Modeling Longitudinal Biomarker Data With Multiple Assays Which Have Different Known Detection Limits

Paul S. Albert
Biometric Research Branch
Division of Cancer Treatment and Diagnosis
National Cancer Institute

Email: albertp@mail.nih.gov

February 27, 2007

Summary

Assays to measure biomarkers are commonly subject to large amounts of measurement error and known detection limits. Studies with longitudinal biomarker measurements may use multiple assays in assessing outcome. I propose an approach for jointly modeling repeated measures of multiple assays when these assays are subject to measurement error and known lower detection limits. A commonly used approach is to perform an initial assay with a larger lower detection limit on all repeated samples, followed by only performing a second more expensive assay with a lower minimum level of detection when the initial assay value is below its lower limit of detection. I show how simply replacing the initial assay measurement with the second assay measurement may be a biased approach and investigate the performance of the proposed joint model in this situation. Additionally, I compare the performance of the joint model with an approach which only uses the initial assay measurements in analysis. Further, I consider alternative designs to only performing the second assay when the initial assay measurement is below its lower detection limit. For example, I show the advantages of performing the second assay at random without regard to the initial assay measurement over a design in which the second assay is only performed when the initial assay is below its lower limit of detection. The methodology is illustrated with a recent study examining the use of a vaccine in treating macaques with simian immunodeficiency virus.

Key Words: Below the level of detection, Measurement error, Linear mixed models, Random effects models, Repeated measures data, Simian immunodeficiency virus.

1 Introduction

Biomarkers are commonly used as the endpoints in longitudinal clinical or laboratory studies. The assays for these biomarkers are often subject to large amounts of measurement error and known detection limits. More sensitive assays, which are more expensive, usually have smaller amounts of measurement error with lower known minimum detection limits. One commonly used design strategy is to perform an initial inexpensive assay on all longitudinally collected samples, followed by a second more expensive assay with a lower minimum level of

detection on the subset of measurements for which the inexpensive assay is below its lower limit of detection. The motivating example for this methodological work is a recent study assessing the effects of a new therapeutic vaccine on maintaining reduced levels of viral load after anti-retroviral therapy (ART) is terminated in simian immunodeficiency virus (SIV) induced macaques (Von Gegerfelt et al., 2007). A group of 12 SIVmac251-infected macaques were treated with ART, and given a therapeutic vaccine at the end of ART. Viral load (RNA copy number) was measured from longitudinal plasma samples during an SIV chronic phase, during the period of ART and vaccination treatment, and after ART was terminated (referred to as the release period). Of interest was assessing whether viral load among vaccinated animals remained low after ART was stopped (release period), or alternatively, showed a quick rebound to levels of viral load seen during the chronic SIV phase. A concurrent control group of 11 SIVmac251-infected macaques was also followed in a similar manner with the exception that no vaccine was given at the end of the ART period. Of interest is comparing the difference between viral load during the release period and the chronic SIV period in vaccinated and unvaccinated macaques.

For both vaccinated and unvaccinated animals, samples of plasma were repeatedly collected at irregularly spaced intervals before and after ART. A median of 45 plasma samples (range: 26 to 61) were taken on each animal during the SIV chronic, ART, and release periods. Initially, viral load was assessed from samples at each time point using a commonly used assay with a known lower limit of detection of 20,000 copies per ml. Viral load was re-assessed with a second assay when (i) the measurement of the initial assay was below its lower limit of detection, and when (ii) there was sufficient blood sample to perform the second assay. In most cases, this second assay had a lower limit of detection of 4,000 copies per ml. Figure 1 shows the log (base 10) transformed viral load profile of the first 4 animals in the vaccinated cohort. Values from the initial assay are denoted as 1's with the top dotted horizontal line being the lower limit of detection for this assay. Values from the second assay are denoted as 2's with the dashed-dotted horizontal line showing the lower limit of detection

for this more sensitive assay. The bottom solid line delineates the SIV chronic phase while the bottom dashed line shows the release period; the area between these two horizonal lines delineates the ART period. The figures visually suggest that for the vaccinated animal, low levels of viral load while being treated with ART are at least partially maintained in the release period. Of interest is on determining whether this is indeed a real finding, or whether it is consistent with the statistical variation in the process. A joint modeling approach will be used to address this question. This modeling approach will then be compared with simpler approaches in which (i) only data on the initial assay is used and (ii) the initial assay is replaced by the second assay when the results of the second assay are available.

