7-25-2007

A Censored Multinomial Regression Model for Perinatal Mother to Child Transmission of HIV

Charlotte C. Gard
University of Washington, gardc@u.washington.edu

Elizabeth R. Brown
University of Washington, elizab@u.washington.edu

1 Introduction

In trials designed to evaluate the efficacy of an intervention to prevent perinatal mother to child transmission (PMTCT) of human immunodeficiency virus (HIV), the primary endpoint is often the cumulative transmission rate at a point in time shortly after birth. To determine HIV status, the infants are usually tested within 48 hours after birth to assess in utero transmission. A second visit is often scheduled 4 to 8 weeks after birth to assess intrapartum transmission. Subsequent visits may be scheduled, but tests at these visits only contribute information about the primary endpoint if the infant has missed earlier scheduled visits. If infants are tested at both scheduled visit times, estimation of the endpoint is straightforward, as are regression models for the endpoint. Unfortunately, missed visits and off-schedule visits are not uncommon. And, even if there are no missed visits, interim analyses may occur when only a fraction of the infants are old enough for the second visit.

In this paper, we propose a censored multinomial regression model for analyzing PMTCT of HIV. The approach is motivated by the HIV Prevention Trials Network (HPTN) 024 study, a multi-site placebo-controlled trial of antiobiotics to prevent chorioamnionitis and, therefore, perinatal transmission of HIV. From the first two testing windows, there were three main outcomes of scientific interest:

A1 The in utero transmission rate, estimated by the fraction of infants testing positive shortly after birth;

A2 The perinatal transmission rate, estimated by the fraction of infants testing positive by 6 weeks (which we extended to 8 weeks, for analysis purposes);

A3 The intrapartum transmission rate, estimated by the fraction of infants testing positive by the end of the perinatal transmission window, given they had a negative test result at birth.

In primary analyses, we are usually interested in obtaining unadjusted estimates of A1, A2, and A3. In this manuscript, we focus on secondary analyses, where adjusted estimates are often desired. Ideally, every subject would be tested in every visit window, and we could use a binary endpoint approach such as logistic regression to model all three outcomes of scientific interest. However, we rarely have complete test result data in PMTCT clinical trials. For example, in HPTN 024, of the 2052 liveborn infants, 1813 (88%) had HIV tests within 48 hours of delivery, 1696 (83%) had tests 4 to 8 weeks after delivery, and only 1584

(77%) had tests both within 48 hours of delivery and 4 to 8 weeks after delivery. While missed visits may be due to the infant's death (which can be accounted for), in some cases the mother simply forgets or is unable to bring the infant in for follow-up. Often, mothers do not deliver at the study hospital and must bring the infant in at a later point in time for the HIV test. Such missed and off-schedule visits make the usual analytical methods problematic.

Table 1 lists a selection of primary papers from trials aimed at reducing PMTCT of HIV and summarizes the methods used for unadjusted and adjusted analyses, as well as how the data were censored in each case. The methods represented are among the more commonly used for estimating PMTCT of HIV. In adjusted analyses, the endpoint is generally modeled as either binary or right-censored continuous, using logistic or Cox proportional hazards (PH) regression, respectively. For both logistic and Cox PH models, methods currently used for handling missing data may be inadequate. For example, when the logistic model is used and a test result is missing for an infant who has not previously tested positive, the observation is dropped, although if subsequent tests are negative, the missing test result may be imputed to be negative. When the Cox PH model is used, an infant's time to HIV infection is right censored at his or her last negative test; however, approaches for addressing the timing of infection when a missing visit is followed by a positive test and there have been no previous positive tests may be inadequate. Some authors use the time of the first positive test as the time of infection while others use the midpoint between the last negative and the first positive tests (or birth and the first positive test) as the time of infection. Instead, we propose a method that reflects the intention of the studies to classify infections according to their timing (in utero, peripartum and postpartum) using a multinomial model. The model accommodates incomplete longitudinal observations by allowing general censoring and also accommodates regression on the three outcomes of interest (A1-A3).

This manuscript proceeds as follows. In Section 2, we describe the censored multinomial model and estimation of the parameters for a single sample. In Section 3, we lay out several strategies for adjusting for covariates. In doing so, we consider estimation for all three scientific endpoints of interest. In Section 4, we describe simulations designed to evaluate the performance of the proposed regression estimators. In Section 5, we present an example based on HPTN 024. Discussion follows in Section 6.

Table 1: Approaches taken in selected papers analyzing PMTCT of HIV

	Unadjusted	Adjusted	Censoring
Wiktor et al. (1999)	KM[a]	–	at last negative test
Guay et al. (1999)	KM	PH[b]	at last negative test
Dabis et al. (1999)	KM	PH	not stated
Shaffer et al. (1999)	KM	logistic regression	not stated
Fawzi et al. (2000)	Chi-square tests	PH	not stated
Dorenbaum et al. (2002)	Fisher's exact tests	logistic regression	–
Moodley et al. (2003)	KM	PH and logistic regression	at last follow-up

[a]Kaplan-Meier.
[b]Cox proportional hazards.

2 Estimation for a Single Sample

In this section, we present the censored multinomial model and maximum likelihood methods for parameter estimation with a single sample. In this general presentation, we assume that there are J visit windows. Usually, when estimating PMTCT, $J = 2$. However, depending upon the study design, J may be larger as in Wiktor et al. (1999), where $J = 3$.

We begin by dividing the follow-up time into windows as follows:

$$\begin{aligned} \text{First visit window} \quad & [t_{11}, t_{12}) \\ \text{Second visit window} \quad & [t_{21}, t_{22}) \\ & \vdots \\ \text{Time following last visit window} \quad & [t_{J+1,1}, \infty). \end{aligned}$$

Here, J is the number of visit windows of interest, and t_{j1} and t_{j2} indicate the times at which the jth visit window starts and ends, respectively. These intervals do not have to be and usually are not contiguous. In other words, t_{j2} is not necessarily equal to $t_{j+1,1}$. Unscheduled or off-schedule visits result in tests that occur in the interval $[t_{j2}, t_{j+1,1})$. An example of potential visit windows is shown in Figure 1 for four subjects **A**, **B**, **C**, and **D**. In this example, $J = 4$.

