysis challenges in the WIHS will be broadly applicable to other longitudinal clinical trials and cohort studies.

It is well known that in the presence of dropout the marginal distribution of the outcome is not identified and untestable assumptions are necessary to make inferences (Little, 1993; Daniels and Hogan, 2008). While all methods for dealing with missing data must make some unavoidable assumptions about the behavior of unobserved outcomes, most methods rely on additional parametric assumptions. For example, in frailty model frameworks, it is usual to make a parametric or proportional hazards assumption about the distribution of dropout times; in mixture model frameworks, an assumption about the functional form of the relationship between the outcome and the distribution or hazard of dropout is often made (Daniels and Hogan, 2008). Inferences can be sensitive to these assumptions (Molenberghs et al., 1997); therefore, having flexible approaches with minimal assumptions about the patterns or distribution of dropout is useful when considering how missing data may influence the results of longitudinal data analyses. In this paper we present how to use a Dirichlet process mixture of simple frailty models to relax assumptions about the distribution of dropout times and the functional form of the relationship between dropout time and the outcome. We use this model to analyze data from the WIHS, a longitudinal HIV cohort study in which the parametric assumptions of standard models for non-ignorable dropout are violated.

1.1 A Preliminary Investigation of the Women's Interagency HIV Study

The WIHS is an ongoing prospective study of the natural and treated histories of HIV infection in women, with behavioral data and specimens collected at semiannual visits by multiple sites since 1994 (Barkan et al., 1998). In contrast to male populations, HIV and AIDS are more prevalent among women of color exposed through heterosexual partners or intravenous drug use (Bacon et al., 2005; Centers for Disease Control and Prevention, 2002). Illicit drug use has been hypothesized to accelerate HIV disease progression by directly enhancing virus replication and by impairing immune responses. While laboratory *in vitro* and animal studies suggest that drug and alcohol use impairs immune function and increases HIV replication, results from epidemiological studies have been mixed (Moore et al., 2017). These conflicting results may be in part linked to differential dropout between drug users and other subjects. Considering the potential impact of non-ignorable dropout on the results of statistical analyses is particularly important in this context.

The goal of this analysis is to compare declines in CD4⁺ T cell count, a measure of immunologic health and HIV disease progression, between untreated hard drug users

(N=130) and other untreated subjects (N=480) over the first five years of the WIHS study. While WIHS follow-up visits were intended to occur every 6 months, exact timing of visits varies greatly between subjects, so that observation times are not aligned. For example, visit 6 occurred anywhere between 1 and 3 years, with a median of 2.5 years. Therefore we planned to model $\ln(\text{CD}4^+)$ as a linear function of time while controlling for baseline $\ln(\text{CD}4^+)$ and its interaction with time. The primary comparison of interest is the difference in the time effect or slope between hard drug users and other subjects.

In our initial investigation into the data, we found several causes for concern. First, we noted that a large proportion of subjects dropped out of the study early, with half of subjects lost by 2.4 years (median of 4 observations, Figure 1a). In addition, drug users tended to drop out of the study earlier than other subjects. Due to the prevalence and differential distribution of dropout, missing data could have a large impact on the results of our analysis. Second, subjects that dropped out of the study had lower mean $\text{CD}4^+$ at their last visit compared to subjects that remained on study (Figure 2a), and both drug users and others that dropped out early tended to have steeper (more negative) subject-specific ordinary least squares (OLS) slopes than subjects that remained on study. This suggests that subjects that dropped out may have done so due to more rapidly deteriorating health, raising concerns of non-ignorable dropout. We also noted that the distribution of dropout times was irregular (Figure 1b), not taking an obvious parametric form, and that the Kaplan Meier curves of dropout for drug users and others cross, potentially violating parametric and proportional hazards assumptions of common time to event models. Due to the potentially large number of visits for each subject as well as the mis-alignment of visit times, we chose to model correlation using a random effects framework; however, subject-specific OLS slopes do not follow the normal distribution typically assumed for random effects, having long tails off to the left (Figure 2b). These deviations from normality and typical model assumptions led us to search for more robust alternatives.

1.2 Accounting for Missing Data in Longitudinal Studies

Dropout is not ignorable and data are missing not at random when missingness depends on the values of the unobserved outcomes, even after conditioning on the available data (Little and Rubin, 2002). There is a large body of literature describing selection (Heckman, 1979; Diggle et al., 1994; Heckman, 1998), frailty (also called “shared parameter”) (Follmann and Wu, 1995; Albert and Follmann, 2000; Wu and Bailey, 1989; Schluchter, 1992; Lancaster and Intrator, 1998) and mixture models (Rubin, 1977; Wu and Bailey, 1989, 1988) to account for non-ignorable dropout in longitudinal studies with a Gaussian response. These methods are based on different factorizations of the joint distribution of the outcome and dropout time (Daniels and Hogan, 2008). Methods that can accommodate non-normal data are less developed (Ibrahim and Molenberghs, 2009). In a frequentist setting, parametric selection (Ibrahim et al., 2001; Wu and Wu, 2007), frailty (Ten Have et al., 1998), and mixture models (Ekholm and Skinner, 1998; Follmann and Wu, 1995; Fitzmaurice et al., 2001) have been proposed to account for dropout in studies with binary and other non-normal outcomes; however, semi-parametric methods are lacking. In a Bayesian framework, methods for binary outcomes have focused largely on marginalized transition models for population level rather than subject-specific in-

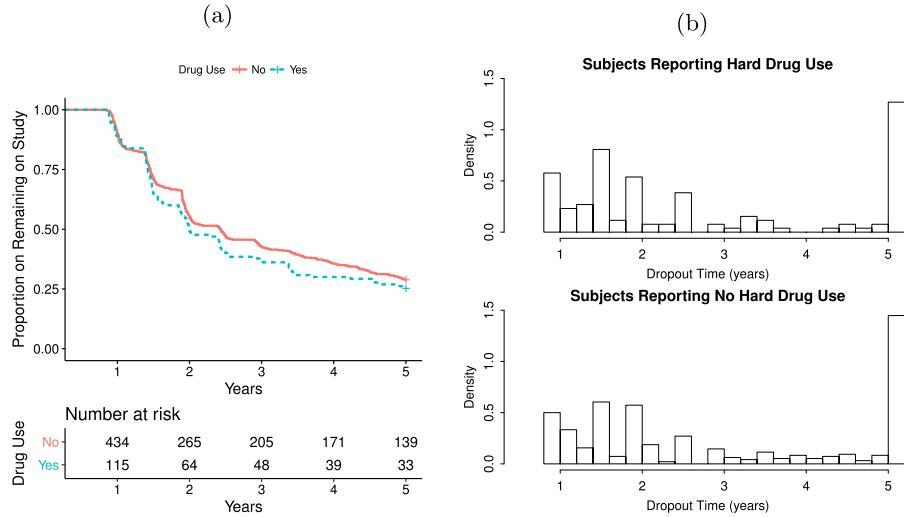


Figure 1: Dropout in the WIHS: (a) Kaplan-Meier Curves for Study Dropout for Subjects Reporting Hard Drug Use (Dashed) and Others (Solid) and (b) Histogram of Dropout Times by Drug Use Group.

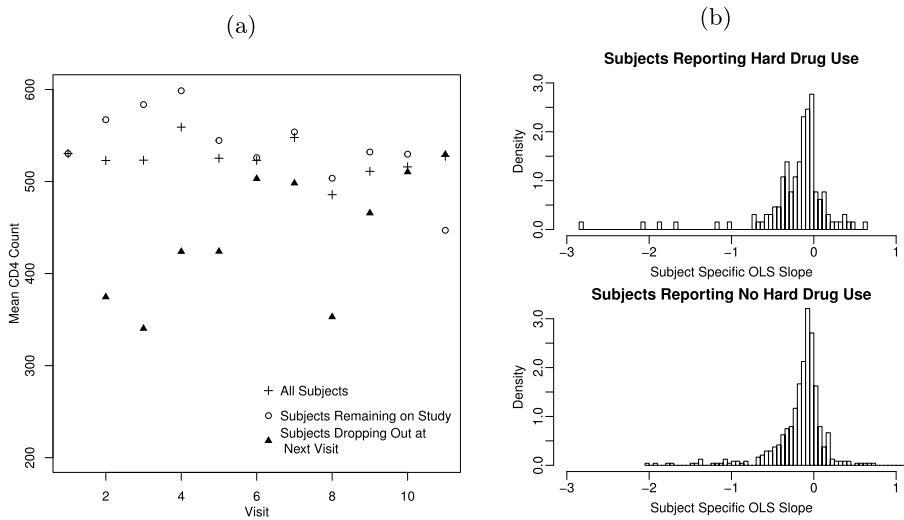


Figure 2: CD4⁺ Over Time in the WIHS. (a) Mean CD4⁺ by Visit for All Subjects (crosses), Subjects Remaining on Study to the Next Visit (open circles) and Subjects Dropping Out at the Next Visit (closed triangles) and (b) Histogram of Subject-Specific Ordinary Least Squares (OLS) Slopes, Representing Change in ln(CD4⁺) per Year by Drug Use Group.

ference (Su and Hogan, 2008, 2010). For subject-specific inference, Kaciroti et al. have described Bayesian pattern mixture models for binary and count data (Kaciroti et al., 2008, 2009, 2012), however, these methods may not be feasible for large numbers of dropout patterns or continuous dropout times.

Related to our model, for studies with a continuous response, aligned observation times and limited dropout patterns, methods proposed by Linero and Daniels (2015) model the joint distribution of dropout time and the longitudinal response as a Dirichlet process mixture of missing at random models in order to relax parametric assumptions about the distribution of the longitudinal response. Our proposed model differs from that of Linero and Daniels (2015) by accounting for correlation in the longitudinal response using a random effects framework and by accommodating continuous dropout times, as well as non-normal outcomes in the exponential family. For a continuous longitudinal response, our model can be represented as a Dirichlet process mixture of the frailty models proposed by Schluchter (1992). By taking a Dirichlet process mixture of frailty models, parametric assumptions about the distributions of the random coefficients and the dropout times are relaxed, by modeling them as mixtures of normal distributions.

1.3 General Approach

Dirichlet process mixture models have been used in semi- and non-parametric Bayesian approaches to model distributions as a mixtures of unknown and potentially infinite numbers of normal distributions (Ghosal, 2010; Christensen et al., 2011). We propose a semi-parametric Dirichlet-process mixture model for dropout in longitudinal studies with exponential family outcomes (DP-Drop) to relax common parametric assumptions made in models that account for non-ignorable dropout.

The DP-Drop can be seen as an extension of the frequentist parametric frailty model proposed by Schluchter et al. (2001). Schluchter considers a two stage model for normally distributed outcomes. The first stage assumes that each subject's responses follow a linear regression with random intercept b_{i0} and slope b_{i1} , which can be written $\mathbf{y}_i = \mathbf{X}_i \mathbf{b}_i + \mathbf{e}_i$, where \mathbf{e}_i is a vector of independent, normally distributed error terms for subject i . In the second stage, the subject-specific random coefficients and the natural log of dropout time, u_i , are modeled with a joint multivariate normal distribution:

$$\begin{pmatrix} b_i \\ \log(u_i) \end{pmatrix} \sim N \left(\boldsymbol{\mu} = \begin{pmatrix} \mu_b \\ \mu_u \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_b & \boldsymbol{\sigma}'_{bu} \\ \boldsymbol{\sigma}_{bu} & \sigma_u^2 \end{pmatrix} \right),$$

where $\boldsymbol{\mu}_b$ is the mean of the random coefficients, μ_u is the mean of the natural log of dropout time, $\boldsymbol{\Sigma}_b$ is the covariance matrix of the random coefficients, $\boldsymbol{\sigma}_{bu}$ is a row vector containing the covariances of u and each random coefficient, and σ_u^2 is the variance of the natural log of the dropout times. This model allows the underlying slope and intercept to be associated with dropout time, via the covariance parameters $\boldsymbol{\sigma}_{bu}$. If these covariances are zero, then the random coefficients and dropout time are independent and dropout is assumed to be non-informative.