There is a sizable literature on modeling longitudinal data from a single known detection limit. Hughes (1999) presented a flexible linear mixed models for a longitudinal outcome with known detection limits. Lyles et al. (2000) presented a linear mixed model for longitudinal data with known detection limits subject to informative dropout. Theibaut et al. (2005) proposed a joint model of bivariate longitudinal data where one assay was subject to a known lower detection limit. Moulton and Halsey (1995) and Moulton et al. (2002) proposed mixture models which incorporated a higher proportion of assay values below the lower detection limit than what would be expected based on a censored Gaussian distribution. There has been little work in simultaneously modeling multiple assays with different known detection limits. In this paper, new methodology is proposed for jointly modeling longitudinal data from two assays when both assays are subject to measurement error as well as different known lower detection limits. Further, I investigate the design issue of whether it is best to perform the second more sensitive assay only when the initial assay is below its lower limit of detection (as was done in the SIV/vaccine study) or, alternatively, whether it is better to measure the second assay on a comparable number of samples without regard to the initial assay values. In Section 2, I present the modeling strategy. Section 3 presents an analysis of the SIV/vaccine dataset using the joint modeling approach along with simpler analysis strategies in which only the initial assay is used and where the initial assay is re-

placed by a second more sensitive assay, when the second assay is performed. In Section 4, I examine properties of the joint modeling approach with simulations corresponding to the example as well as some additional targeted simulations. I also compare the joint modeling approach with the simpler approaches in terms of bias and efficiency. A discussion follows in Section 5.

2 Modeling Framework

Initially, I present the model in the general framework of Laird and Ware (1982),

$$Y_{ij} = \underline{\boldsymbol{X}}_{ij}\underline{\boldsymbol{\beta}} + \underline{\boldsymbol{Z}}_{ij}\underline{\boldsymbol{b}}_i + \epsilon_{ij}, \tag{1}$$

where Y_{ij} is the jth measurement for the ith individual, $\underline{\boldsymbol{\beta}}$ is a vector of fixed effects, and $\underline{\boldsymbol{b}}_i$ is a vector of random effects for the ith subject. Let n_i be the number of longitudinal measurements on the ith subject ($j = 1, 2, ..., n_i$) and I be the number of individuals ($i = 1, 2, ..., I$). I denote, $\underline{\boldsymbol{X}}_{ij}$ and $\underline{\boldsymbol{Z}}_{ij}$ as design matrices associated with the fixed and random effects, respectively. Further, I assume that $\underline{\boldsymbol{b}}_i$ and ϵ_{ij} are statistically independent and that $\underline{\boldsymbol{b}}_i \sim N(\underline{\mathbf{0}}, \underline{\boldsymbol{D}})$ and $\epsilon_{ij} \sim N(0, \sigma^2)$, where $\underline{\boldsymbol{D}}$ is the variance matrix for the random effects and σ^2 is a scalar variance for the error term (i.e, the residual variance). The measurement Y_{ij} is not observed directly, but rather is measured using either one or two assays which are subject to measurement error and lower detection limits. Let A_{1ij} and A_{2ij} be two assay measurements which are attempting to measure Y_{ij}. I assume that the two assays have measurement errors δ_{1ij} and δ_{2ij}, which are assumed independent from each other as well as from $\underline{\boldsymbol{b}}_i$ and ϵ_{ij}, and are assumed to follow Gaussian distributions with mean zero and variances σ_1^2 and σ_2^2. Thus, the distribution of the difference in the two assays is

$$A_{2ij} - A_{1ij} \sim \mathrm{N}(\Delta, \sigma_1^2 + \sigma_2^2), \tag{2}$$

where the parameter Δ measures a systematic difference in the two assays that may be due to lack of calibration between the two assays. Equation (2) assumes that the difference between

A_{2ij} and A_{1ij} is constant and does not depend on the values of the two assays. Effectively, this implies that the slope between the two measurements is 1. Since in the SIV/vaccine study, we measure the second assay only when the first is below detectable limits, and therefore there are no cases when both assays are simultaneously directly measured, it is very difficult to estimate more than a shift in the means between the two assays. The two assay values are subject to different lower limits of detection C_1 and C_2. Specifically, we observe $A^*_{1ij} = A_{1ij}$ when $A_{1ij} > C_1$ and $A^*_{1ij} = C_1$ when $A_{1ij} \leq C_1$. Further, we observe $A^*_{2ij} = A_{2ij}$ when $A_{2ij} > C_2$ and $A^*_{2ij} = C_2$ when $A_{2ij} \leq C_2$.