We define a complete response vector for the ith subject as $Y_i^* = (Y_{i1}^*, \ldots, Y_{iJ}^*)'$ where $Y_{ij}^* = 1$, $j = 1, \ldots, J$, if the subject tests positive for the first time at the jth visit and 0 otherwise. The vector Y_i^* represents a multinomial response as the ith subject can only test positive for HIV for the first time once. When a subject misses a scheduled visit (including

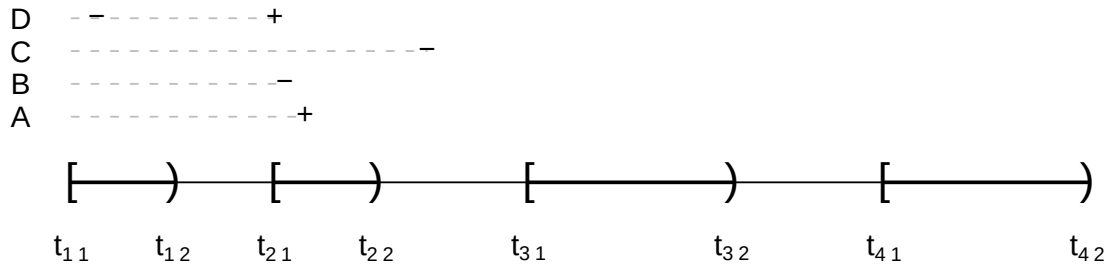

Figure 1: Example of possible scheduling of testing visits.

if he or she is late or early for the visit), we observe a censored version of Y_i^*. Let Y_i denote the observed endpoint, which is defined as

$$Y_{i1} = \begin{cases} 1, & t_i^p \in [0, t_{12}) \\ 1, & t_i^p \geq t_{12} \text{ and } t_i^n < t_{11} \\ 0, & otherwise \end{cases} \quad (1)$$

$$Y_{ij} = \begin{cases} 1, & t_i^p \in [t_{j-1,2}, t_{j2}) \\ 1, & t_i^p \geq t_{j2} \text{ and } t_i^n < t_{j1} \\ 0, & otherwise \end{cases} \quad , \ j = 2, \ldots, J, \quad (2)$$

$$Y_{i,J+1} = \begin{cases} 1, & t_i^p \geq t_{J2} \\ 0, & otherwise \end{cases} . \quad (3)$$

Here, t_i^p and t_i^n are the time of the first positive test result for subject i and the time of the last negative test result for subject i, respectively. For subjects with no positive test result during follow-up, we take t_i^p equal to ∞. For subjects with no negative test result during follow-up, we take t_i^n equal to -∞. We assume that each subject has at least one (non-missing) test result during the follow-up period.

To illustrate how the observed vector Y relates to the unobserved but true outcome Y^*, we look at four possible visit and outcome patterns. Following Figure 1, we assume that the visit windows are not contiguous. First, we examine the effect of a missed visit. Suppose subject **A** is not tested until the second visit at which point he or she tests positive. We do not know if subject **A** would have tested positive had he or she come in for the first visit. We can say, however, that the subject would have tested positive for the first time at the first visit or at the second visit if tested at both visits. In other words, Y^* for this subject

may be $(0,1)'$ or $(1,0)'$ but is not $(0,0)'$. Therefore, by (1), (2), and (3), $Y = (1,1,0)'$.

Next, we consider subject **B** who missed the first visit and tested negative at the second visit. Here we assume that, if a subject tests negative at the end of the study, he or she was negative throughout the study; therefore, $Y^* = (0,0)'$ and $Y = (0,0,1)'$ for this subject. In this case, even though the subject was not tested in every visit window, we have complete information regarding his or her outcome. This illustrates another difficulty involving missed visits. If a subject is uninfected and misses all visits except the last (the Jth visit), we still have complete information about him or her (as in **B**); however, if the subject is infected at the last visit (as in **A**), we have incomplete information about him or her. Returning to the example, we also have complete information for subjects **C** and **D**, for whom $Y^* = (0,0)'$ and $Y^* = (0,1)'$, corresponding to $Y = (0,0,1)'$ and $Y = (0,1,0)'$, respectively.

Letting $\mathcal{J}_j$ be a vector of length J with jth element equal to 1 and all other elements equal to zero, we define the probability that a subject's first positive test occurs in the jth visit window as $p_j = P(Y_i^* = \mathcal{J}_j),\ j = 1, \ldots, J$, and the probability that a subject's first positive test occurs after the last visit window as $p_{J+1} = 1 - \sum_{j=1}^{J} p_j$. Each subject's contribution to the likelihood is given by

$$f(y_i) = Y_i' p,$$

where $p = (p_1, \ldots, p_{J+1})'$. The log-likelihood can be written as

$$l(p_1, \ldots, p_J) = \sum_{i=1}^{N} \log \Big[\sum_{j=1}^{J} Y_{ij} p_j + Y_{i,J+1}(1 - \sum_{j=1}^{J} p_j) \Big], \tag{4}$$

where the outer summation is over all individuals in the sample. Maximum likehood estimates of $p_1, \ldots, p_J$, denoted $\hat{p} = (\hat{p}_1, \ldots, \hat{p}_J)'$, are obtained by maximizing (4) using numerical optimization techniques. In this maximization, probabilities $p_1, \ldots, p_J$ are constrained to lie between 0 and 1 and $\sum_{j=1}^{J} p_j$ is constrained to be less than 1.

We focus now on estimating the cumulative probabilities of transmission at the birth and 4 to 8 week visits. These correspond to the in utero and perinatal endpoints (A1 and A2) described in Section 1. The birth transmission rate, P_1, is estimated by $\widehat{P}_1 = \hat{p}_1$. The 4 to 8 week transmission rate, $P_2 = p_1 + p_2$, is estimated by $\widehat{P}_2 = \hat{p}_1 + \hat{p}_2$. In general, the cumulative transmission rate is obtained as $P_j = \sum_{k=1}^{j} p_k,\ j \leq J$, with maximum likelihood estimate $\widehat{P}_j = \sum_{k=1}^{j} \hat{p}_k$. By the multivariate delta method (Agresti, 2002, page 579), the asymptotic variance of $\widehat{P}_j$ is given by $v_j = \sum_{k=1}^{j} \sum_{l=1}^{j} V_{kl}$, where V_{kl} denotes the element in the kth row and lth column of V, the covariance matrix of $\hat{p}_1, \ldots, \hat{p}_J$. We estimate V using the negative

of the inverse of the Hessian matrix.

We turn now to consider endpoint A3, the intrapartum transmission rate. The intrapartum transmission rate represents the probability that an infant is infected during delivery and is given by

$$p_{2|1^-} = \frac{p_2}{1-p_1}.$$

We estimate $p_{2|1^-}$ as the fraction of infants who test positive by the end of the second visit window given that they had a negative test result at birth. That is, we estimate $p_{2|1^-}$ as

$$\widehat{p_{2|1^-}} = \frac{\hat{p_2}}{1-\hat{p_1}}.$$

In a breastfeeding population, where HIV transmission may also occur via breastfeeding, $\widehat{p_{2|1^-}}$ represents the rate of delivery or early postnatal transmission.

Using the multivariate delta method, we derive the asymptotic variance of $\widehat{p_{2|1^-}}$ as

$$w = (\frac{p_2}{(1-p_1)^2})^2 \times \text{var}(\hat{p_1}) + \frac{2p_2}{(1-p_1)^3} \times \text{cov}(\hat{p_1}, \hat{p_2}) + \frac{1}{(1-p_1)^2} \times \text{var}(\hat{p_2}).$$

We estimate this quantity using

$$\hat{w} = (\frac{\hat{p_2}}{(1-\hat{p_1})^2})^2 \times \widehat{\text{var}}(\hat{p_1}) + \frac{2\hat{p_2}}{(1-\hat{p_1})^3} \times \widehat{\text{cov}}(\hat{p_1}, \hat{p_2}) + \frac{1}{(1-\hat{p_1})^2} \times \widehat{\text{var}}(\hat{p_2}),$$

where $\widehat{\text{var}}(\hat{p_1})$, $\widehat{\text{cov}}(\hat{p_1}, \hat{p_2})$, and $\widehat{\text{var}}(\hat{p_2})$ are obtained using the negative of the inverse of the Hessian matrix.