In addition to assuming a log-normal distribution of dropout times, the parametric frailty model assumes that there is a linear relationship between the log of dropout time

and the dropout time specific intercepts and slopes, since the random coefficients are related to dropout time through the covariance parameters. We illustrate how to relax these assumptions using a Dirichlet process mixture of frailty models, rather than a single model, to allow for non-parametric modeling of the distribution of dropout times and random effects. There is no need to specify a parametric form for the relationship between dropout time and the subject-specific effects, as in the conditional linear model (CLM) (Wu and Bailey, 1989), which models the effect of time on the longitudinal response as a low-order polynomial function of dropout time. Our DP mixture of frailty models also does not require any a priori grouping of subjects with subjectively similar patterns of missing data, as in a pattern mixture model (PMM) approach (Paule et al., 2003). We see the main benefits of our approach as the robust modeling of the subject-specific coefficients and dropout time distributions, as well as the relationship between dropout time and the subject-specific coefficients, since it is known that inferences can be sensitive to parametric assumptions, particularly in the presence of dropout (Moore et al., 2017; Forster et al., 2012; Hogan et al., 2004a; Molenberghs et al., 1997; Linero and Daniels, 2015). In addition, our model allows the number and timing of responses to vary from subject to subject, as is common in long-term cohort studies, and accommodates longitudinal binary and count responses.

1.4 Outline

In Section 2, we describe the DP-Drop method, how to estimate the model using Markov chain Monte Carlo, and how to calculate marginal effects, averaged over the distribution of dropout times. In Section 3, we present results of a simulation study which show that the DP-Drop method reduces bias and mean squared error for the marginal slope compared to 1) standard generalized linear mixed models (GLMM), 2) simple frailty models for non-ignorable dropout and 3) the conditional linear model (CLM) for non-ignorable dropout when data are missing not at random. In Section 4, we demonstrate the benefits of the DP-Drop method by examining the impact of hard drug use on CD4⁺ T cell decline in untreated HIV infected subjects enrolled in the WIHS.

2 Model Description

2.1 The Dirichlet Process Mixture Model for Dropout

For m subjects with n_i longitudinal observations, let $\mathbf{y}_i = (y_{i1}, \dots, y_{in_i})$ be the vector of outcomes for subject i observed at times $\mathbf{t}_i = (t_{i1}, \dots, t_{in_i})$, for $i = 1, \dots, m$. Let u_i be the dropout time for subject i , which may occur at any continuous point in time. u_i is observed for subjects that drop out and censored for those that complete the study, so that we observe $u_i^o = \min(u_i, C_i)$, where C_i is the amount of time subject i is followed on the study in the absence of early withdraw or dropout. Let $\delta_i = 1$ if u_i is censored and 0 else.

Similar to Linero and Daniels (2015), we model the joint distribution of the outcomes, \mathbf{y} , and dropout times, \mathbf{u} , as a mixture of parametric models, such that:

$$p(\mathbf{y}, \mathbf{u}|\boldsymbol{\omega}, \boldsymbol{\theta}) \sim \int f(\mathbf{y}, \mathbf{u}|\boldsymbol{\omega}, \boldsymbol{\theta}) F(d\boldsymbol{\theta}), \quad (2.1)$$

F is modeled with a Dirichlet process with base measure H and mass $\alpha > 0$ (Escobar and West, 1995). Sethuraman (Sethuraman, 1994) showed that the Dirichlet process mixture is a prior on latent class models, with an unknown, potentially infinite number of classes, so that (1) is equivalent to:

$$p(\mathbf{y}, \mathbf{u} | \boldsymbol{\omega}, \boldsymbol{\theta}) = \sum_{k=1}^{\infty} p_k f(\mathbf{y}, \mathbf{u} | \boldsymbol{\omega}, \boldsymbol{\theta}^{(k)}), \quad (2.2)$$

where $p_k = V_k \prod_{l=1}^{k-1} (1 - V_l)$, $V_k \sim \text{Beta}(1, \alpha)$ and $\boldsymbol{\theta}^{(k)} \sim H(d\theta)$. The specification in (2) is known as the “stick breaking” construction of the Dirichlet process. $\{p_k\}_{k=1}^{\infty}$ are the probabilities of belonging to latent class or cluster k . $\{V_k\}_{k=1}^{\infty}$, known as “stick breaking weights,” are the conditional probabilities of belonging to latent class k given that the subject was not in classes 1 to $k-1$. $\boldsymbol{\theta}^{(k)}$ are the parameters associated with joint distribution of \mathbf{y} and \mathbf{u} in latent class or cluster k , while parameters common across clusters are included in $\boldsymbol{\omega}$. We take $f(\mathbf{y}, \mathbf{u} | \boldsymbol{\omega}, \boldsymbol{\theta}^{(k)})$ to follow a frailty or shared parameter model similar to the model proposed by Schluchter (1992), in which the distributions of \mathbf{y} and \mathbf{u} are linked through subject-specific random coefficients, $\mathbf{b}_1 \dots \mathbf{b}_m$. The main advantage of using the stick breaking formulation is that it allows for straight forward estimation of the posterior with Gibbs sampling.

Model for Latent Class k

Using the stick breaking construction, we can think of each subject as belonging to a latent class, with data from subjects clustered together in class k following a common distribution with parameters $\boldsymbol{\theta}^{(k)}$. To facilitate posterior computation, we use a truncation of the stick breaking model by replacing the infinite sum in (2) with a sum across the first N terms (see Section 2.3 Estimation) (Ishwaran and James, 2001). We introduce latent variables $\mathbf{K} = \{K_1, \dots, K_m\}$ in $\{1 \dots N\}$, which indicate the latent class or cluster assignment of each subject. Let 1_k be an indicator function for class k . We may then write our model:

$$\begin{aligned} \mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}_C, \phi &\sim \text{EF}(\mathbf{X}_i \mathbf{b}_i + \mathbf{Z}_i \boldsymbol{\beta}_C, \phi) \\ K_i | \mathbf{p} &\sim \sum_{k=1}^N p_k 1_k \\ \mathbf{b}_i | \boldsymbol{\mu}_b^{(K_i)}, \boldsymbol{\Sigma}_b &\sim \text{N}(\boldsymbol{\mu}_b^{(K_i)}, \boldsymbol{\Sigma}_b) \\ \log(u_i) | \mathbf{b}_i, \boldsymbol{\mu}_u^{(K_i)}, \sigma_u^2, \boldsymbol{\sigma}_{bu}, \boldsymbol{\mu}_b^{(K_i)}, \boldsymbol{\Sigma}_b &\sim \text{N}(\mu_u^{(K_i)} + \boldsymbol{\sigma}_{bu} \boldsymbol{\Sigma}_b^{-1} (\mathbf{b}_i - \boldsymbol{\mu}_b^{(K_i)}), \sigma_u^2 - \boldsymbol{\sigma}_{bu} \boldsymbol{\Sigma}_b^{-1} \boldsymbol{\sigma}'_{bu}). \end{aligned}$$

For \mathbf{y} , we assume a hierarchical random coefficient model to account for correlation due to repeated measurements made on the subjects over time, where EF is an exponential family distribution with a scale parameter ϕ . The linear predictor for subject i is $\mathbf{X}_i \mathbf{b}_i + \mathbf{Z}_i \boldsymbol{\beta}_C$, where \mathbf{X}_i is an $n_i \times p$ design matrix for the subject-specific coefficients, \mathbf{b}_i , and \mathbf{Z}_i is the design matrix for covariate effects $\boldsymbol{\beta}_C$. For continuous data, a typical choice for EF is the normal distribution with $\phi = \Sigma = I\sigma_{\epsilon}^2$, where I is an $n_i \times n_i$ identity

matrix. For \mathbf{b}_i , we typically include a random intercept, b_{0i} , and time effect, b_{1i} , for each subject.

Within latent class k , subjects share common random coefficient and dropout time distributions, with class specific parameters $\boldsymbol{\theta}^{(k)} = (\mu_u^{(k)}, \boldsymbol{\mu}_b^{(k)})'$. In class k , we assume that the random coefficients have a multivariate normal distribution, with cluster specific mean $\boldsymbol{\mu}_b^{(k)}$ and covariance $\boldsymbol{\Sigma}_b$. Random coefficients account for correlation when the measurement times are misaligned between subjects or there are large number of observation times. This differs from the approach of Linero and Daniels, who modeled correlation by allowing Σ to be an unstructured covariance matrix, which is better suited to studies with a small number of common observation times.

Since dropout times take on positive values, we assume a log normal distribution for the dropout time within latent class k . We note that others working with discrete dropout times or patterns of missingness have chosen to use Bayesian bootstraps or to model the discrete hazard of dropout at time j sequentially with logistic regression (Linero, 2017; Linero and Daniels, 2015). While it has been applied to continuous dropout time distributions (Su and Hogan, 2010), the Bayesian bootstrap assumes that dropout time values that are not observed in the data have zero posterior probability. Rubin (1981) notes that using the Bayesian bootstrap for a variable X effectively makes the assumption that all possible distinct values of X have been observed in the data and that the Bayesian bootstrap is “clearly inappropriate if X can take on many more than [the] n ” observed values, as inferences may be sensitive to this assumption. While the Bayesian bootstrap is well suited to modeling dropout time in studies where dropout occurs at discrete visits, the assumptions of the Bayesian bootstrap are likely not met in studies such as the WIHS where the timing of subjects’ repeated measurements is mis-aligned and dropout may occur at any continuous point in time. Following Schluchter (1992), we allow a subject’s dropout time, u_i , to depend on \mathbf{y}_i through the random coefficients, \mathbf{b}_i , as we believe that subjects with poorer outcomes are more likely to dropout. In our notation, $\mu_u^{(k)}$ is the intercept of the mean function for latent class k and $\boldsymbol{\sigma}_{bu}$ is a row vector containing the covariances of u and each random coefficient.

For simplicity in our model write up, we have not included covariate effects in the distributions of \mathbf{b}_i or u_i ; however, our model does allow $\boldsymbol{\mu}_b^{(k)}$ and $\mu_u^{(k)}$ to depend on categorical covariates. For example, in our application, we allow separate $\boldsymbol{\mu}_b^{(k)}$ and $\mu_u^{(k)}$ for hard drug users and others.

Relationship to Parametric Frailty Models

Although perhaps an unusual presentation, we can write the joint distribution of dropout times and the subject-specific random effects as a multivariate normal distribution to highlight which parameters are included in the Dirichlet process mixture and to clearly show our model as a natural extension of Schluchter et al.:

$$\begin{pmatrix} \mathbf{b}_i \\ \log(u_i) \end{pmatrix} | \boldsymbol{\omega}, \boldsymbol{\theta}^{(k)} \sim N \left(\boldsymbol{\mu}^{(k)} = \begin{pmatrix} \boldsymbol{\mu}_b^{(k)} \\ \mu_u^{(k)} \end{pmatrix}, \boldsymbol{\Sigma} = \begin{pmatrix} \boldsymbol{\Sigma}_b & \boldsymbol{\sigma}_{bu} \\ \boldsymbol{\sigma}_{bu} & \sigma_u^2 \end{pmatrix} \right).$$

Within class k the data are modeled using the parametric frailty model proposed by Schluchter (1992). We note that this joint distribution, $p(u_i, \mathbf{b}_i | \boldsymbol{\mu}^{(K_i)}, \boldsymbol{\Sigma}) = p(u_i | \mathbf{b}_i, \boldsymbol{\mu}^{(K_i)}, \boldsymbol{\Sigma}) p(\mathbf{b}_i | \boldsymbol{\mu}^{(K_i)}, \boldsymbol{\Sigma}_b)$, is part of the posterior distribution as factored in Section 2.2: Priors and Posterior. Our approach non-parametrically models the distribution of dropout times as a mixture of log-normal distributions and the distribution of subject-specific coefficients as a mixture of multivariate normal distributions, while also clustering subjects with similar dropout times and coefficient values.