A special case of this formulation can be used for analyzing the SIV longitudinal data. I denote Y_{ij} as the *true* log-transformed viral load at the jth time point for the ith animal in either the vaccinated or control group. I consider the simple linear mixed model (which is a special case of (1)) for analysis,

$$Y_{ij} = \beta_0 + \beta_1 D_{ij} + \beta_2 R_{ij} + b_i + \epsilon_{ij} \tag{3}$$

where $b_i \sim N(0, \sigma_b^2)$ and $\epsilon_{ij} \sim N(0, \sigma^2)$. Further, D_{ij} is an indicator which is equal to one when the ith animal at the time of the jth observation is on ART and R_{ij} is an indicator which is equal to one when the ith animal at the time of the jth measurement is in the release period after ART is terminated. For the SIV/vaccine study, the inexpensive assay A_{1ij} measured from samples at all time points has measurement error δ_{1ij} and a known lower detection limit of 20,000 copies per ml on the original scale or 4.8 on the log-base 10 scale ($C_1 = 4.8$). The second assay, which has the lower detection limit of only 4,000 copies per ml or 3.6 on the log-base 10 scale ($C_2 = 3.6$) was only performed when the initial assay was below the lower limit of detection ($C_1 = 4.8$) and there was sufficient additional plasma sample to perform the assay.

Without replication of an assay at a particular time point, it is difficult or impossible to uniquely identify measurement error from residual variance without making unverifiable additivity assumptions about the two sources of variation. However, we are able to reliably iden-

tify, $\sigma_1^{2^*} = \sigma_1^2+\sigma^2$, $\sigma_2^{2^*} = \sigma_2^2+\sigma^2$, and $\rho^* = corr(\delta_{1ij}+\epsilon_{ij}, \delta_{2ij}+\epsilon_{ij}) = \sigma^2/(\sqrt{\sigma_1^2+\sigma^2}\sqrt{\sigma_2^2+\sigma^2}$. Note that $\sigma^2 = \rho^*\sigma_1^*\sigma_2^*$, $\sigma_1^2 = \sigma_1^{2^*} - \rho^*\sigma_1^*\sigma_2^*$ and $\sigma_2^2 = \sigma_2^{2^*} - \rho^*\sigma_1^*\sigma_2^*$.

Denote $\mu_{ij} = \underline{\boldsymbol{X}}_{ij}\underline{\boldsymbol{\beta}} + \underline{\boldsymbol{Z}}_{ij}\underline{\boldsymbol{b}}_i$ and let S_{1ij} and S_{2ij} be indicators of whether the initial and second assays are performed for the jth measurement on the ith subject, respectively. In our formulation, the initial assay is performed at all measurements (i.e., $S_{1ij} = 1$ for all i and j). For the general model (1 and 2), maximum-likelihood estimation can be performed by maximizing the log-likelihood $logL = \prod_{i=1}^{I} \log L_i$, where

$$
\begin{aligned}
L_i \;=\; & \int_{\underline{\boldsymbol{b}}} \Big[\prod_{j=1}^{n_i} \Big\{ f_{(\mu_{ij},\mu_{ij}+\Delta),(\sigma_1^{2^*},\sigma_2^{2^*},\rho^*)}(A_{1ij}^*, A_{2ij}^*|\underline{\boldsymbol{b}})^{(A_{1ij}^*>C_1)(A_{2ij}^*>C_2)S_{2ij}} \\
& \times F_{(\mu_{ij},\mu_{ij}+\Delta),(\sigma_1^{2^*},\sigma_2^{2^*},\rho^*)}(A_{1ij}^*, A_{2ij}^*|\underline{\boldsymbol{b}})^{(A_{1ij}^*=C_1)(A_{2ij}^*=C_2)S_{2ij}} \\
& \times \Big\{ F_{(\mu_{ij}+\rho^*\frac{\sigma_1^*}{\sigma_2^*}(A_{2ij}^*-\mu_{ij}-\Delta),\sigma_1^{2^*}(1-\rho^{*2})}(C_1|\underline{\boldsymbol{b}}) f_{\mu_{ij}+\Delta,\sigma_2^{2^*}}(A_{2ij}^*|\underline{\boldsymbol{b}}) \Big\}^{(A_{1ij}^*=C_1)(A_{2ij}^*>C_2)S_{2ij}} \\
& \times \Big\{ F_{\mu_{ij}+\Delta+\rho^*\frac{\sigma_2^*}{\sigma_1^*}(A_{1ij}^*-\mu_{ij}),\sigma_2^{2^*}(1-\rho^{*2})}(C_2|\underline{\boldsymbol{b}}) f_{\mu_{ij},\sigma_1^{2^*}}(A_{1ij}^*|\underline{\boldsymbol{b}}) \Big\}^{(A_{1ij}^*>C_1)(A_{2ij}^*=C_2)S_{2ij}} \\
& \times f_{\mu_{ij},\sigma_1^{2^*}}(A_{1ij}^*|\underline{\boldsymbol{b}})^{(A_{1ij}^*>C)(1-S_{2ij})} F_{\mu_{ij},\sigma_1^{2^*}}(C_1|\underline{\boldsymbol{b}})^{(A_{1ij}^*=C_1)(1-S_{2ij})} \Big\} \Big] g(\underline{\boldsymbol{b}}) d\underline{\boldsymbol{b}}, \qquad (4)
\end{aligned}
$$