In the case where there are more than two visit windows, we can denote the probability of testing positive at the j^{th} visit given that the first $j-1$ tests are negative as

$$p_{j|(j-1)^-} = Pr\left\{Y_i^* = \mathcal{J}_j | Y_i^* \neq \mathcal{J}_1, \ldots, \mathcal{J}_{j-1}\right\},$$

which can be written as

$$p_j = p_{j|(j-1)^-} \times (1 - \sum_{k=1}^{j-1} p_k).$$

The asymptotic variance of $\widehat{p_{j|(j-1)^-}}$ is given by

$$
\begin{aligned}
w_j &= \sum_{l=1}^{j-1}\sum_{l'=1}^{j-1}\Big[\Big(\frac{p_j}{(1-\sum_{k=1}^{j-1}p_k)^2}\Big)^2 \times V_{ll'}\Big] \\
&\quad +2\sum_{l'=1}^{j-1}\Big[\frac{p_j}{(1-\sum_{k=1}^{j-1}p_k)^3}\times V_{jl'}\Big] \\
&\quad +\frac{1}{(1-\sum_{k=1}^{j-1}p_k)^2}\times V_{jj}. \qquad (5)
\end{aligned}
$$

Here, $V_{ll'}$ denotes the element in the lth row and l'th column of V, the covariance matrix of $\hat{p}_1, \ldots, \hat{p}_J$. $V_{jl'}$ denotes the element in the jth row and l'th column of V, and V_{jj} denotes the element in the jth row and jth column of V. We use the negative of the inverse of the Hessian matrix to estimate $V_{ll'}$, $V_{jl'}$, and V_{jj}. We substitute these estimates, with $\hat{p}_1, \ldots, \hat{p}_j$, into (5) to obtain estimates of w_j.

3 Regression Approaches

In the previous section, we presented methods for estimating one sample or unadjusted probabilities of HIV transmission, corresponding to endpoints A1 through A3. In this section, we use a regression approach to determine how cumulative and conditional probabilities are associated with predictors. In the spirit of assessing PMTCT of HIV, models are presented for the case of two visit windows, birth and 4 to 8 weeks. We also address how these models can be extended to analyze data from studies with more than two visit windows.

3.1 Cumulative Probabilities

We define the probability that subject i's first positive test occurs in the jth visit window as π_{ij} and the probability that subject i's first positive test occurs after the last visit window as $\pi_{i,J+1} = 1 - \sum_{j=1}^{J}\pi_{ij}$. Unlike in the previous section, here we do not assume that the probability of testing positive for the first time in the jth visit window is the same for all subjects. To examine the relationship between a set of predictors, $(X_{i1}, \ldots, X_{im})$, and the probability that subject i tests positive at or before the jth visit, we define the following regression model:

$$g(\pi_{i1}) = X_i'\beta_1 \tag{6}$$

$$g(\sum_{k=1}^{j}\pi_{ik}) = X_i'\beta_j \quad , \ j = 2, \ldots, J, \tag{7}$$

where $g(\cdot)$ is a link function that specifies the relationship between the predictors, $X_i = (1, X_{i1}, \ldots, X_{im})'$, and the response, through the parameter vector $\beta_j = (\beta_{j0}, \ldots, \beta_{jm})'$, of length m+1. For ease of exposition, we assume that a predictor is relevant for all visit windows. Therefore, if a predictor is included in the model, it is included for all regressions.

When modeling cumulative probabilities, two appropriate choices for the link function are the log link, where $g(p) = \log(p)$, and the logit link, where $g(p) = \text{logit}(p) = \log\{p/(1-p)\}$. Here, we focus on the logit link, where β_{jl}, $l = 1 \ldots, m$, is interpreted as the change in the log odds of testing positive at or before the jth visit window per one unit increase in X_{il}, $l = 1, \ldots, m$.

We consider the case of two visit windows, corresponding to birth and 4 to 8 weeks. The regression model is defined as

$$\text{logit}(\pi_{i1}) = X_i'\beta_1$$
$$\text{logit}(\pi_{i1} + \pi_{i2}) = X_i'\beta_2,$$

where β_1 and β_2 are parameter vectors of length $m + 1$ linking the predictors to the odds of testing positive at the birth visit and the odds of testing positive at or before the 4 to 8 week visit, respectively.

Returning to the general case, we combine and re-write equations (6) and (7) to obtain the following expressions for π_{ij}:

$$\pi_{i1} = g^{-1}(X_i'\beta_1) \tag{8}$$

$$\pi_{ij} = g^{-1}(X_i'\beta_j) - g^{-1}(X_i'\beta_{j-1}) \quad , \ j = 2, \ldots, J. \tag{9}$$

The log-likelihood is given by

$$l(\pi_{i1}, \ldots, \pi_{iJ}) = \sum_{i=1}^{N} \log\Big[\sum_{j=1}^{J} Y_{ij}\pi_{ij} + Y_{i,J+1}(1 - \sum_{j=1}^{J}\pi_{ij})\Big], \tag{10}$$

where the outer summation is over all individuals in the sample.

To obtain maximum likelihood estimates of the regression parameters, we maximize (10) using numerical optimization techniques (discussed in Section 3.3). During optimization, we must ensure that $\pi_{i1}, \ldots, \pi_{iJ}$ lie between 0 and 1 and that $\sum_{j=1}^{J} \pi_{ij} < 1$ for all i. The logit link imposes the constraints that $0 < \pi_{i1} < 1$ and $\sum_{j=1}^{J} \pi_{ij} < 1$; however, it does not guarantee that $0 < \pi_{ij} < 1, j = 2, \ldots, J$. Instead, we impose this set of constraints through the optimization procedure in the form of non-linear constraints on the coefficients. Further implications of the constraints are presented in the discussion.

3.2 Conditional Probabilities

We now consider regressions on conditional probabilities in order to estimate endpoint A3, the intrapartum transmission rate. We begin by considering the case where there are two visit windows, corresponding to birth and 4 to 8 weeks. To examine the relationship between a set of predictors, $(X_{i1}, \ldots, X_{im})$, and $\pi_{i2|1^-}$, the probability that subject i tests positive at the 4 to 8 week visit given he or she tested negative at birth, we define the following regression model:

$$\begin{aligned} g(\pi_{i1}) &= X_i'\beta_1 \\ g(\pi_{i2|1^-}) &= X_i'\beta_{2|1^-}, \end{aligned}$$

where $g(\cdot)$ is a link function that specifies the relationship between the predictors, $X_i = (1, X_{i1}, \ldots, X_{im})'$, and the response, through the parameters vectors β_1 and $\beta_{2|1^-}$, each of length $m+1$. If we choose the logit link, $\beta_{2|1^-}$ represents the change in the log odds of testing positive at the 4 to 8 week visit, given a negative result at birth, per one unit increase in X_{il}, $l = 1, \ldots, m$.