In order to account for non-ignorable dropout, we must allow for dependence between the outcome model and the dropout model. Clustering is a flexible method to achieve this dependence without requiring that the relationship between the random coefficients and dropout time take any particular functional form. This relaxes the assumption of a linear relationship between dropout time and subject-specific slopes common in both the parametric frailty model and the conditional linear model. Pattern mixture models use a similar strategy to account for non-ignorable dropout, by fitting separate models for a series of dropout patterns. However, unlike PMM's, the DP model does not require a-priori grouping of subjects into subjective patterns or clusters. A similar strategy has been implemented by Linero and Daniels (2015) for discrete dropout times. Linero (2017) also notes that an alternative to his proposed method using the Bayesian bootstrap would be to model the joint distribution of the outcomes and the missing data as a mixture, treating the outcomes and missing data as conditionally independent given class membership, with each class having unique parameters for both the dropout distribution and the distribution of the data. We propose a similar mixture; however, since within each mixture component we assume a normal distribution for the subject specific effects as well as $\log(u)$, it is natural to model these quantities using a multivariate normal distribution, similar to that proposed by Schluchter (1992), as we believe that subjects with poorer outcomes are more likely to dropout. This allows for potential covariance between the subject specific effects and dropout time within cluster. It is possible for this covariance to be estimated at or near 0, in which case the subject specific effects and dropout time would be conditionally independent given cluster membership, similar to the model for discrete dropout times proposed by Linero and Daniels (2015).

2.2 Priors and Posterior

Above we derived the model for $p(\mathbf{y}, \mathbf{u} | \boldsymbol{\omega}, \boldsymbol{\theta})$. However, we are interested in the posterior distribution, $p(\boldsymbol{\omega}, \boldsymbol{\theta} | \mathbf{y}, \mathbf{u}) \propto p(\mathbf{y}, \mathbf{u} | \boldsymbol{\omega}, \boldsymbol{\theta}) p(\boldsymbol{\omega}, \boldsymbol{\theta})$. In our model specification, $p(\mathbf{y}, \mathbf{u} | \boldsymbol{\omega}, \boldsymbol{\theta}) p(\boldsymbol{\omega}, \boldsymbol{\theta})$ factorizes as follows:

$$\begin{aligned} & \left[\prod_{i=1}^m p(\mathbf{y}_i | \mathbf{b}_i, \boldsymbol{\beta}_C, \phi) p(u_i | \mathbf{b}_i, \boldsymbol{\theta}^{(K_i)}, \boldsymbol{\Sigma}_b, \sigma_{bu}^2) p(\mathbf{b}_i | \boldsymbol{\theta}^{(K_i)}, \boldsymbol{\Sigma}_b) \right] \\ & \times \left[\prod_{k=1}^N p(\boldsymbol{\theta}^{(k)} | \boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0) \right] \times p(\mathbf{K} | \mathbf{V}) p(\mathbf{V} | \alpha) p(\boldsymbol{\beta}_C) p(\boldsymbol{\Sigma}) p(\alpha) p(\mu_0) p(\Sigma_0) p(\phi). \end{aligned}$$

As mentioned in Section 2.1, $V_k \sim \text{Beta}(1, \alpha)$ and $\boldsymbol{\theta}^{(k)} = (\mu_u^{(k)}, \boldsymbol{\mu}_b^{(k)})' \sim H$, where α is the concentration parameter and H is the baseline distribution of the Dirichlet

process. In practice, we use a normal baseline distribution, so that the parameters $\boldsymbol{\theta}^{(k)} \sim N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$. In addition, the following priors are assigned:

$$\begin{aligned}\boldsymbol{\beta}_C &\sim N(\mathbf{0}, \mathbf{R}_0) \\ \boldsymbol{\Sigma} | \nu_0, \mathbf{T}_0 &\sim IW(\nu_0, \mathbf{T}_0) \\ \boldsymbol{\mu}_0 | \mathbf{m}_b, \mathbf{S}_b &\sim N(\mathbf{m}_b, \mathbf{S}_b) \\ \boldsymbol{\Sigma}_0 | \nu_b, \mathbf{T}_b &\sim IW(\nu_b, \mathbf{T}_b) \\ \alpha &\sim \text{Gamma}(a_0, b_0),\end{aligned}$$

where IW is the inverse Wishart distribution. For normally distributed outcomes, the within-subject error variance is assigned an inverse gamma prior.

2.3 Estimation

The posterior distribution is not analytically tractable and must be approximated. Markov chain Monte Carlo (MCMC) is a common approach (Ishwaran and James, 2001; MacEachern, 1994; Jain and Neal, 2004), although other alternatives, including partial predictive recursion (Newton and Zhang, 1999), variational Bayes approximations (Blei and Jordan, 2006), sequential importance sampling (MacEachern et al., 1999), and weighted Chinese restaurant sampling (Ishwaran and Takahara, 2002) have been proposed. A blocked Gibbs sampler for non-parametric random effects distributions has been described in detail by Ishwaran and James (Ishwaran and James, 2001). The method approximates the posterior using a truncation of the stick breaking model of (2.2). Since the probability weights assigned to each cluster decrease as the number of clusters increases, the infinite sum can be replaced by a sum across the first N terms by letting $V_N = 1$, and a good approximation can be achieved for fairly modest N of 20 to 50 (Dunson, 2010). For non-Gaussian exponential family outcomes, we utilize a hybrid Gibbs Metropolis-Hastings sampler based on the truncated stick breaking approach (Ishwaran and James, 2001). For Gaussian outcomes, the posterior can be estimated using Gibbs sampling alone. We use data augmentation to simulate the censored dropout times for subjects that complete the study or are administratively censored ($\delta_i = 1$) (Wei and Tanner, 1990). To initialize the algorithm, we set $u_i = u_i^o$ for the first iteration. Details of our sampler are described in Appendix A of the Supplementary Material (Moore et al., 2019). An R package to implement these models is available at <https://github.com/kreidles/informativeDropout>.

2.4 Calculation of Marginal and Dropout Time Specific Effects

Marginal Effects, Averaged Over Clusters

At each iteration of the MCMC algorithm, subjects are assigned to clusters or latent classes, with class specific means for their random coefficients. However, we are typically interested in the expected value of the coefficients, averaged across clusters. The stick breaking construction of the Dirichlet process makes the calculation of these marginal effects straight forward, since the probability of belonging to cluster k , p_k , is calculated

at each iteration s of the MCMC, we calculate the marginal or expected value of the subject-specific coefficients as follows: $E(\mathbf{b}_i^{(s)}) = \sum_{k=1}^N p_k^{(s)} \boldsymbol{\mu}_b^{(k)(s)}$.

Dropout Time Specific Effects

In order to understand the relationship between dropout time and the subject-specific coefficients, we are also interested in the expected values of the random coefficients at specific dropout times, $E(\mathbf{b}_i|u_i = u)$. These can also be estimated at each iteration as $E(\mathbf{b}_i|u)^{(s)} = \sum_{k=1}^N p_{k|u}^{(s)} E(\mathbf{b}_i|\boldsymbol{\theta}^{(k)(s)}, \boldsymbol{\omega}^{(s)}, u)$, where $p_{k|u}^{(s)}$ is the probability of belonging to cluster k given a particular dropout time and $E(\mathbf{b}_i|\boldsymbol{\theta}^{(k)(s)}, \boldsymbol{\omega}^{(s)}, u)$ is the expected value of the random coefficients in cluster k for a subject with dropout time u at iteration s . $p_{k|u}$ can be estimated at iteration s as follows:

$$p_{k|u}^{(s)} = \frac{p_k^{(s)} f(\log(u)|\mu_u^{(k)(s)}, \sigma_u^{2(s)})}{\sum_{k=1}^N p_k^{(s)} f(\log(u)|\mu_u^{(k)(s)}, \sigma_u^{2(s)})},$$

where f is the normal probability density function with mean $\mu_u^{(k)(s)}$ and variance $\sigma_u^{2(s)}$. Using properties of the multivariate normal distribution, $E(\mathbf{b}_i|k, u)$ at iteration s can be calculated:

$$E(\mathbf{b}_i|k, u)^{(s)} = \boldsymbol{\mu}_b^{(k)(s)} + \boldsymbol{\sigma}_{bu}^{(s)} \sigma_u^{-2(s)} (\log(u) - \mu_u^{(k)(s)}).$$

2.5 Specification of Hyperparameters and Sensitivity to Priors

Dunson (2010) has discussed implementation issues when using Dirichlet process priors to non-parametrically model the random effects distribution in hierarchical models. The posterior distribution of the number of clusters is influenced by both α and the Dirichlet process baseline distribution. The data inform strongly about α , so robust results can be obtained by choosing a hyperprior for α ; a common choice is Gamma(1,1). However, the posterior distribution is sensitive to the choice of baseline distribution and a poor choice can lead to a model with a single cluster or very few clusters. For example, for the $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ baseline distribution, choosing the diagonal variance components of $\boldsymbol{\Sigma}_0$ to be extremely large in order to be “non-informative” may result in all of the subjects being allocated to the same cluster, since the conditional posterior probability of allocating subjects to unoccupied clusters is very low. This would result in the parametric frailty model. Interestingly, allocating all subjects to separate clusters also results in a parametric model since the random effects distribution is then the $N(\boldsymbol{\mu}_0, \boldsymbol{\Sigma}_0)$ baseline distribution. To properly characterize a non-normal distribution of subject-specific effects, subjects should be assigned to a few distinct clusters (Dunson, 2010). This makes it important to choose the baseline distribution with careful thought about realistic values given the problem.

One solution that we have employed in this paper is to use a hyperpriors for the baseline distribution parameters, although the hyperparameters, \mathbf{m}_b , \mathbf{S}_b , ν_b , and \mathbf{T}_b , must still be chosen. For HIV cohort studies, there is a wealth of information in the

literature that can help inform these choices. For example, in our application, we considered the range of typical CD4⁺ T cell counts observed in untreated HIV⁺ subjects and the mean and range of dropout times in our dataset to help inform our choices for \mathbf{m}_b and \mathbf{S}_b , the hyperprior parameters for the mean of the normal baseline distribution. Since we planned to include baseline log(CD4⁺ T cell count) as a covariate in our models, setting the first element of \mathbf{m}_b , which relates to the subject specific intercepts, to 0 was a reasonable choice. Since declines in log(CD4⁺ T cell count) are typically small (less than 1 log per year) and our primary interest was in comparing changes in CD4⁺ T cell count over time between groups, we set the second element of \mathbf{m}_b to 0 to be non-informative. For the third element of \mathbf{m}_b , we used the log(mean dropout time) in our dataset. \mathbf{S}_b quantifies our prior confidence in our estimates of \mathbf{m}_b , the mean of the normal hyperprior on the mean of the baseline distribution. We set \mathbf{S}_b to a 3 x 3 identity matrix, as this covered the range of subject specific intercepts, slopes and dropout times we believed were plausible.