where $f_{(\mu_y,\mu_z),(\sigma_y^2,\sigma_z^2,\rho)}(y,z)$ and $F_{(\mu_y,\mu_z),(\sigma_y^2,\sigma_z^2,\rho)}(y,z)$ denote bivariate Gaussian densities and cumulative distribution functions with means (μ_y,μ_z), variances (σ_y^2, σ_z^2), and correlation ρ. Further, $f_{\mu_y,\sigma_y^2}(y)$ and $F_{\mu_y,\sigma_y^2}(y)$ denote the Gaussian density and cumulative distribution functions with mean μ_y and variance σ_y^2, and $g(\underline{\boldsymbol{b}})$ denote a multivariate normal density function with mean $\underline{\boldsymbol{0}}$ and variance $\underline{\boldsymbol{D}}$.

When only one assay is available on each measurement as in Hughes (1999), an individual's contribution to the likelihood corresponding to model (1) reduces to

$$
L_i = \int_{\underline{\boldsymbol{b}}} \prod_{j=1}^{n_i} \Big\{ f_{\mu_{ij},\sigma_1^{2^*}}(A_{1ij}^*|\underline{\boldsymbol{b}})^{A_{1ij}^*>C} F_{\mu_{ij},\sigma_1^{2^*}}(C_1|\underline{\boldsymbol{b}})^{A_{1ij}^*=C_1} \Big\} g(\underline{\boldsymbol{b}}) d\underline{\boldsymbol{b}}. \qquad (5)
$$

For high dimensional random effects, Monte-Carlo techniques such as a Monte-Carlo EM algorithm (McCulloch, 1997) can be used for maximizing (4) or (5). In fact, Hughes (1999) proposed just such a technique for maximizing (5). For a single random intercept as in (3), I found that a simple trapezoidal rule worked well for evaluating the integrals in (4) and (5). Specifically, I summed over 800 equally spaced points between -4 and 4 standard deviations in the random intercept. Further, the likelihoods were maximized using quasi-Newton-Rhapson algorithm implemented in GAUSS (Aptech, 1992).

For the SIV/vaccine data analysis, variance estimation were based on asymptotic approximations (i.e., inverting the information matrix).

3 Application

I applied various modeling approaches to the vaccinated animals ($I = 12$) in order to estimate the mean change in log-tranformed viral load from the chronic SIV period to the period on ART and to the release period after ART was terminated. Specifically, using (3), I fit (i) a joint model, (ii) a model which used only the initial assay (single assay), and (iii) a model which replaced the initial assay measurement with the second more sensitive assay measurement when the initial assay result was below its lower detection limit and when there was sufficient sample available to perform the second assay (replacement assay). Samples were available in 48% of the samples for which the initial assay was below the lower-limit of detection. The last approach involves maximizing a likelihood, where the individual contribution is

$$\begin{aligned} L_i &= \int_{\underline{\boldsymbol{b}}} \Big[\prod_{j=1}^{n_i} \Big\{ f_{\mu_{ij},\sigma_1^{2*}}(A^*_{1ij}|\underline{\boldsymbol{b}})^{(A^*_{1ij}>C)(1-S_{2ij})} F_{\mu_{ij},\sigma_1^{2*}}(C_1|\underline{\boldsymbol{b}})^{(A^*_{1ij}=C_1)(1-S_{2ij})} \\ &\times f_{\mu_{ij},\sigma_2^{2*}}(A^*_{2ij}|\underline{\boldsymbol{b}})^{(A^*_{2ij}>C)S_{2ij}} F_{\mu_{ij},\sigma_2^{2*}}(C_2|\underline{\boldsymbol{b}})^{(A^*_{2ij}=C_2)S_{2ij}} \Big\} \Big] g(\underline{\boldsymbol{b}}) d\underline{\boldsymbol{b}}. \end{aligned} \tag{6}$$