We calculate π_{i2} for use in the log-likelihood as

$$\begin{aligned} \pi_{i2} &= \pi_{i2|1^-} \times (1 - \pi_{i1}) \\ &= g^{-1}(X_i'\beta_{2|1^-}) \times (1 - g^{-1}(X_i'\beta_1)). \end{aligned} \tag{11}$$

Maximum likelihood estimates of the regression parameters are obtained by maximizing (10), with π_{i1} as in (8) and π_{i2} as in (11). As for the cumulative model, for this maximization, probabilities $\pi_{i1}, \ldots, \pi_{iJ}$ must lie between 0 and 1 and $\sum_{j=1}^{J} \pi_{ij}$ must be < 1 for all i. The logit link imposes constraints that $0 < \pi_{i1} < 1$ and $0 < \pi_{i2|1^-} < 1$ which, together, imply that $0 < \pi_{i2} < 1$. Thus, for the conditional model with $J = 2$, the optimization procedure

requires only that $\pi_{i1} + \pi_{i2} < 1$.

More generally, we can link covariates to the probability of testing positive in an interval given all previous tests were negative by setting

$$\pi_{ij} = \pi_{ij|(j-1)^-} \times (1 - \sum_{k=1}^{j-1} \pi_{ik}),$$

where

$$\pi_{ij|(j-1)^-} = g^{-1}(X_i' \beta_{j|(j-1)^-}). \tag{12}$$

3.3 Obtaining MLEs

In Sections 3.1 and 3.2, we derived the log-likelihood equations for the cumulative and conditional regression models. To obtain maximum likelihood estimates of the regression parameters, we maximize the log-likelihood equations using numerical optimization techniques. For the analyses presented here, numerical optimization was carried out using a quasi-Newton algorithm with non-linear constraints on the coefficients. The algorithm is an efficient modification of Powell's Variable Metric Constrained WatchDog algorithm, which is available in SAS PROC NLP (SAS Institute Inc., 2004a). Additional details regarding our implementation are available upon request.

4 Simulations

We performed simulations to assess the properties of the proposed regression estimators, first by comparing them to the more commonly used logistic regression approaches for performing adjusted analyses described in Section 1 and, second, by assessing how well they captured the effect of a treatment on in utero and delivery transmission given potential misclassification due to early breastfeeding transmission. We considered the case of two visit windows, corresponding to birth and 4 to 8 weeks. For each simulated dataset, we randomly generated a set of covariates for each observation that we used to simulate a subject's time of detectable infection. Next, we randomly generated a set of visit times for each observation. We determined each subject's observed endpoints by comparing his or her simulated time of detectable infection to his or her simulated visit times. We considered several scenarios, allowing for different treatment effects and different visit processes. Results for each scenario

are provided, based on 1000 datasets of 1500 observations each.

4.1 Simulation of Time of Detectable Infection

In this section, we describe how we simulated time of detectable infection, which was used together with information on visit timing (discussed below) to determine a simulated subject's sequence of test results. First, we simulated mode of transmission as either in utero, during delivery or early breastfeeding. In doing so, we assumed that timing of infection is subject to the effects of a binary and a continuous predictor. The binary predictor (X_1) acts as a treatment with potentially different effects in utero and in the peripartum period and was drawn from a Bernoulli distribution with probability 0.5. We simulated the continuous predictor (X_2) from a normal distribution with mean 4.3 and standard deviation 0.8 to mimic the observed distribution of log 10 viral load in the HPTN 024 data.

We simulated mode of transmission under two frameworks, one based on a cumulative regression model and the other based on a conditional regression model. We calculated the probabilities of in utero infection, π_{i1}, and perinatal infection, $\pi_{i1} + \pi_{i2}$, for subject i according to equations (8) and (9) with $\beta_1 = (\beta_{10}, \beta_{11}, \beta_{12})'$, $\beta_2 = (\beta_{20}, \beta_{21}, \beta_{22})'$, and $g(\cdot)$ taken to be the logit link. We calculated the probability of intrapartum infection, $\pi_{i2|1^-}$, for subject i according to equation (12) with $\beta_{2|1^-} = (\beta_{2|1^-,0}, \beta_{2|1^-,1}, \beta_{2|1^-,2})'$ and $g(\cdot)$ the logit link. We used a subject's probability of perinatal infection (along with his or her probability of in utero infection) to determine π_{i2} under the cumulative framework. We determined π_{i2} under the conditional framework using a subject's probability of intrapartum infection, according to equation (11). To simulate whether a subject became infected in utero, at birth, or neither, we used a multinomial distribution with probabilities π_{i1}, π_{i2}, and $1 - (\pi_{i1} + \pi_{i2})$.

We allowed for the imperfect sensitivity of the test by generating the time of detectable infection. This reflects the fact that an intrapartum transmission is unlikely to be detected at the birth visit. We generated each subject's time of detectable infection as number of days since birth. For subjects who became infected in utero, we assigned a time of detectable infection equal to zero days. For subjects who became infected during delivery, we generated time of detectable infection according to a uniform (0,14) distribution. An upper limit of 14 days was chosen to accommodate the lag time inherent in detecting HIV infection (Balasubramanian and Lagakos, 2001). We also allowed the simulations to reflect additional positive test results at the 4 to 8 week visit due to breastfeeding. For subjects infected neither in utero nor at birth, we generated time of detectable infection according

Table 2: Simulation of time of detectable infection for each treatment effect scenario

			Cumulative				Conditional	
			Treatment		Control		Treatment	Control
Treatment Effect	β_{11}	β_{21}=$\beta_{2\|1^-,1}$	π_{i1}	$\pi_{i1}+\pi_{i2}$	π_{i1}	$\pi_{i1}+\pi_{i2}$	$\pi_{i2\|1^-}$	$\pi_{i2\|1^-}$
TE 1	-0.55	-0.54	0.03	0.11	0.05	0.18	0.05	0.09
TE 2	-0.55	0.00	0.03	0.18	0.05	0.18	0.09	0.09
TE 3	-0.02	-0.38	0.05	0.13	0.05	0.18	0.06	0.09
TE 4	0.27	-0.27	0.07	0.14	0.05	0.18	0.07	0.09

to an exponential distribution. This added an average of 30 infections to the 0 to 8 week period for the cumulative model and an average of 32 infections to the 0 to 8 week period for the conditional model.

We simulated time of detectable infection under four treatment effect scenarios, denoted TE1 through TE4. Under the cumulative framework, we took (β_{10}, β_{12}, β_{20}, β_{22}) = (-4, 0.25, -2.6, 0.25) for each scenario while under the conditional framework, we took (β_{10}, β_{12}, $\beta_{2|1^-,0}$, $\beta_{2|1^-,2}$) = (-4, 0.25, -3.4, 0.25) for each scenario. The effects of treatment on in utero, perinatal, and intrapartum transmission, represented by β_{11}, β_{21}, and $\beta_{2|1^-,1}$ respectively, were allowed to vary across scenarios as described in Table 2. For all scenarios, we assumed that treatment had the same effect on perinatal transmission (cumulative model) as on intrapartum transmission (conditional model), that is, we took β_{21}=$\beta_{2|1^-,1}$. Table 2 also provides the probabilities of in utero, perinatal, and intrapartum transmission for the treatment and control groups for each treatment effect scenario when the continuous predictor is taken to be equal to its average value.