Specifying ν_b and \mathbf{T}_b , the hyperparameters for the inverse Wishart hyperprior on the covariance of the normal baseline distribution, can be more challenging. We chose to use a 3 x 3 identity matrix for \mathbf{T}_b and then set ν_b so that the expected variances of the inverse Wishart distribution would cover the plausible range of intercepts, slopes, and dropout times in our dataset. For example, the expected value of the inverse Wishart distribution is $\mathbf{T}_b/(\nu_b-4)$, if a subject specific intercept and slope are included in the cluster specific coefficients. Setting ν_b to 5, resulted in prior expected variances of 1 for the baseline distribution, which covered the range of values expected or seen in our dataset well. For example, log(dropout time) ranged from approximately -0.6 to 1.6 log(years), so a single standard deviation above and below the expected mean of the baseline distribution would cover all the of the observed values in our dataset. Since the log of dropout times, intercepts and changes in log(CD4)⁺ over time were on a similar scale, this strategy worked well. If this were not the case, a more complex \mathbf{T}_b , with unequal diagonal elements, may have been more appropriate. Shi et al. (2019) have proposed methods for developing low information priors for Dirichlet process mixture models, which involve scaling the data before analysis. This offers another potential solution when there is less subject-specific information available to choose hyperparameter values. We also suggest fitting additional models varying the hyperparameter values for the hyperpriors for α and the mean and variance of the baseline distribution to better understand sensitivity of inferences to these choices.

Other hyperparameters that must be specified include R_0 , the prior variance of covariate effects, and ν_0 and T_0 , the prior degrees of freedom and scale matrix for the inverse Wishart prior on Σ , the within cluster the covariance matrix for the subject specific intercept, slope and log of dropout time. For R_0 we typically use a covariance matrix with zeros for the off-diagonal elements and large values on the diagonal, in order to be uninformative. In our simulation and applications, we set the diagonal elements to 100, but in general “large” will depend on the scale of the data. Again, we set ν_0 and T_0 to 5 and a 3 x 3 identity matrix, respectively, as this would cover the range of means for subject specific intercepts, slopes and dropout times we believed were plausible in our dataset.

2.6 Identifiability of Model Parameters

The identifiability of parameters is an issue both in statistical models for missing data and in Bayesian non-parametric modelling in general. Likelihood invariance under label switching is a well-known problem in Bayesian mixture modeling, which complicates inference on the parameters of the individual components of mixture models (Stephens, 2000; Celeux et al., 2000; Makov et al., 1985). However, in our proposed approach, we do not attempt to assign meaning to or make inference on individual components; rather, we simply use the mixture model approach to flexibly model the marginal distribution of dropout times and subject-specific effects. As noted in Linero (2017), in the presence of missing data, identification of model parameters and causal effects requires assumptions about the missing data which cannot be checked, making sensitivity analysis important. In our model, we assume that after accounting for dropout time, data are missing at random and that subject specific slopes remain unchanged after dropout (i.e. subjects continue on the same linear trajectory after dropout). In practice, we perform sensitivity analyses to determine if results are robust to this assumption by re-fitting the model assuming that slopes change by a multiplicative factor of Δ after dropout. We consider a range of Δ and examine results to understand the extent to which Δ changes estimates of the expected value of the outcome over time. We perform a sensitivity analysis in the analysis of the WIHS data in Section 4.2.

3 Simulation

We assess the performance of our model to estimate the marginal slope (expected change in the outcome over time) as well as dropout time specific slopes. Data were generated under four different scenarios, including two normally distributed outcomes and two binary outcomes. In these four scenarios the subject-specific slopes were related to dropout time by two different dropout mechanisms: (i) a continuous and smooth function of dropout time, resulting in a left skewed distribution of subject-specific slopes and (ii) a discontinuous step function, resulting in a bimodal distribution. In addition, simulations for a linear relationship and no dropout effect are presented in Appendix B of the Supplementary Material to show the method can also fit CLMs and GLMMs.

More specifically, the following form for generating simulated data was assumed: $\eta_{ij} = \beta_0 + \beta_1(u_i)t_{ij} + \gamma_{0i} + \gamma_{1i}t_{ij}$, $i = 1, \dots, m$, $j = 1, \dots, n_i$ for m subjects with n_i observations for the i th subject, where $(\gamma_{0i}, \gamma_{1i})' \sim N(0, \Sigma_\gamma)$. For the Gaussian simulations, $y_{ij} \sim N(\eta_{ij}, \sigma^2)$, and $\beta_0 = 0$. Dropout times were $u = U/15 \in [0, 1]$, resulting in 16 time points spaced equally from 0 to 1. Dropout was created from a beta-binomial where $p \sim \text{Beta}(1.5, 1.5)$ and $U \sim \text{Bin}(15, p)$. The within-subject variance, σ^2 , was set at 0.067. The elements of Σ_γ were as follows: $\sigma_0^2 = 0.4$, $\sigma_1^2 = 0.01$ and $\sigma_{01} = -0.01$. For the binary simulations, $y_{ij} \sim \text{Bernoulli}(\text{logit}^{-1}(\eta_{ij}))$, $\beta_0 = -3$, and for stability, dropout began at the third observation. The elements of Σ_γ were as follows: $\sigma_0^2 = 0.4$, $\sigma_1^2 = 0.1$ and $\sigma_{01} = -0.01$. The forms of the dropout-varying slope were: Normal (i): $\beta_1(u) = -3 \exp(-4u)$, Normal (ii): $\beta_1(u) = I_{(u>2/3)}$, Binary (i): $\beta_1(u) = 10[1 - 2 \exp(-4u)]$, Binary (ii): $\beta_1(u) = 4 + 6I_{(u>2/3)}$, where $I_{(u>2/3)}$ is an indicator function that equals 1 if $u > 2/3$. The forms of the dropout-varying slopes are

represented by the solid lines in Figure 3. For each simulation scenario, 1,000 datasets with 400 subjects each were created. The magnitude of the dropout effects in these scenarios are similar to those seen in HIV cohort studies (Forster et al., 2013; Moore, 2013).

3.1 Methods of Evaluation

DP-Drop models, frailty models (Schluchter et al., 2001), and CLM's (Wu and Bailey, 1989) to account for drop out, as well as standard generalized linear mixed models (GLMM's) were fit to each dataset. The DP-Drop was implemented as described in Section 2.1, including subject-specific intercepts and time effects for \mathbf{b}_i . Similarly, the frailty model was implemented as shown in Section 1.2, also including subject-specific intercepts and time effects for \mathbf{b}_i . The conditional linear model was implemented as follows:

$$\begin{aligned} \mathbf{y}_i &\sim EF(\boldsymbol{\eta}_i, \phi) \\ \boldsymbol{\mu}_i &= g(\boldsymbol{\eta}_i) \\ \boldsymbol{\eta}_i &= \mathbf{1}_i(\beta_0 + b_{0i}) + \mathbf{t}_i(\beta_1 + \beta_2 u_i + b_{1i}) \\ \begin{pmatrix} b_{0i} \\ b_{1i} \end{pmatrix} &\sim N\left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \boldsymbol{\Sigma}_b\right), \end{aligned}$$

where $\mathbf{1}_i$ is an $n_i \times 1$ vector of 1's and b_{0i} and b_{1i} are subject-specific random effects centered at 0. Marginal effects for the CLM were calculated by averaging over the empirical distribution of dropout times. The GLMM was similar to the CLM, with the exception that the $\beta_2 u_i$ term was excluded from the linear predictor. The performance of the methods was compared graphically and in terms of bias, variance, and mean square error for the marginal slope, as well as for the sixteen dropout time specific slopes.

3.2 Implementation

Non-informative priors, with large variances were used for all models. The priors for $\boldsymbol{\Sigma}$ and $\boldsymbol{\Sigma}_0$ were $IW(5, I_{3 \times 3})$, where $I_{3 \times 3}$ is a 3×3 identity matrix, and the prior for the mean of the baseline distribution was $N((0, 0, \log(0.5))', I_{3 \times 3})$. The hyperparameters for α were both set to 1, and an inverse gamma prior with both hyperparameters set to 0.001 was used for the within-subject variance for the normal distribution simulations. A truncation level of $K = 20$ was used and subjects were randomly assigned to one of the clusters to initialize the MCMC chain. The MCMC chains were run for 20,000 iterations, with the first 5,000 iterations used as burn in. Trace plots were used to check for convergence. Similar MCMC algorithms were run for the frailty model, the CLM, and the standard generalized linear mixed model. The DP-Drop, CLM, frailty model and GLMM methods were implemented in R using a custom MCMC algorithm utilizing the MASS (Venables and Ripley, 2002), mvtnorm (Genz et al., 2014), MCMCpack (Martin et al., 2011), and gtools (Warnes et al., 2014) packages.

Method	Normal (i)					Normal (ii)				
	Slope	Bias	Var.	MSE	Coverage	Slope	Bias	Var.	MSE	Coverage
DP-Drop	-0.56	0.14 (20%)	0.006	0.03	0.62	0.33	-0.02 (-6%)	0.006	0.006	0.96
CLM	-0.45	0.25 (36%)	0.003	0.06	0.20	0.30	-0.05 (-14%)	0.028	0.03	0.23
Frailty	-0.52	0.19 (27%)	0.004	0.04	0.18	0.16	-0.19 (-54%)	0.005	0.04	0.22
GLMM	-0.21	0.49 (70%)	0.001	0.24	0	0.65	0.31(89%)	0.002	0.10	0

Method	Binary (i)					Binary (ii)				
	Slope	Bias	Var.	MSE	Coverage	Slope	Bias	Var.	MSE	Coverage
DP-Drop	7.66	0.30 (4%)	0.21	0.30	0.69	6.36	-0.36 (-5%)	0.27	0.39	0.70
CLM	8.02	0.66 (9%)	0.11	0.54	0.50	6.47	-0.24 (-4%)	0.14	0.20	0.90
Frailty	8.14	0.78 (11%)	0.14	0.75	0.36	6.04	-0.96 (-14%)	0.18	1.11	0.60
GLMM	9.09	1.74 (24%)	0.09	3.11	0	8.56	1.85 (28%)	0.14	3.55	0

Table 1: Simulation Study Comparing the DP-Drop, Parametric Frailty Model, CLM and GLMM Methods: Posterior Mean Marginal Slope Estimates, Bias (Absolute, (Relative)), Variance (Var.), Mean Squared Error (MSE), and 95% Credible Interval Coverage. Negative bias values indicate the slope was underestimated, while positive values indicate overestimation.

3.3 Results

Model performance was quantified in terms of bias, variance, mean squared error (MSE) and 95% credible interval coverage for the marginal slope (Table 1). The median number of non-empty clusters for the DP-Drop model was 10.5, 12.0, 6.5, and 8.1 for the Normal (i), Normal (ii), Binary (i), and Binary (ii) scenarios, respectively. The GLMM had the lowest variance, likely because the method makes unmet assumptions that simplify the model and also has the fewest parameters. The DP-Drop method had the lowest bias and MSE, as well as the highest coverage probability, of all the methods, except in the Binary (ii) scenario, where the CLM had better performance. However, in this scenario, the CLM poorly modeled the dropout time specific slopes, underestimating slopes at early dropout times and over-estimating at later dropout times. This combination of over and under estimation, rather than appropriate model fit, resulted in the minimized bias values.

While the DP-Drop had the best performance in terms of credible interval coverage, we note that the DP-Drop 95% credible intervals do not always maintain 95% frequentist coverage probabilities. While in certain scenarios, Bayesian credible intervals can achieve nominal frequentist coverage probabilities, in general, there is no guarantee that a 95% Bayesian credible interval will maintain 95% frequentist coverage probability, as coverage probability depends on the prior distribution and sample size, as well as other factors (Wasserman, 2011; Gray et al., 2015). In these simulations with non-ignorable dropout, coverage is lower than expected since the estimates of change over time for subjects that dropout early, and therefore have fewer observations, are necessarily shrunk towards subjects with more information. For example, for the Normal (i) scenario, subjects that dropout after their first visit have the smallest change over time, but only have one observation, which is not enough information to estimate a slope. Therefore, it is not surprising that the DP-Drop method estimates the marginal slope to be higher than the true simulation value and that the coverage is less than 95% due to this shrinkage.