Although appealing for its simplicity, the replacement analysis is biased even with $\Delta = 0$ since it does not explicitly account for the probabilistic mechanism of selecting samples for

performing the second assay. Note that the likelihood for the joint model given by equation (4) does account for this selection, while (6) does not. The selection bias in the replacement analysis is demonstrated with a simple example in the Appendix.

Table 1 presents the results of parameter estimation from the three different modeling approaches fit to the vaccinated animals' data. The joint model which incorporated both the initial assay (with a known detection limit of 20,000 copies per ml) and the second more sensitive assay (with a known detection limit of 4,000 copies per ml) showed a substantial reduction in viral load from the SIV chronic phase to the ART period (i.e., $\widehat{\beta}_1 = -2.42$). Although, there was a slight rebound in viral load during the release period (period after ART is terminated), viral load was still substantially lower than during the chronic SIV period ($\widehat{\beta}_2 = -0.97$), suggesting a benefit to the therapeutic vaccine. The joint model showed a large correlation between the two assays ($\text{logit}\rho^* = 2.88$), suggesting that for both assays, the measurement errors for the two assays (σ_1^2 and σ_2^2) are small relative to the residual variation (σ^2). In addition, the estimate of Δ suggested that there is a systematic difference between the initial assay and the second assay. Namely, the mean value for the second assay was estimated as 0.402 less than the mean value for the initial assay.

Table 1 presents parameter estimates from two simpler models. Similar to the joint model, the model which incorporates only the initial assay (i.e., single assay) showed a substantial reduction in viral load from the SIV chronic phase to the ART phase, with a partially sustained reduction in the release phase. The estimates of β_0, β_1 and β_2 were nearly identical for this simpler model as compared with the joint model. Interestingly, for this analysis, there was also little efficiency gain in estimating β_0, β_1 and β_2 by using the joint model over using the simpler model. Specifically, the standard errors were only slightly smaller for the joint model as compared with the single assay model. One possible reason for the lack of an efficiency gain in using the joint model is the difficulty in estimating ρ^*, which is reflected in the large standard error associated with this estimate. Further, the variability in estimating Δ, a mean shift between assays, was large. Thus, the results suggest that it

may be difficult to estimate the association between the two assays when both assays are never simultaneously observed above their known lower detection limits. This raises the possibility that the joint model may be substantially more efficient when, on at least some occasions, both assays are observed above these limits. This will be examined in more detail in the simulation section.

I examined the fit of the joint model by comparing the quantiles from the actual values relative to those obtained by simulating data based on the model with estimated model parameters. Figure 2A shows a Q-Q plot for the single assay fit, while Figure 2B shows a Q-Q plot for the second assay (only observed in cases where the initial assay was below its detection limit and there was sufficient plasma sample to perform this assay). The plots were generated by simulating data according to the model (with model parameters given by estimates presented in Table 1) with the same data structure (numbers of patients and follow-up times) as in the vaccinated animal group. The expected quantiles under the assumed model were evaluated by taking the mean value of the empirical distribution across 500 simulated datasets. The Q-Q plot was constructed by plotting the observed empirical distribution from the actual dataset against the expected quantiles obtained through simulation. In constructing Figure 2B, in each simulation, I took a random sample of second assay values corresponding to the number of second assay measurements in the SIV/vaccine dataset. Figure 2A shows that the model adequately describes the initial assay data in the vaccine group. Figure 2B shows that the observed quantiles of second assay values are often larger than the expected quantiles (i.e., observations appear above the 45 degree line). However, it is important not to over emphasize this Q-Q plot since it is based on only a relatively small number of second assay values. Further, additional calculations showed that the number of second assays which were below the lower detection limit was 49 of 66. The expected number based on a simulation of 5,000 datasets under the estimated model was 54.8 (95% of the number below the lower detection limits ranged between 48 and 60). Thus, the observed number of second assay measurements which were below detectable limits is

within sampling error of what would be expected under a correctly specified joint model.