4.2 Simulation of Visit Process and Determination of Test Results

In simulating each subject's visit process, we considered two visit windows, birth and 4 to 8 weeks, and the corresponding periods birth, between birth and 4 to 8 weeks, 4 to 8 weeks, and after 4 to 8 weeks. We simulated whether or not a subject was tested during each of these periods using a binomial distribution according to three visit process scenarios, denoted VP1 through VP3 (Table 3). VP1 corresponds to an analysis that might be done at the end of a study, when all of the data that are expected have been collected. For VP2, we assumed that a smaller percentage of subjects are tested at the 4 to 8 week visit and after the 4 to 8 week visit. For VP3, we reduced these percentages even further, as well as the percentage of subjects tested at birth. VP2 and VP3 were designed to represent interim analyses.

We assigned time of visit as number of days since birth. Time of birth visit was generated

Table 3: Probability tested during/following each visit window for each visit process

Visit Process	Birth	Between birth and 4-8 weeks	4-8 weeks	After 4-8 weeks
VP 1	0.85	0.05	0.75[a]	0.80
VP 2	0.85	0.05	0.50	0.25
VP 3	0.50	0.05	0.25	0.10

[a]Probability 0.85 used if infection detected on day 0.

according to a multinomial distribution with probability 0.4 for days 0 and 1 and probability 0.04 for days 2 through 6. For visits between birth and 4 to 8 weeks, time of visit was assigned according to a multinomial distribution with probability 0.05 for all days. For visits during the 4 to 8 week visit window, time of visit was assigned according to a multinomial distribution with probability 0.04 for all days. For visits after 4 to 8 weeks, time of visit was assigned according to a multinomial distribution, with days 275 through 325 weighted more heavily (probability = 0.008) than other days in the period (probability = 0.002). Each subject was allowed at most one visit in each of the windows described.

Finally, we compared each subject's simulated time of detectable infection to his or her simulated visit times to generate the subject's vector of observed results (Y). In assessing infection after 4 to 8 weeks, we considered only simulated times of detectable infection before day 500.

4.3 Methods Evaluated

We estimated the effect of treatment using logistic regression models and the proposed cumulative (CM-CUM) and conditional (CM-COND) censored multinomial regression models. We considered two sets of logistic models: the first (L-CUM) modeled infection at birth and infection at 4 to 8 weeks among subjects for whom HIV status at birth and HIV status at 4 to 8 weeks could be determined and the second (L-COND) modeled infection at birth and infection at 4 to 8 weeks among subjects known to the HIV negative at birth. The logistic models were chosen to represent those used in the analysis of PMTCT of HIV. Cox PH models, although often used, do not specifically address treatment effects on A1-A3 but instead estimate the average treatment effect over the observation period; therefore, we did not assess them in our simulations.

For the cumulative and conditional approaches, we compared the effect of treatment as obtained from the regression model to the "true" effect of treatment according to which the data were generated. We determined bias, mean squared error (MSE), 95% coverage

probability (CP), and power for each estimator, averaging across datasets for a given set of simulations. Because we allowed for imperfect sensitivity and early breastfeeding transmission, we would not expect to see zero bias in the estimates from our simulations. Our design, however, allows us to judge how well a data analysis estimates the associations in terms of how they are interpreted (as in utero, intrapartum, etc.).

In carrying out numerical optimization, we chose the convergence criteria for the proposed cumulative and conditional regression models to coincide with the convergence criteria for the logistic regression models in SAS PROC LOGISTIC (SAS Institute Inc., 2004b).

4.4 Simulation Results

Table 4 provides simulation results for the following combinations of treatment effect and visit process: TE1/VP1, TE2/VP1, TE3/VP1, TE4/VP1, TE4/VP2, and TE4/VP3. These combinations allow us to assess the impact of treatment effect on estimator performance for a given visit process as well as the impact of visit process on estimator performance for a given treatment effect. Table 4 consists of six subtables each corresponding to one of the above TE/VP scenarios. Results for the estimators of treatment effect at 4 to 8 weeks (cumulative models) and the estimators of treatment effect at 4 to 8 weeks among infants with a negative test result at birth (conditional models) are combined under the heading "4-8 weeks." Models are denoted using the abbreviations provided in Section 4.3.

We find that the CM-CUM model performs comparably to, or better than, the L-CUM model across all performance measures, for all but the TE1/VP1 scenario. Here, the birth estimate obtained from the L-CUM model is less biased than the birth estimate obtained from the CM-CUM model, although the bias is small in both cases (0.013 versus 0.027, respectively). In addition, power at birth is slightly higher for the L-CUM model than for the CM-CUM model. For TE1/VP1, we assumed that treatment reduces the odds of both in utero and perinatal transmission, and that the effect of treatment on the two endpoints is roughly the same.

Across all scenarios, MSE is consistently lower (albeit only slightly in most cases) for the CM-CUM model than for its logistic counterpart. The CPs for the competing cumulative models are similar while power at 4 to 8 weeks is higher for the CM-CUM model than for the L-CUM model for all scenarios where power was assessed. In general, we observe that power for the cumulative models is higher at 4 to 8 weeks than at birth. This is not surprising given the smaller probability of infection at birth as well as the fact that the simulations

Table 4: Simulation results for selected treatment effect/visit process scenarios

	Bias		MSE		CP		Power	
Scenario	Birth	4-8 weeks	Birth	4-8 weeks	Birth	4-8 weeks	Birth	4-8 weeks
TE1/VP1								
L-CUM	0.013	0.045	0.069	0.030	0.948	0.943	0.518	0.845
CM-CUM	0.027	0.048	0.068	0.027	0.948	0.939	0.499	0.882
L-COND	-0.015	0.107	0.077	0.077	0.959	0.914	0.539	0.439
CM-COND	-0.007	0.106	0.072	0.072	0.956	0.901	0.541	0.461
TE2/VP1								
L-CUM	0.173	-0.020	0.096	0.022	0.879	0.952	0.309	–
CM-CUM	0.168	-0.009	0.094	0.020	0.873	0.952	0.339	–
L-COND	0.098	0.004	0.083	0.050	0.938	0.956	0.398	–
CM-COND	0.101	0.006	0.079	0.046	0.941	0.952	0.405	–
TE3/VP1								
L-CUM	-0.078	0.055	0.058	0.029	0.952	0.931	0.057	0.517
CM-CUM	-0.064	0.049	0.054	0.025	0.954	0.933	0.056	0.605
L-COND	-0.050	0.064	0.062	0.059	0.946	0.941	0.062	0.263
CM-COND	-0.042	0.063	0.058	0.055	0.944	0.944	0.057	0.264
TE4/VP1								
L-CUM	-0.123	0.059	0.065	0.028	0.921	0.936	0.096	0.271
CM-CUM	-0.105	0.049	0.057	0.023	0.931	0.942	0.111	0.307
L-COND	-0.043	0.034	0.058	0.060	0.953	0.937	0.160	0.173
CM-COND	-0.038	0.033	0.055	0.056	0.954	0.936	0.186	0.187
TE4/VP2								
L-CUM	-0.110	0.116	0.064	0.045	0.927	0.898	0.103	0.138
CM-CUM	-0.096	0.043	0.059	0.032	0.927	0.940	0.128	0.261
L-COND	-0.073	0.041	0.061	0.084	0.941	0.949	0.138	0.123
CM-COND	-0.066	0.042	0.054	0.078	0.946	0.945	0.146	0.134
TE4/VP3								
L-CUM	-0.124	0.192	0.113	0.105	0.923	0.890	0.068	0.064
CM-CUM	-0.088	0.017	0.097	0.058	0.939	0.954	0.096	0.193
L-COND	-0.054	-0.009	0.099	0.292	0.944	0.967	0.095	0.077
CM-COND	-0.040	0.002	0.087	0.211	0.946	0.964	0.102	0.086