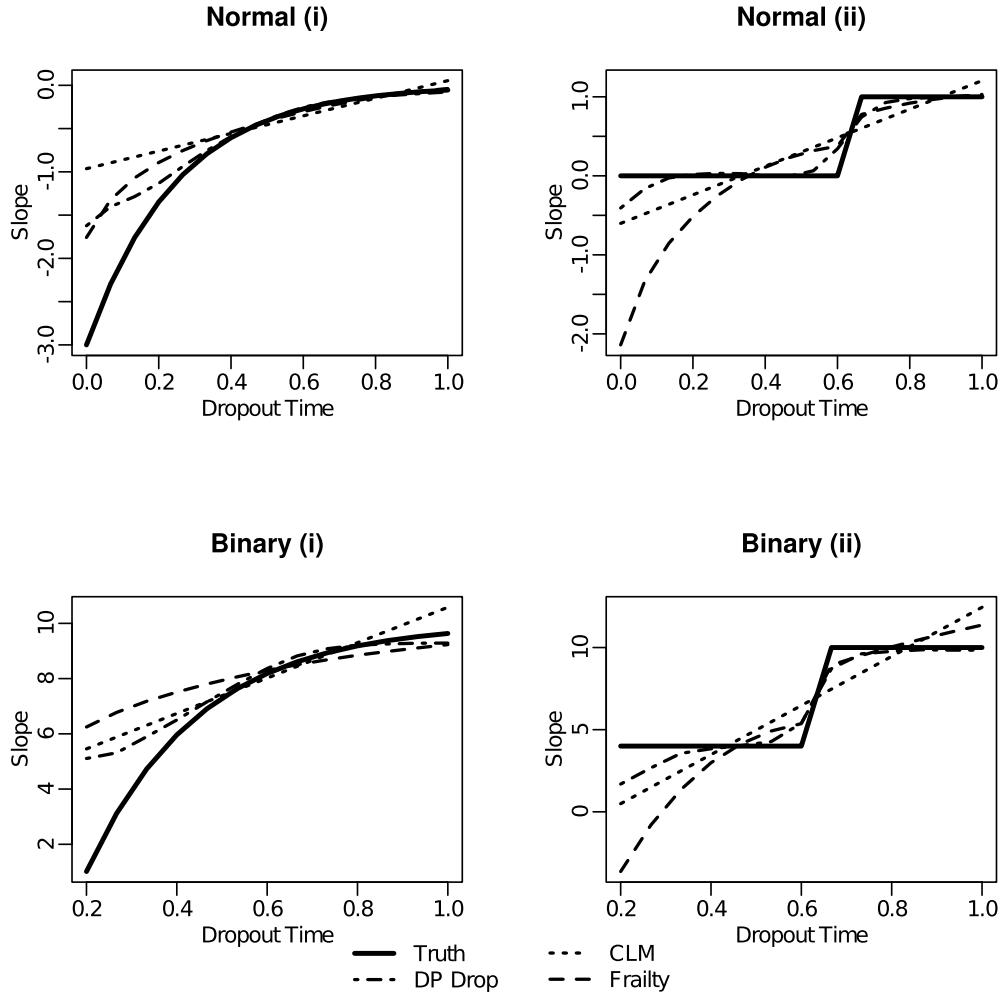


Figure 3: Simulation Study Comparing the DP-Drop, CLM and Frailty Model Methods to Account for Dropout: Posterior Mean Estimates of the Dropout-Varying Slope, Averaged Over 1,000 Simulations. In all cases the DP-Drop method fits the simulated dropout-varying slope more closely than the frailty model.

Graphs of the predicted DP-Drop frailty model, and CLM slopes at each dropout time are presented in Figure 3. The DP-Drop method was able to more accurately describe the relationship between dropout time and the subject-specific slope compared to the frailty model and the CLM, which always fits a linear relationship. While the frailty model and DP-Drop seemed to provide a similar fit for the Normal (i) scenario, where the dropout-varying slope could be relatively well represented by the logarithmic curve of the frailty model, the DP-Drop method provided a much better fit for form (ii) and in the binary simulations. The densities of the subject-specific slopes are presented

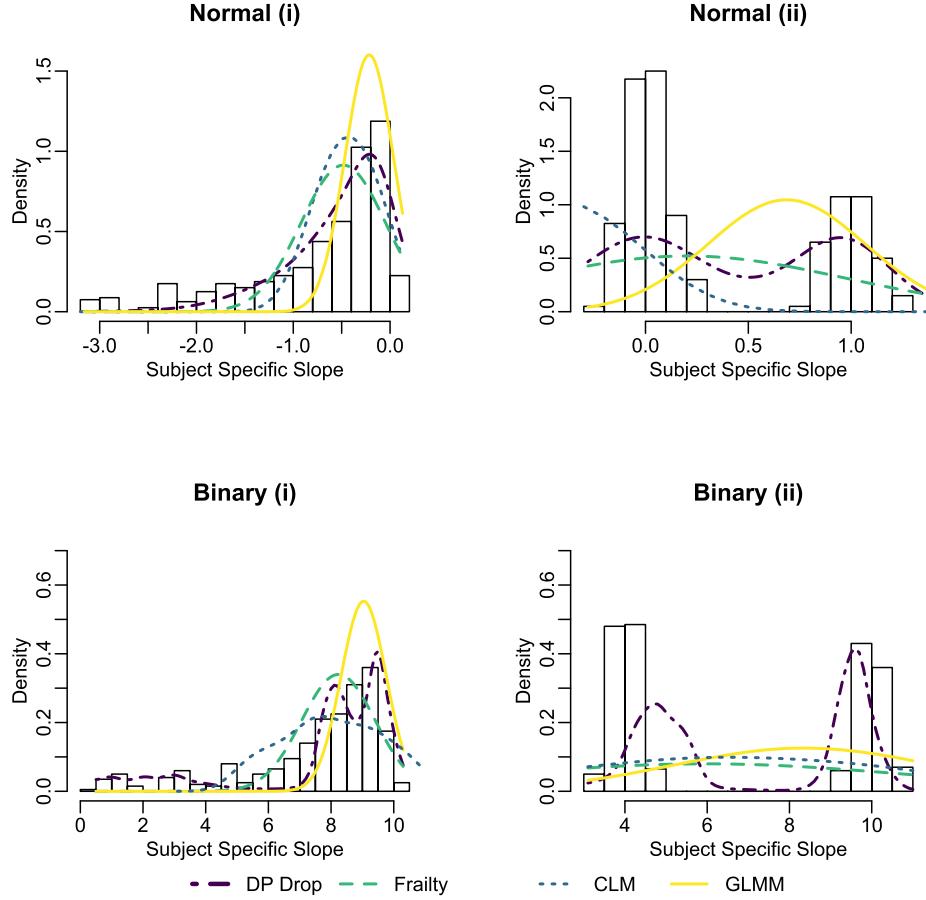


Figure 4: Simulation Study Comparing the DP-Drop, CLM, Parametric Frailty Model, and GLMM Methods: Estimated Densities and Actual Distribution of the Subject-Specific Slope for a Single Simulated Dataset. The DP-Drop is able to model the left skew in the distribution of subject specific effects in scenario (i) and the bi-modal nature of the distribution in scenario (ii).

in Figure 4. The DP-Drop method was able to capture the skewness of the form (i) distributions and the bimodal nature of the form (ii) distributions. The CLM, frailty and GLMM models fit normal distributions to these data, resulting in shrinkage of the estimates of subjects with early dropout times towards the mean for form (i) and large variances in form (ii). For the DP-Drop form (ii), in particular for the binary case, we see that while the density is bimodal, the modes of the estimated density are shifted towards one another compared to the histogram of the simulated data, and are slightly shrunk towards the overall mean. This is not surprising, as there is uncertainty in cluster assignment, particularly for subjects that dropout early and have few observations.

In Appendix B, we present additional simulation study results, including a scenario when dropout is ignorable and there is no relationship between dropout time and the subject specific slope. If there is no relationship between the subject specific slopes and dropout time, the DP-Drop is more complex than needed (as are other models that account for dropout) and the variance and MSE are increased compared to GLMM's.

4 Application

We use our methodology to analyze data from the WIHS. Recall that we are primarily interested in comparing the expected change in $\ln(\text{CD}4+)$ over time between hard drug users and other subjects in the cohort, while controlling for baseline $\ln(\text{CD}4^+)$ and its interaction with time. At all study visits, subjects in the WIHS were questioned regarding illicit drug use since their previous visit. Subjects that reported the use of any hard drug, such as cocaine, amphetamines, heroin or other injection drugs, at more than half of visits were classified as consistent hard drug users and comprised the hard drug use group for this analysis. We define dropout time as the time of the last observation plus 1 day.

4.1 Comparison to Alternatives and Assessing Model Fit

In addition to the DP-Drop, we consider two parametric models to account for non-ignorable dropout, the CLM and the parametric frailty model. We also fit a standard linear mixed model (LMM) that did not account for dropout for comparison. Baseline $\ln(\text{CD}4^+)$ and its interaction with time were included as common covariates in all models. All models also included random subject-specific intercepts and time effects. The CLM and the LMM included separate intercepts and time effects for hard drug users and others. The CLM also included a time by dropout time by hard drug use interaction to allow a separate effect of dropout time for hard drug users and other subjects. For the frailty model the means of the subject-specific coefficients and dropout times were allowed to differ for drug users and others. For the DP-Drop, distributions of the subject-specific coefficients and dropout times were allowed to differ for drug users and others. For full details of these models, see Appendix C of the Supplementary Material.

For the DP-Drop, when possible, we used uninformative priors with large variances. An inverse gamma prior with both hyperparameters set to 0.001 was used for the within-subject variance. The prior for the covariates was $N(0, 100)$. The hyperparameters for α were both set to 1. Priors for Σ and Σ_0 were both $IW(5, I_{3 \times 3})$. The prior for the mean of the baseline distribution was $N((0, 0, \log(1.9))', I_{3 \times 3})$. These values were chosen using the approaches outlined in Section 2.5. A truncation level of $N = 60$ was used. The MCMC chain was run for 200,000 iterations with a burn in of 50,000. We fit additional models varying the priors for α and the mean and variance of the baseline distribution. As our dataset is fairly large, we did not see significant changes in the results. Similar MCMC algorithms were run for the frailty model, the conditional linear model, and the standard generalized linear mixed model.

	CV Loss	Others			Hard Drug Users			Difference		
		PM	95% CI	PM	95% CI	PM	95% CI	PP		
DP-Drop	17.6	-0.16	(-0.21, -0.11)	-0.28	(-0.39, -0.19)	-0.12	(-0.24, -0.01)	0.98		
Frailty Model	18.3	-0.19	(-0.22, -0.15)	-0.27	(-0.33, -0.22)	-0.08	(-0.15, -0.02)	0.99		
CLM	18.0	-0.20	(-0.22, -0.17)	-0.27	(-0.32, -0.22)	-0.07	(-0.13, -0.02)	0.99		
LMM	18.5	-0.16	(-0.18, -0.13)	-0.24	(-0.29, -0.18)	-0.08	(-0.14, -0.02)	0.99		

Table 2: Estimated Change in $\ln(\text{CD}4^+)$ per year for Untreated Subjects in the WIHS Using the DP-Drop, CLM, Frailty, and LMM Methods, Assuming a Baseline $\text{CD}4^+$ count of 478. PM=posterior mean, CI = Credible Interval, PP=posterior probability of a difference < 0 /lower slopes among drug users, CV = cross-validation.

To compare model fit between the methods, we performed a 10-fold cross validation. We first created 10 independent datasets consisting of all observations on 61 subjects (13 subjects reporting hard drug use, 48 others). In each fold of the cross-validation, models were trained on 90% of the data, and predicted values were generated for the remaining 10% of the data in the test set. We calculated a squared error loss function as follows: we calculated the mean squared error of the predicted values for each subject in the test set and summed these values across subjects. Values of the loss function were then averaged over the 10 folds. Lower values of the loss function indicate better model performance. The DP-Drop model had significantly lower loss than any of the alternatives (Table 2), likely because the parametric assumptions of the CLM and frailty models were not met. For example, we see in Panel a of Figure 5 that the CLM assumes a linear relationship between dropout time and the slopes, the frailty model assumes a log-linear relationship, and the DP-Drop model is more flexible. However, this increased flexibility comes at the price of wider credible intervals. In Figure 5b, we see that the histograms of the subject specific slopes estimated using the DP-Drop method reasonably approximates the distributions seen in Figure 1b.

4.2 Results and Sensitivity Analysis

For subjects that reported hard drug use, $\text{CD}4^+$ counts declined by an average of 24.4% per year compared to 15.2% per year for other subjects (Table 2). The DP-Drop method had a posterior mean of 7 clusters (95% credible interval: 4-11) for subjects reporting hard drug use and 13 clusters (95% credible interval: 10-17) for others. Using a linear mixed model that does not account for dropout, these declines are estimated to be 21.0% and 14.4% respectively. Accounting for dropout, we estimate larger declines in $\text{CD}4^+$ count over time, particularly for hard drug users, who were more likely to drop out. In addition, we find larger differences in declines between hard drug users and other subjects (Table 2). Failing to account for dropout could result in underestimating the effect of hard drug use (Figure 6). In Table 3, we see that subjects that reported hard drug use had lower posterior mean slopes across the range of the dropout times, and significantly steeper declines for subjects that dropped out at years 2 and 3. This indicates that the significant marginal difference in $\text{CD}4^+$ decline (averaged over dropout time distribution) between subjects reporting hard drug use and others is not only driven by differences in dropout time distribution, but also by more rapid declines in $\text{CD}4^+$ at specific dropout times.

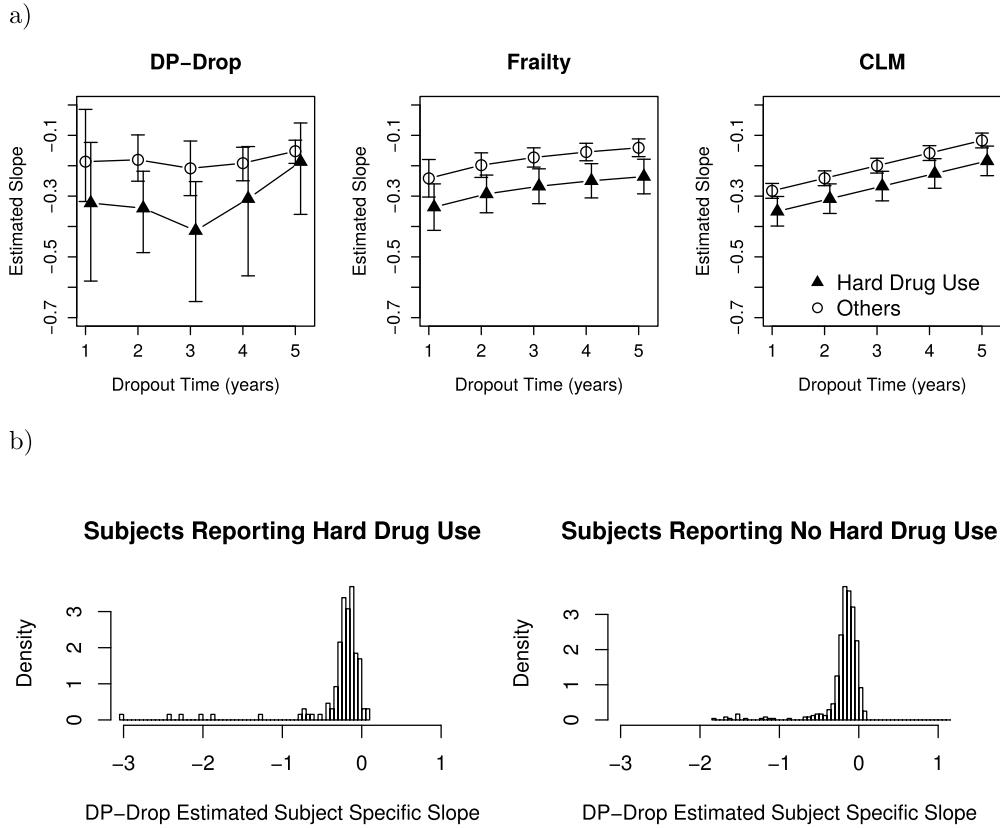


Figure 5: a) DP-Drop, CLM and Frailty Model Posterior Mean and 95% Credible Interval for Change in CD4⁺ Count per Year by Dropout Time for Untreated Subjects in the WIHS, Assuming a Baseline CD4⁺ Count of 478. b) Histogram of the DP-Drop Posterior Mean Estimates of the Subject Specific Slopes.

Our estimates of CD4⁺ count over time rely on the untestable assumption that subjects continue to have the same rate of CD4⁺ decline after they dropout of the study. We evaluate the sensitivity of our results to this assumption by estimating CD4⁺ count over time assuming that the rate of decline changes by a factor of Δ after dropout. $\Delta > 1$ would indicate more rapid declines after dropout, and $0 < \Delta < 1$ would indicate an attenuation of the rate of decline after dropout. $\Delta = 0$ represents the assumption that subjects experience no further declines in CD4⁺ count after dropout, which is unrealistic in a cohort of untreated subjects. For Δ of 0, 0.5, and 0.75, differences in CD4⁺ count between hard drug users and other subjects are estimated to be smaller, but are still statistically significant at years 2, 3, 4, and 5, with drug users having significantly lower CD4⁺ than other subjects (Figure 7, Table 4). This sensitivity analysis suggests that our results are relatively robust to violations of the assumption that subjects' CD4⁺ counts continue to decline at the same rate after dropout.

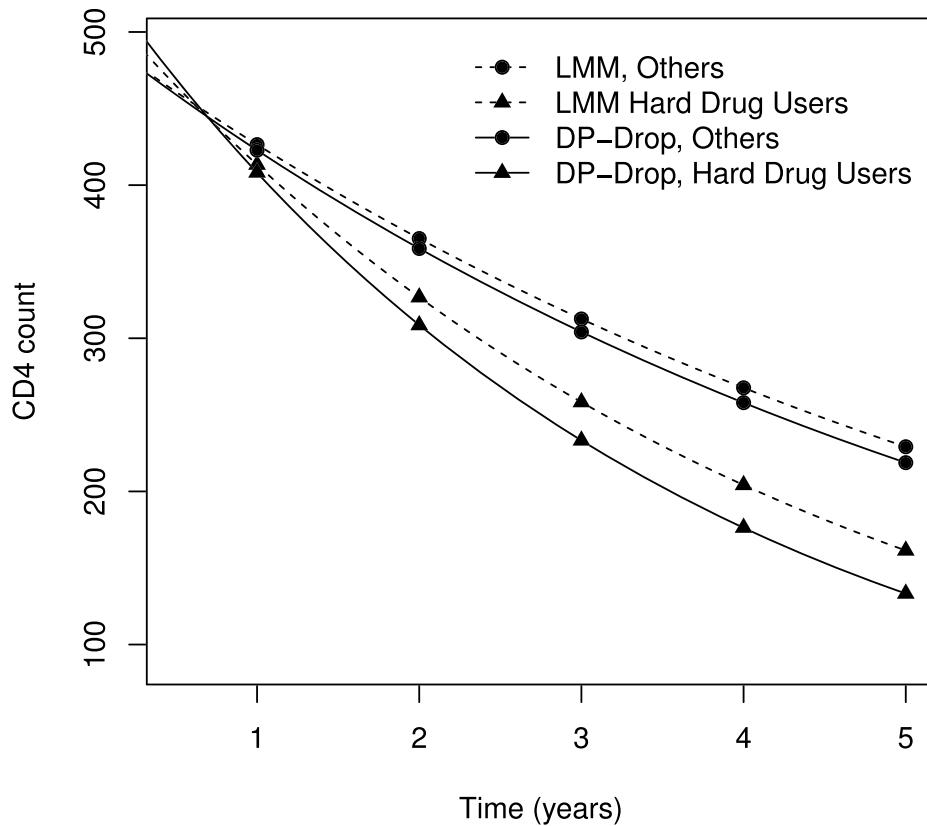


Figure 6: DP-Drop and LMM Posterior Mean CD4⁺ Count Over Time in Untreated Subjects in the WIHS, Assuming a Baseline CD4⁺ Count of 478.

Dropout Time (years)	Others		Hard Drug Users		Difference		PP
	PM	95% CI	PM	95% CI	PM	95% CI	
1	-0.19	(-0.32, -0.01)	-0.32	(-0.58, -0.12)	-0.14	(-0.44, 0.11)	0.85
2	-0.18	(-0.25, -0.10)	-0.34	(-0.49, -0.22)	-0.16	(-0.32, -0.02)	0.99
3	-0.21	(-0.30, -0.12)	-0.41	(-0.65, -0.25)	-0.21	(-0.45, -0.02)	0.99
4	-0.19	(-0.25, -0.14)	-0.31	(-0.56, -0.14)	-0.12	(-0.38, 0.07)	0.87
5	-0.15	(-0.19, -0.12)	-0.19	(-0.36, -0.06)	-0.03	(-0.21, 0.10)	0.66

Table 3: Estimated Dropout Time Specific Change in ln(CD4⁺) per year for Untreated Subjects in the WIHS Using the DP-Drop, Method, Assuming a Baseline CD4⁺ count of 478. PM=posterior mean, CI = Credible Interval, PP=posterior probability of a difference < 0/lower slopes among drug users.

One potential limitation of this analysis is that subjects in our study may have dropped out due to death. Subjects who die do not have potential outcomes after their dropout. Estimates of CD4⁺ presented in marginal plots, such as Figure 6, are averaged over all subjects, including those who may have died. Therefore in the presence of death,

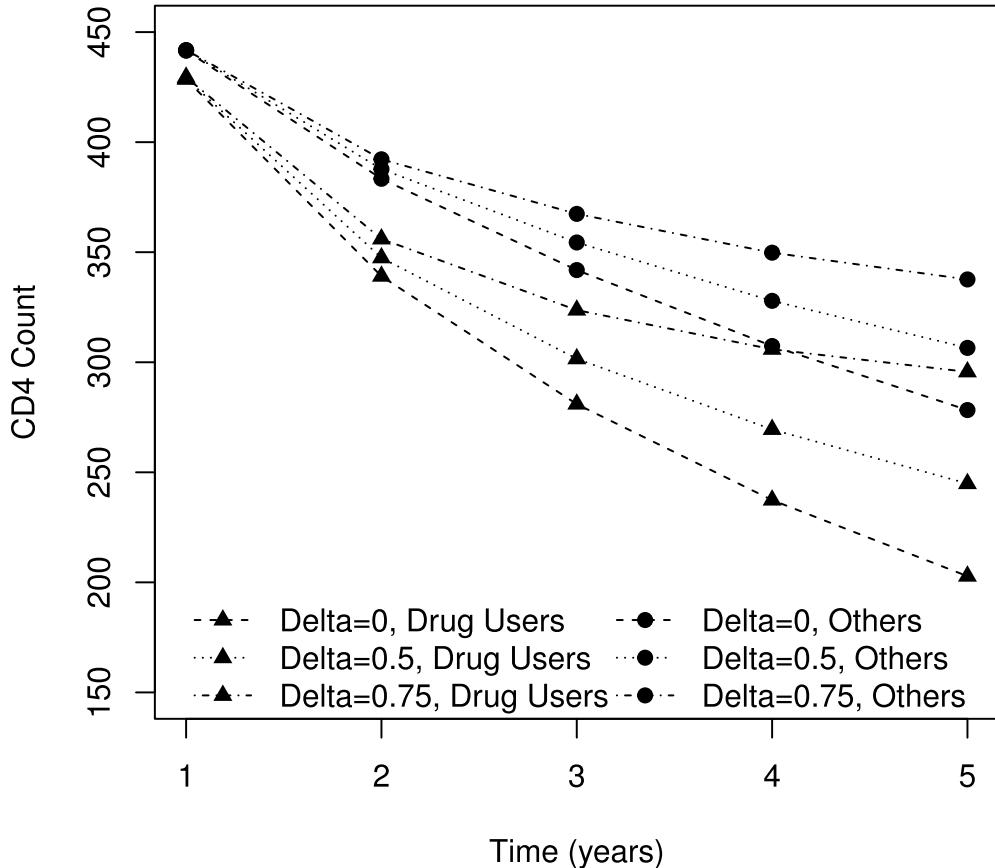


Figure 7: A Sensitivity Analysis of Posterior Mean CD4^+ Count Over Time in Untreated Subjects in the WIHS for $\Delta = 0, 0.5$, and 0.75 .

marginal estimates and plots should not be interpreted as the expected CD4^+ across the surviving patients in the cohort. If dropout due to death is known, it may be more relevant to report the proportion of subjects surviving to the end of the study in each group and the expected longitudinal trajectories for those that survive, or to compare estimates of dropout time specific slopes between groups, as in Figure 5, which do not depend on extrapolating beyond the observed dropout times.