were designed based on HPTN 024, which was powered to detect a difference in cumulative transmission rates at 4 to 8 weeks.

The CM-CUM model performs most impressively for the TE4/VP1, TE4/VP2, and TE4/VP3 scenarios, where it consistently outperforms the L-CUM model in terms of bias, MSE, CP, and power at birth and at 4 to 8 weeks. In addition, as the amount of missing test result data increases, the improvement offered by the CM-CUM model increases. We see this, for example, in the bias estimates at 4 to 8 weeks. Recall that VP2 and VP3 were designed to represent interim analyses, as described in Section 4.2. For TE4/VP1, the bias for the L-CUM model is 1.2 times the bias for the CM-CUM model. For TE4/VP2, the bias for the L-CUM model is 2.7 times the bias for the CM-CUM model. For TE4/VP3, which assumes the highest percentage of missing data of the visit processes considered, the bias for the L-CUM model is 11.3 times the bias for the CM-CUM model. We note that, of the treatment effects considered, TE4 is most similar to that observed in HPTN 024.

On the whole, the CM-COND model performs comparably to the L-COND model. MSE is consistently lower, although again only slightly, for the CM-COND model than for the L-COND model. While power tends to be low for the conditional models, power at birth and at 4 to 8 weeks is slightly higher for the CM-COND model than for the L-COND model for the TE4/VP1, TE4/VP2, and TE4/VP3 scenarios.

In addition to comparing the proposed regression estimators to the more commonly used logistic regression approaches, our simulations assessed the bias associated with early breastfeeding transmission when estimating the treatment effect. Breastfeeding transmission appears to have had little effect on 4 to 8 week estimates of treatment effect, with the magnitude of most bias estimates at 4 to 8 weeks (for the CM-CUM and CM-COND models) less than 0.05. We would expect early breastfeeding contamination to bias the treatment effect to the null, by adding infections to both the treatment and control arms. This is consistent with our findings of positive bias at 4 to 8 weeks for TE1, TE3, and TE4, where we assumed the effect of treatment on birth transmission is negative.

5 Example

In this section, we analyzed data from HPTN 024, a multi-site double-blinded placebo controlled trial of antiobiotics to prevent chorioamnionitis and, therefore, perinatal transmission of HIV. The trial enrolled pregnant, HIV positive women receiving care in hospitals and clinics in Malawi, Tanzania, and Zambia. Women were randomized to receive either treatment or

placebo. Treatment consisted of two courses of antibiotics, with the first course administered at enrollment (20 to 24 weeks gestation) and the second at the onset of contractions and/or premature rupture of membranes. All women and their liveborn infants were offered single dose nevirapine per the HIV Network for Prevention Trials (HIVNET) 012 protocol (Guay et al., 1999). Women were followed during their pregnancies, and their infants were followed postnatally. Visit windows for determining in utero and delivery/early postnatal transmission in this breastfeeding population were 0 to 48 hours and 4 to 6 weeks, respectively. Because over half of the visits scheduled to occur between 4 and 6 weeks actually took place between 6 and 8 weeks, we extended the second visit window to 4 to 8 weeks for analysis purposes. We also extended the birth visit window to 0 to 7 days.

Recruitment began in July 2001. The trial was monitored for safety and efficacy by the NIAID Vaccine and Prevention Data and Safety Monitoring Board (DSMB). In February 2003, the DSMB reviewed trial progress in a scheduled interim analysis and concluded that, while statistical evidence neither established benefit nor harm, the available evidence ruled out targetted levels of benefit. The DSMB recommended that HPTN 024 stop recruitment and continue follow-up of enrolled women and infants. Further randomization and distribution of study drugs was halted at all clinical sites in early March 2003. Additional details regarding the 024 study are provided by Taha et al. (2006).

For this example, we defined the treatment and control groups as infants born to mothers randomized to antibiotics who delivered prior to termination of the study drug and infants born to mothers randomized to placebo or to mothers randomized to antibiotics who delivered after termination of the study drug, respectively. Additional covariates of interest were log maternal viral load, maternal CD4 count, and infant gender. In the birth model, we adjusted for mother's use of nevirapine and, in the 4 to 8 week model, for mother's and infant's use of nevirapine. To account for unmeasured differences between hospitals and clinics, we included study site in both models.

Of 2052 firstborn infants born alive to HIV positive mothers, 1758 had complete data with respect to the covariates of interest. Of these, 1696 had a test result at some point during follow-up and were included in the analysis of HIV infection. 1739 had a test result or are known to have died during follow-up and were, thus, included in the analysis of HIV infection or death. Descriptive statistics for the 1739 subjects included in the analysis of HIV infection or death are provided in Table 5. Figure 2 provides the complete testing profile for the 1758 subjects with complete covariate data, according to treatment group.

We used the proposed regression methods to analyze the outcomes infection and infection

Table 5: Descriptive statistics for 1739 subjects included in analysis of HIV infection or death

Covariate	Treatment		Control	
	Mean/N	SD/%	Mean/N	SD/%
Maternal viral load (1 log10 unit)	4.338	0.836	4.242	0.817
CD4 count (100 units)	3.697	2.067	3.796	2.234
Female	296	47%	566	51%
Mother nevirapine	603	95%	1066	96%
Mother and infant nevirapine	558	88%	1001	90%

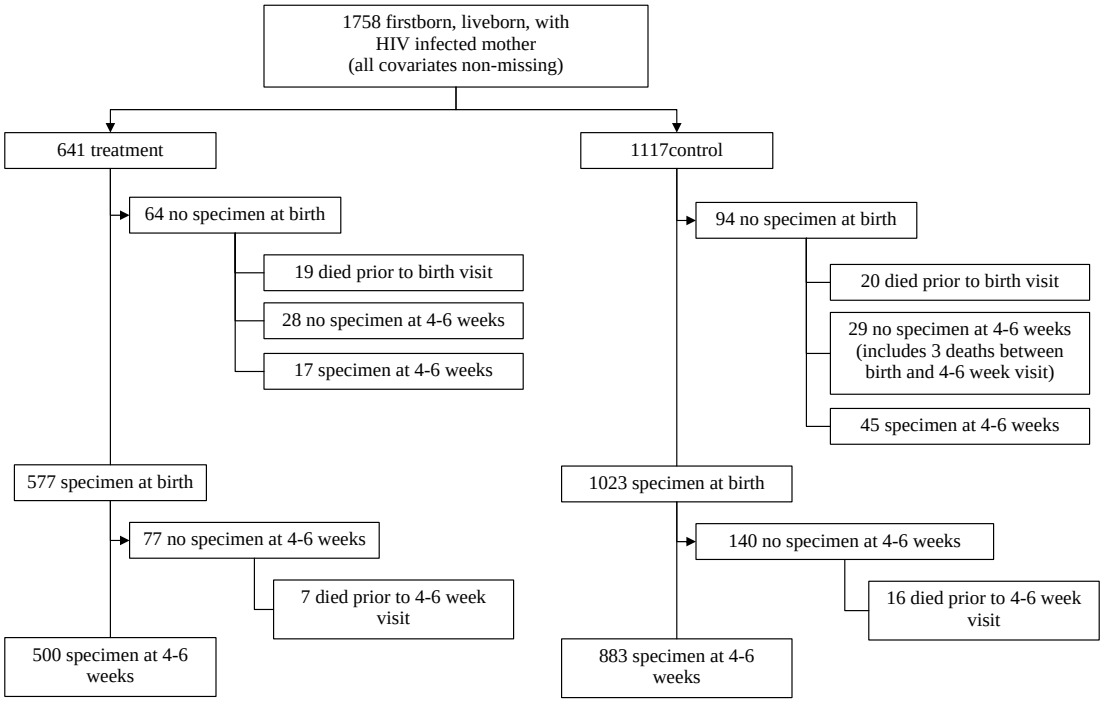

Figure 2: HPTN 024 testing profile.

Table 6: Adjusted odds ratios at birth, 4 to 8 weeks

Risk factor	Birth		4-8 Weeks		4-8 Weeks |- Birth	
	OR	95% CI	OR	95% CI	OR	95% CI
Infection endpoint[a]						
Treatment vs. control	0.691	[0.469, 1.017]	0.987	[0.736, 1.323]	1.404	[0.928, 2.123]
Maternal viral load (1 log10 unit)	2.552	[1.895, 3.437]	2.820	[2.221, 3.579]	2.873	[2.023, 4.082]
CD4 count (100 units)	1.037	[0.939, 1.145]	1.099	[1.015, 1.190]	1.172	[1.033, 1.331]
Female	0.938	[0.658, 1.339]	1.011	[0.764, 1.338]	1.101	[0.733, 1.654]
Mother nevirapine	0.549	[0.248, 1.215]				
Mother and infant nevirapine			0.628	[0.387, 1.019]	0.562	[0.262, 1.206]
Infection/death endpoint[b]						
Treatment vs. control	0.873	[0.632, 1.205]	1.025	[0.789, 1.332]	1.245	[0.850, 1.825]
Maternal viral load (1 log10 unit)	2.299	[1.786, 2.959]	2.568	[2.086, 3.161]	2.597	[1.895, 3.558]
CD4 count (100 units)	1.043	[0.959, 1.135]	1.098	[1.024, 1.178]	1.175	[1.048, 1.318]
Female	0.883	[0.650, 1.201]	1.066	[0.829, 1.370]	1.336	[0.916, 1.949]
Mother nevirapine	0.221	[0.128, 0.381]				
Mother and infant nevirapine			0.452	[0.302, 0.678]	0.546	[0.277, 1.075]

[a]Analysis based on 612 treatment and 1084 control.
[b]Analysis based on 632 treatment and 1107 control.

or death. We estimated the cumulative odds of infection at 4 to 8 weeks using the cumulative model of Section 3.1 and the odds of infection at 4 to 8 weeks among infants with a negative test result at birth using the conditional model of Section 3.2. Results are provided in Table 6.

We found that treatment does not significantly reduce HIV infection at birth, HIV infection at 4 to 8 weeks, or HIV infection at 4 to 8 weeks among subjects who test negative at birth. These findings are consistent with those presented by Taha et al., where treatment was defined based on intent-to-treat and the analysis was limited to women who delivered prior to study termination. The trend of estimates here is also consistent with that in Taha et al., suggesting that treatment decreases the odds of infection at birth (OR 0.691, 95% CI [0.469, 1.017]) while increasing the odds of infection at 4 to 8 weeks among those who test negative at birth (OR 1.404, 95% CI [0.928, 2.123]). When we defined first positive test or death at or before a given visit as the endpoint, we observed the same trend in the estimates of treatment effect as for the infection endpoint, although the trend was slightly weaker in this case.

Birth estimates in Table 6 were obtained from the conditional regression model. Birth estimates obtained from the cumulative model were comparable for all covariates except

mother's use of nevirapine in the infection or death model. Using the cumulative regression model, we obtained an odds ratio estimate of 0.326 (95% CI [0.203, 0.524]), which is slightly higher than the odds ratio estimate obtained from the conditional model. The difference is likely a consequence of the fact that almost all mothers received nevirapine (95% of mothers in the treatment group and 96% of mothers in the control group).

6 Discussion

Many statistical techniques are available for estimating PMTCT of HIV while adjusting for covariates. Among the more commonly used are logistic regression models and Cox proportional hazards models. While these methods are relatively straightforward to implement, they do not easily accommodate missed or unscheduled visits while allowing for a time-varying treatment effect. Cox models can be modified to allow the effect of treatment to depend upon time but do not fully solve the problem of how to handle missed or unscheduled visits. Interval censored models, which use a subject's time to last negative test and time to first positive test to form an interval around his or her (unknown) time of infection, may better accommodate the missing data, but software is not generally available for regression with interval censored data unless we are willing to make parametric assumptions about the distribution of the event times.

Recently, Bang and Spiegelman (2004) proposed a likelihood approach for a dichotomous outcome to estimate mother to child transmission when infection status is missing for some infants due to fetal loss. However, this approach does not address all three endpoints of interest or missing data due to incomplete follow-up. Balasubramanian and Lagakos (2001) provide methods for estimating the distribution of the timing of in utero and peripartum transmission in a non-breastfeeding population that accounts for the imperfect sensitivity of the HIV assay. Because we examine a breastfeeding population and are interested in categorizing infection timing, this approach would not be suitable. Little and Rubin (2002, pages 169, 170) describe an approach for maximum likelihood estimation in a multinomial setting based on the Expectation Maximization (EM) algorithm. With this approach, data are categorized according to infection timing and missing values are imputed through an iterative procedure. While it solves the problem of incomplete data, the approach addresses only single sample estimation.

Here, we propose a censored multinomial approach for estimating PMTCT that accommodates missing test result data, regression on the three outcomes of interest, and time-varying

treatment effects. Through simulation, we investigated the performance of the estimators obtained from the more commonly used logistic regression approaches and compared them to the proposed estimators. We also looked at the robustness of the estimators to contamination of the endpoint due to early breastfeeding transmission. We found that both the proposed cumulative model and conditional models performed well when compared to their logistic counterparts. Performance of the proposed cumulative model was particularly strong under scenarios designed to represent interim analyses. Power for the proposed models was consistently higher at 4 to 8 weeks, which is to be expected given that the logistic models used only data for subjects whose endpoints were non-missing or could be imputed based on subsequent negative tests.