5 Discussion

We propose a flexible Dirichlet process mixture method to account for dropout in a Bayesian framework. The DP-Drop method allows for dropout occurring at any continuous point in time and avoids making parametric assumptions about the distribution of dropout times or the functional form of dropout-varying slope, by modeling both the

Δ	Year	Difference (Drug users – Others)	
		PM	95% CI
0	1	-0.03	(-0.06, 0.00)
	2	-0.10	(-0.14, -0.06)
	3	-0.13	(-0.17, -0.09)
	4	-0.13	(-0.18, -0.09)
	5	-0.13	(-0.18, -0.08)
	0.5	-0.03	(-0.06, 0.00)
	2	-0.11	(-0.16, -0.06)
	3	-0.16	(-0.22, -0.10)
	4	-0.20	(-0.27, -0.12)
	5	-0.22	(-0.32, -0.13)
0.75	1	-0.03	(-0.06, 0.00)
	2	-0.12	(-0.18, -0.07)
	3	-0.20	(-0.28, -0.11)
	4	-0.26	(-0.37, -0.14)
	5	-0.32	(-0.46, -0.17)
1	1	-0.03	(-0.06, 0.00)
	2	-0.14	(-0.20, -0.07)
	3	-0.23	(-0.34, -0.12)
	4	-0.32	(-0.47, -0.17)
	5	-0.41	(-0.61, -0.21)

Table 4: Sensitivity Analysis. Estimated Differences in $\ln(\text{CD4}^+)$ Between Hard Drug Users and Others, for Untreated Subjects in the WIHS for $\Delta = 0, 0.5, 0.75$ and 1.

distribution of the log of dropout time and the subject specific effects as a Dirichlet process mixture of normals. The model flexibly clusters subjects with similar random effects and dropout times in order to model skewed, multi-modal and other non-normal distributions of subject-specific effects and to more accurately characterize longitudinal outcomes when non-ignorable dropout is present. Results of our simulation studies show that the DP-Drop reduces bias and mean squared error for the marginal slope compared to standard GLMMs, frailty models and conditional linear models when non-ignorable dropout is present. Natural extensions to this model might include adding continuous covariates to the class specific model for the random effects or dropout time or additions to account for differing dropout reasons among subjects (Moore et al., 2017; Pauder et al., 2003). In longitudinal studies, subjects may dropout for a variety of reasons, such as loss to follow up, declines or improvements in health, or death. Since subjects with similar dropout reasons are also likely to have similar subject-specific coefficients, we expect that the DP-Drop will cluster subjects with similar dropout reasons together and that the heterogeneity of effects across different dropout reasons would be taken into account. However, as the model does not explicitly include reason, we cannot compare differences in declines between subjects dropping out for different reasons.

Applying the DP-Drop method to the analysis of untreated CD4⁺ count in the WIHS, we find important differences in estimates of CD4⁺ decline for both hard drug

users and others compared to a standard LMM and the CLM and frailty models that account for dropout. In addition, since differential dropout is present, we also find a larger difference in CD4⁺ declines between hard drug users and others. Since we must assume that subjects continue on the same trajectory after dropout in order to identify model parameters, we performed a sensitivity analysis and found that the results of the WIHS analysis are relatively robust to violations of this assumption. Our preliminary investigation of the WIHS data suggested that additional parametric assumptions of common methods to account for non-ignorable dropout, such as log-normally distributed dropout times and/or normally distributed subject specific effects, may be violated. As inferences can be sensitive to these assumptions, having flexible approaches with minimal assumptions, such as the DP-Drop, is important when considering how missing data may influence the results of longitudinal data analyses.

Supplementary Material

Supplementary Material: A Dirichlet Process Mixture Model for Non-Ignorable Dropout
(DOI: [10.1214/19-BA1181SUPP](https://doi.org/10.1214/19-BA1181SUPP); .pdf).

References

- Albert, P. and Follmann, D. (2000). “Modeling repeated count data subject to informative dropout.” *Biometrics*, 56: 667–677. [1141](#)
- Bacon, M. C., von Wyl, V., Alden, C., Sharp, G., Robison, E., Hessol, N., Gange, S., Barranday, Y., Holman, S., Weber, K., and Young, M. A. (2005). “The Women’s Interagency HIV Study: an observational cohort brings clinical sciences to the bench.” *Clinical and Diagnostic Laboratory Immunology*, 12(9): 1013–1019. URL <http://dx.doi.org/10.1128/CDLI.12.9.1013-1019.2005>. [1140](#)
- Barkan, S. E., Melnick, S. L., Preston-Martin, S., Weber, K., Kalish, L. A., Miotti, P., Young, M., Greenblatt, R., Sacks, H., and Feldman, J. (1998). “The Women’s Interagency HIV Study. WIHS Collaborative Study Group.” *Epidemiology*, 9(2): 117–125. [1140](#)
- Blei, D. and Jordan, M. (2006). “Variational inference for Dirichlet process mixtures.” *Bayesian Analysis*, 1: 121–144. [MR2227367](#). doi: <https://doi.org/10.1214/06-BA104>. [1148](#)
- Celeux, G., Hurn, M., and Robert, C. P. (2000). “Computational and inferential difficulties with mixture posterior distributions.” *Journal of the American Statistical Association*, 95(451): 957–970. [MR1804450](#). doi: <https://doi.org/10.2307/2669477>. [1151](#)
- Centers for Disease Control and Prevention (2002). “HIV/AIDS Surveillance Report.” [1140](#)
- Christensen, R., Johnson, W., Branscum, A., and Hanson, T. (2011). *Bayesian Ideas and Data Analysis: An Introduction for Scientists and Statisticians*. CRC Press. [MR2682928](#). [1143](#)

- Daniels, M. and Hogan, J. (2008). *Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis*. Boca Raton: Chapman and Hall/CRC. [MR2459796](#). doi: <https://doi.org/10.1201/9781420011180>. 1140, 1141
- Diggle, P., Liang, K., and Zeger, S. (1994). *Analysis of Longitudinal Data*. Clarendon Press: Oxford. 1141
- Dunson, D. (2010). “Nonparametric Bayes applications to biostatistics.” In Hjort, N., Holmes, C., Muller, P., and Walker, S. (eds.), *Bayesian Nonparametrics*. Cambridge: Cambridge University Press. [MR2730665](#). 1148, 1149
- Ekholm, A. and Skinner, C. (1998). “The Muscatine children’s obesity data reanalysed using pattern mixture models.” *Journal of Applied Statistics*, 47: 251–263. 1141
- Escobar, M. and West, M. (1995). “Bayesian density estimation and inference using mixtures.” *Journal of the American Statistical Association*, 90: 577–588. [MR1340510](#). 1145
- Fairclough, D., Peterson, H., and Chang, V. (1998). “Why are missing quality of life data a problem in clinical trials of cancer therapy?” *Statistics in Medicine*, 17: 667–677. 1140
- Fitzmaurice, G., Laird, N., and Shneyer, L. (2001). “An alternative parameterization of the general linear mixture model for longitudinal data with non-ignorable drop-outs.” *Statistics in Medicine*, 20: 1009–1021. 1141
- Follmann, D. and Wu, M. (1995). “An approximate generalized linear model with random effects for informative missing data.” *Biometrics*, 51(1): 151–168. [MR1341233](#). doi: <https://doi.org/10.2307/2533322>. 1141
- Forster, J., MaWhinney, S., Ball, E., and Fairclough, D. (2012). “A varying-coefficient method for analyzing longitudinal clinical trials data with nonignorable dropout.” *Contemporary Clinical Trials*, 33: 378–385. 1144
- Forster, J., MaWhinney, S., and Wang, X. (2013). “A natural b-spline varying-coefficient method for longitudinal binary response data with nonignorable dropout.” American Statistical Association, Joint Statistical Meetings. August, 2013. Montreal CA.
- Genz, A., Bretz, F., Miwa, T., Mi, X., Leisch, F., Scheipl, F., and Hothorn, T. (2014). *mvtnorm: Multivariate Normal and t Distributions*. R package version 1.0-2. URL <http://CRAN.R-project.org/package=mvtnorm>. 1152 1152
- Ghosal, S. (2010). “The Dirichlet process, related priors and posterior asymptotics.” In Hjort, N., Holmes, C., Muller, P., and Walker, S. (eds.), *Bayesian Nonparametrics*. Cambridge: Cambridge University Press. [MR2730660](#). 1143
- Gray, K., Hampton, B., Silveti-Falls, T., McConnell, A., and Bausell, C. (2015). “Comparison of Bayesian Credible Intervals to Frequentist Confidence Intervals.” *Journal of Modern Applied Statistical Methods*, 14(1): 8. 1153
- Heckman, J. (1979). “Sample selection bias as a specification error.” *Econometrica*, 47: 153–161. [MR0518832](#). doi: <https://doi.org/10.2307/1912352>. 1141

- Heckman, J. (1998). "Selection models for repeated measurements with non-random dropout: an illustration of sensitivity." *Statistics in Medicine*, 17: 2723–2732. [doi: https://doi.org/10.1002/\(SICI\)1099-0256\(199809\)17:21.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-0256(199809)17:21.0.CO;2-1) 1141
- Hogan, J., Lin, X., and Herman, B. (2004a). "Mixtures of varying-coefficient models for longitudinal data with discrete or continuous nonignorable dropout." *Biometrics*, 60: 854–864. [MR2133537](https://doi.org/10.1111/j.0006-341X.2004.00240.x). doi: <https://doi.org/10.1111/j.0006-341X.2004.00240.x>. 1144
- Hogan, J., Roy, J., and Korkontzelou, C. (2004b). "Tutorial in biostatistics, handling drop-out in longitudinal studies." *Statistics in Medicine*, 23: 1455–1497. [doi: https://doi.org/10.1002/\(SICI\)1099-0256\(200407\)23:11.0.CO;2-1](https://doi.org/10.1002/(SICI)1099-0256(200407)23:11.0.CO;2-1) 1140
- Ibrahim, J., Chen, M., and Lipsitz, S. (2001). "Missing responses in generalised linear mixed models when the missing data mechanism is nonignorable." *Biometrika*, 88: 551–564. [MR1844851](https://doi.org/10.1093/biomet/88.2.551). doi: <https://doi.org/10.1093/biomet/88.2.551>. 1141
- Ibrahim, J. and Molenberghs, G. (2009). "Missing data methods in longitudinal studies: a review." *Test*, 18: 1–43. [MR2495958](https://doi.org/10.1007/s11749-009-0138-x). doi: <https://doi.org/10.1007/s11749-009-0138-x>. 1141
- Ishwaran, H. and James, L. (2001). "Gibbs Sampling Methods for Stick-Breaking Priors." *Journal of the American Statistical Association*, 96: 161–173. [MR1952729](https://doi.org/10.1198/016214501750332758). doi: <https://doi.org/10.1198/016214501750332758>. 1145, 1148
- Ishwaran, H. and Takahara, G. (2002). "Independent and identically distributed Monte Carlo algorithms for semiparametric linear mixed models." *Journal of the American Statistical Association*, 97: 1154–1166. [MR1951267](https://doi.org/10.1198/016214502388618951). doi: <https://doi.org/10.1198/016214502388618951>. 1148
- Jain, S. and Neal, R. (2004). "A split-merge Markov chain Monte Carlo procedure for the Dirichlet process mixture model." *Journal of Computational and Graphical Statistics*, 13: 158–182. [MR2044876](https://doi.org/10.1198/1061860043001). doi: <https://doi.org/10.1198/1061860043001>. 1148
- Kaciroti, N., Raghunathan, T., Schork, M., and Clark, N. (2008). "A Bayesian model for longitudinal count data with non-ignorable dropout." *Journal of the Royal Statistical Society (Series C) Applied Statistics*, 57: 521–534. [MR2528669](https://doi.org/10.1111/j.1467-9876.2008.00628.x). doi: <https://doi.org/10.1111/j.1467-9876.2008.00628.x>. 1143
- Kaciroti, N., Raghunathan, T., Taylor, J., and Julius, S. (2012). "A Bayesian model for time-to-event data with informative censoring." *Biostatistics*, 13: 341–354. [doi: https://doi.org/10.1093/biostatistics/kqr031](https://doi.org/10.1093/biostatistics/kqr031) 1143
- Kaciroti, N., Schork, M., Raghunathan, T., and Julius, S. (2009). "A Bayesian sensitivity model for intention-to-treat analysis on binary outcomes with dropouts." *Statistics in Medicine*, 28: 572–585. [MR2655731](https://doi.org/10.1002/sim.3494). doi: <https://doi.org/10.1002/sim.3494>. 1143
- Lancaster, T. and Intrator, O. (1998). "Panel data with survival: hospitalization of HIV-positive patients." *Journal of the American Statistical Association*, 93: 46–53. [doi: https://doi.org/10.1080/01621459808818670](https://doi.org/10.1080/01621459808818670) 1141
- Lanoya, E., Mary-Krause, M., Tattevinb, P., Dray-Spirac, R., Duvivierd, C., Fischere, P., Obadiaf, Y., and Lert, F. (2006). "Predictors identified for losses to follow-up among HIV-seropositive patients." *Journal of Clinical Epidemiology*, 59: 829–835. [doi: https://doi.org/10.1016/j.jclinepi.2006.01.014](https://doi.org/10.1016/j.jclinepi.2006.01.014) 1139

- Linero, A. and Daniels, M. (2015). “A Flexible Bayesian Approach to Monotone Missing Data in Longitudinal Studies with Nonignorable Missingness with Application to an Acute Schizophrenia Clinical Trial.” *Journal of the American Statistical Association*, 110: 45–55. MR3338485. doi: <https://doi.org/10.1080/01621459.2014.969424>. 1143, 1144, 1146, 1147
- Linero, A. R. (2017). “Bayesian nonparametric analysis of longitudinal studies in the presence of informative missingness.” *Biometrika*, 104(2): 327–341. MR3698257. doi: <https://doi.org/10.1093/biomet/asx015>. 1146, 1147, 1151
- Little, J. and Rubin, D. (2002). *Statistical Analysis with Missing Data, Second Edition*. Hoboken, New Jersey: John Wiley & Sons, Inc. MR1925014. doi: <https://doi.org/10.1002/9781119013563>. 1141
- Little, R. (1993). “Pattern-mixture models for multivariate incomplete data.” *Journal of the American Statistical Association*, 88: 125–134. 1140
- MacEachern, S. (1994). “Estimating normal means with a conjugate style Dirichlet process prior.” *Communications in Statistics: Simulation and Computation*, 23: 727–741. MR1293996. doi: <https://doi.org/10.1080/03610919408813196>. 1148
- MacEachern, S., Clyde, M., and Liu, J. (1999). “Sequential importance sampling for nonparametric Bayes models: The next generation.” *Canadian Journal of Statistics*, 27: 251–267. MR1704407. doi: <https://doi.org/10.2307/3315637>. 1148
- Makov, D., Smith, A., and Titterington, D. (1985). “Statistical Analysis of Finite Mixture Distributions.” Chichester-New York: J. Wiley & Sons. MR0838090. 1151
- Martin, A. D., Quinn, K. M., and Park, J. H. (2011). “MCMCpack: Markov Chain Monte Carlo in R.” *Journal of Statistical Software*, 42(9): 22. URL <http://www.jstatsoft.org/v42/i09/>. 1152
- Molenberghs, G., Kenward, M., and Lesaffre, E. (1997). “The analysis of longitudinal ordinal data with non-random dropout.” *Biometrika*, 84: 33–44. 1140, 1144
- Molenberghs, G., Thijs, H., Jansen, I., Beunckens, C., Kenward, M., Mallinckrodt, C., and Carroll, R. (2004). “Analyzing incomplete longitudinal clinical trial data.” *Biostatistics*, 5: 445–464. 1140
- Moore, C. (2013). “Knot Selection Strategies for Semiparametric Varying-Coefficient Models Applied to Longitudinal Cohorts with Multiple Dropout Reasons.” Master’s thesis, Colorado School of Public Health, University of Colorado Denver. 1152
- Moore, C., MaWhinney, S., Forster, J., Carlson, N., Allshouse, A., Wang, X., Routy, J., Conway, B., and Connick, E. (2017). “Accounting for dropout reason in longitudinal studies with nonignorable dropout.” *Statistical Methods in Medical Research*, 1854–1866. MR3687182. doi: <https://doi.org/10.1177/0962280215590432>. 1140, 1144, 1161
- Moore, C., Carlson, N., MaWhinney, S., and Kreidler, S. (2019). “Supplementary Material: A Dirichlet Process Mixture Model for Non-Ignorable Dropout.” *Bayesian Analysis*. doi: <https://doi.org/10.1214/19-BA1181SUPP>. 1148

- Newton, M. and Zhang, Y. (1999). “A recursive algorithm for nonparametric analysis with missing data.” *Biometrika*, 86: 15–26. [MR1688068](#). doi: <https://doi.org/10.1093/biomet/86.1.15>. 1148
- Pauler, D., 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. 1144, 1161
- Rubin, D. (1977). “Formalizing subjective notions about the effect of nonrespondents in sample surveys.” *Journal of the American Statistical Association*, 72: 538–543. [MR0445668](#). 1141
- Rubin, D. (1981). “The Bayesian bootstrap.” *Annals of Statistics*, 9: 130–134. [MR0600538](#). 1146
- Schluchter, M. (1992). “Methods for the analysis of informatively censored longitudinal data.” *Statistics in Medicine*, 11: 1861–1870. 1141, 1143, 1145, 1146, 1147
- Schluchter, M., Greene, T., and Beck, J. (2001). “Analysis of change in the presence of informative censoring: application to a longitudinal clinical trial of progressive renal disease.” *Statistics in Medicine*, 20: 989–1007. 1143, 1152
- Sethuraman, J. (1994). “A constructive definition of Dirichlet priors.” *Statistica Sinica*, 4: 639–650. [MR1309433](#). 1145
- Shi, Y., Martens, M., Banerjee, A., Laud, P., et al. (2019). “Low Information Omnibus (LIO) Priors for Dirichlet Process Mixture Models.” *Bayesian Analysis*, 14(3): 677–702. [MR3960766](#). doi: <https://doi.org/10.1214/18-BA1119>. 1150
- Stephens, M. (2000). “Dealing with label switching in mixture models.” *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 62(4): 795–809. [MR1796293](#). doi: <https://doi.org/10.1111/1467-9868.00265>. 1151
- Su, L. and Hogan, J. W. (2008). “Bayesian semiparametric regression for longitudinal binary processes with missing data.” *Statistics in Medicine*, 27(17): 3247–3268. [MR2523915](#). doi: <https://doi.org/10.1002/sim.3265>. 1143
- Su, L. and Hogan, J. W. (2010). “Varying-coefficient models for longitudinal processes with continuous-time informative dropout.” *Biostatistics*, 11(1): 93–110. [MR2843515](#). doi: <https://doi.org/10.1155/2012/734341>. 1143, 1146
- Ten Have, T., Kunselman, A., Pulkstenis, E., and Landis, J. (1998). “Mixed effects logistic regression models for longitudinal binary response data with informative dropout.” *Biometrics*, 54: 367–383. 1141
- Venables, W. N. and Ripley, B. D. (2002). *Modern Applied Statistics with S*. New York: Springer, fourth edition. ISBN 0-387-95457-0. URL <http://www.stats.ox.ac.uk/pub/MASS4>. [MR1337030](#). doi: https://doi.org/10.1007/978-1-4899-2819-1_1152
- Warnes, G. R., Bolker, B., and Lumley, T. (2014). *gtools: Various R programming tools*. R package version 3.4.1. URL <http://CRAN.R-project.org/package=gtools>. 1152

- Wasserman, L. (2011). “Frasian inference.” *Statistical Science*, 322–325. [MR2917980](#). doi: <https://doi.org/10.1214/11-STS352C>. 1153
- Wei, G. and Tanner, M. (1990). “Posterior computations for censored regression data.” *Journal of the American Statistical Association*, 85: 829–839. [MR1294083](#). 1148
- Wu, K. and Wu, L. (2007). “Generalized linear mixed models with informative dropouts and missing covariates.” *Metrika*, 66: 1–18. [MR2306373](#). doi: <https://doi.org/10.1007/s00184-006-0083-6>. 1141
- Wu, M. and Bailey, K. (1988). “Analysing changes in the presence of informative right censoring caused by death and withdrawal.” *Statistics in Medicine*, 7: 337–346. 1141
- Wu, M. and Bailey, K. (1989). “Estimation and comparison of changes in the presence of informative right censoring; Conditional linear model.” *Biometrics*, 45: 939–955. [MR1029611](#). doi: <https://doi.org/10.2307/2531694>. 1141, 1144, 1152

Acknowledgments

The authors thank Professor W.O. Johnson for inspiring this work.

Data in this manuscript were collected by the Women’s Interagency HIV Study (WIHS). The contents of this publication are solely the responsibility of the authors and do not represent the official views of the National Institutes of Health (NIH). WIHS (Principal Investigators): Bronx WIHS (Kathryn Anastos), U01-AI-035004; Brooklyn WIHS (Howard Minkoff and Deborah Gustafson), U01-AI-031834; Chicago WIHS (Marge Cohen and Audrey French), U01-AI-034993; Metropolitan Washington WIHS (Seble Kassaye), U01-AI-034994; Northern California (Ruth Greenblatt, Bradley Aouizerat, and Phyllis Tien), U01-AI-034989; WIHS Data Management and Analysis Center (Stephen Gange and Elizabeth Golub), U01-AI-042590. The WIHS is funded primarily by the National Institute of Allergy and Infectious Diseases (NIAID), with additional co-funding from the Eunice Kennedy Shriver National Institute of Child Health and Human Development (NICHD), the National Cancer Institute (NCI), the National Institute on Drug Abuse (NIDA), and the National Institute on Mental Health (NIMH). Targeted supplemental funding for specific projects is also provided by the National Institute of Dental and Craniofacial Research (NIDCR), the National Institute on Alcohol Abuse and Alcoholism (NIAAA), the National Institute on Deafness and other Communication Disorders (NIDCD), and the NIH Office of Research on Women’s Health. WIHS data collection is also supported by UL1-TR000004 (UCSF CTSA) and UL1-TR000454 (Atlanta CTSA).