The censored multinomial regression approach is not without limitations. Both the proposed cumulative and the proposed conditional models impose non-linear constraints on the coefficients, which can complicate interpretation of the estimates if maximization of the likelihood occurs on the boundary of the parameter space. In the case of the conditional model, however, only a single constraint is imposed, which is no more than would be imposed for a general multinomial model (Agresti, 2002, page 21). In numerous simulations (beyond those presented here), we saw no evidence of bias due to maximization on the boundary.

Our approach relies on the assumption that missingness is non-informative and, thus, may be more appropriate for some endpoints (infection/death) than for others (infection). While valid for breastfeeding populations, our approach does not allow us to separate intrapartum transmission from early transmission due to breastfeeding. In addition, our approach assumes that infants are at risk for breastfeeding transmission throughout the postnatal period and, thus, does not allow for the possibility of weaning during this period. Our approach does not account for misclassification due to the imperfect sensitivity of testing. However, as we demonstrated in our simulations, the bias in the 4 to 8 week estimates for the proposed models is quite small, suggesting that the impact of misclassification and contamination due to breastfeeding on our estimates of treatment effect is minimal. Finally, our approach does not provide a mechanism for estimating relative risks, which are often of interest in PMTCT trials. Our model could easily be adapted to this context through use of a log link or a complementary log-log link. In this case, valid comparison models would include relative risk regression models and time-dependent Cox models. To our knowledge, these models have not been used in the PMTCT setting but would be worthy of further exploration.

Here, we have studied the problem of estimating the effect of treatment on perinatal mother to child transmission of HIV when outcome data are incomplete. We provide methods

that give consistent and asymptotically normal estimators using maximum likelihood and are easily programmed using standard statistical software. Through simulation, we have shown that the proposed methods outperform standard logistic regression methods in terms of bias, mean squared error, coverage probability, and power under a range of treatment effect and visit process scenarios. While demonstrated for HIV transmission, the approach has broader applicability to problems of estimating treatment effects on disease incidence when data are collected at multiple visits spaced in time and outcome for some subjects, at some time points, is unknown.

References

Agresti, A. (2002). *Categorical Data Analysis*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2nd edition.

Balasubramanian, R. and Lagakos, S. W. (2001). Estimation of the timing of perinatal transmission of HIV. *Biometrics* **57,** 1048–1058.

Bang, H. and Spiegelman, D. (2004). Estimating treatment effects in studies of perinatal transmission of HIV. *Biostatistics* **5,** 31–43.

Dabis, F., Msellati, P., Meda, N., Welffens-Ekra, C., You, B., Manigart, O., Leroy, V., Simonon, A., Cartoux, M., Combe, P., Ouangre, A., Ramon, R., Ky-Zerbo, O., Montcho, C., Salamon, R., Rouzioux, C., de Perre, P. V., and Mandelbrot, L. (1999). 6-month efficacy, tolerance, and acceptability of a short regimen of oral zidovudine to reduce vertical transmission of HIV in breastfed children in Cote D'Ivoire and Burkina Faso: a double-blind placebo-controlled multicentre trial. *Lancet* **353,** 786–792.

Dorenbaum, A., Cunningham, C. K., Gelber, R. D., Culnane, M., Mofenson, L., Britto, P., Rekacewicz, C., Newell, M. L., Delfraissy, J. F., Cunningham-Schrader, B., Mirochnick, M., and Sullivan, J. L. (2002). Two-dose intrapartum/newborn nevirapine and standard antiretroviral therapy to reduce perinatal HIV transmission-A randomized trial. *Journal of the American Medical Association* **288,** 189–198.

Fawzi, W. W., Msamanga, G., Hunter, D., Urassa, E., Renjifo, B., Mwakagile, D., Hertzmark, E., Coley, J., Garland, M., Kapiga, S., Antelman, G., Essex, M., and Spiegelman, D. (2000). Randomized trial of vitamin supplements in relation to vertical transmission of HIV-1 in Tanzania. *Journal of Acquired Immune Deficiency Syndromes* **23,** 246–254.

Guay, L. A., Musoke, P., Fleming, T., Bagenda, D., Allen, M., Nakabiito, C., Sherman, J., Bakaki, P., Ducar, C., Deseyve, M., Emel, L., Mirochnick, M., Fowler, M. G., Mofenson, L., Miotti, P., Dransfield, K., Bray, D., Mmiro, F., and Jackson, J. B. (1999). Intrapartum and neonatal single-dose nevirapine compared with zidovudine for prevention of mother-to-child transmission of HIV-1 in Kampala, Uganda: HIVNET 012 randomised trial. *Lancet* **354,** 795–802.

Little, R. J. A. and Rubin, D. B. (2002). *Statistical Analysis with Missing Data*. John Wiley & Sons, Inc., Hoboken, New Jersey, 2nd edition.

Moodley, D., Moodley, J., Coovadia, H., Gray, G., McIntyre, J., Hofmyer, J., Nikodem, C., Hall, D., Gigliotti, M., Robinson, P., Boshoff, L., and Sullivan, J. L. (2003). A multicenter randomized controlled trial of nevirapine versus a combination of zidovudine and lamivudine to reduce intrapartum and early postpartum mother-to-child transmission of Human Immunodeficiency Virus type 1. *Journal of Infectious Diseases* **187,** 725–735.

SAS Institute Inc. (2004a). *SAS/OR 9.1.2 User's Guide: Mathematical Programming*. SAS Institute Inc., Cary, North Carolina.

SAS Institute Inc. (2004b). *SAS/STAT 9.1 User's Guide*. SAS Institute Inc., Cary, North Carolina.

Shaffer, N., Chuachoowong, R., Mock, P. A., Bhadrakom, C., Siriwasin, W., Young, N. L., Chotpitayasunondh, T., Chearskul, S., Roongpisuthipong, A., Chinayon, P., Karon, J., Mastro, T. D., and Simonds, R. J. (1999). Short-course zidovudine for perinatal HIV-1 transmission in Bangkok, Thailand: a randomised controlled trial. *Lancet* **353,** 773–780.

Taha, T. E., Brown, E. R., Hoffman, I. F., Fawzi, W., Read, J. S., Sinkala, M., Martinson, F. E. A., Kafulafula, G., Msamanga, G., Emel, L., Adeniyi-Jones, S., Goldenberg, R., and the HPTN024 Team (2006). A phase III clinical trial of antiobiotics to reduce chorioamnionitis-related perinatal HIV-1 transmission. *AIDS* **20,** 1313–1321.

Wiktor, S. Z., Ekpini, E., Karon, J. M., Nkengasong, J., Maurice, C., Severin, S. T., Roels, T. H., Kouassi, M. K., Lackritz, E. M., Coulibaly, I. M., and Greenberg, A. E. (1999). Short-course oral zidovudine for prevention of mother-to-child transmission of HIV-1 in Abidjan, Cote D'Ivoire: a randomised trial. *Lancet* **353,** 781–785.