Table 1 presents the results of the approach in which the second assay replaced the initial assay when it was available (estimates were obtained by maximizing (6)). Although, the approach gave qualitatively similar results to the other two approaches, estimates of β_1 and β_2 were attenuated as compared with the other two approaches. One has to be careful in interpreting the results since there are inherent biases with the replacement approach. First, if there is a systematic difference between the two assays (i.e., Δ not equal to zero) then, clearly, the replacement analysis will result in biased estimation. Second, even if there is no systematic difference between the two assays (i.e., $\Delta = 0$), the fact that I only replace the initial measurements with the second assay when the initial assay is below the level of detection may result in selection bias. I will investigate these biases in more detail with simulations in the next section.

I found that, in the vaccine group, all methods show a sizable and statistically significant reduction in viral load while the animal is being treated with ART. Of more interest, is what happens after ART is terminated. In the vaccinated group, all methods show that there was substantial reduction in viral load in the follow-up period relative to the SIV chronic phase. This suggested that the vaccine may have induces changes in viral load that are partially maintained after ART termination. In order to be assured that this was not just the natural history of SIV disease in these animals, data from a control group with longitudinal viral load measurements were obtained. Estimates of β_0, β_1 and β_2 from the joint model fit to the control data ($I = 11$) were $\widehat{\beta}_0 = 5.88$ (SE= 0.174), $\widehat{\beta}_1 = -2.26$ (0.100), and $\widehat{\beta}_2 = -0.006$ (0.077). Figure 2C shows a Q-Q plot of the initial assay viral load data in the control group, which was constructed in a similar way to Figure 2A for the vaccinated group. The figure illustrates the good fit of the model to the initial assay measurements in the control group. Since there were very few measurements taken with the second assay in the control group, I did not assess the goodness of fit for the second assay in this group. Specifically, only 2% of samples with initial assay measurements below the lower limit of detection were measured

with the second assay in the control group. Although estimates of β_0 and β_1 were nearly the same for the vaccination and control groups, the estimate of β_2 was very different between the two groups. Specifically, this estimate was nearly zero for the control group, suggesting a near immediate rebound in viral load after ART is terminated in this group. This is in contrast with the sizable negative estimate of β_2 in the vaccination group.

For the control group, I also fit the single assay model (using only the initial assay) and the replacement model (where I replaced the initial assay by the second more sensitive assay). Estimates of β_0, β_1, and β_2 (and their standard errors) were nearly identical across the three methods (data not shown). Primarily, this was due to the fact that there were very few assays performed with the second assay in the control group.

The data analysis in this section suggested that there may be very little pay-off in measuring a second more expensive assay only when measurements are below the level of detection on the initial assay. I will examine this further with simulations in the next section.

4 Simulations

I begin by presenting the results from a simulation similar to the example. Specifically, I simulated data with the same data structure (i.e., same number of observations per patient and number of patients) as the vaccine group in the example. I did not perform simulations corresponding to the control group since there were very few second assays performed in this group. I examine the properties of the joint modeling approach and compare these results with two simpler approaches under different designs. The estimators are compared under a design in which the second assay is only performed when the initial assay is below the level of detection (as in the SIV/vaccine study) and when the second assay is performed at random without regard to the initial assay values. For computation simplicity, these simulations were done assuming that there was no between-subject variation ($\sigma_b = 0$) in the simulation and estimation. The effect of accounting for between-subject variation in the models will be presented in a simulation later in this section. Table 2 shows the results

of simulations for estimating β_0, β_1, β_2, Δ, and logitρ^* for different true values of Δ, ρ^*, and initial assay variation (σ_1). Other parameters were chosen similar to the estimates obtained for the vaccination group (see Table 1). Further, values of C_1 and C_2 were chosen to correspond to the example ($C_1 = \log(20,000) = 4.8$ and $C_2 = \log(4,000) = 3.6$). I compared the following models/designs: A) the joint modeling approach where the second assay is performed only when the initial assay is below the level of detection ($A_{1ij} \leq C_1$), B) the joint modeling approach where the second assay is performed at random with the same proportion of second assays as in A, C) the joint modeling approach where both the initial and second assay are measured at all time points, D) the single assay approach where only the initial assay is preformed, and E) the replacement approach where the initial assay measurement is replaced by the second assay measurement when the initial value is below its lower detection limit. For all simulation scenarios, the second more sensitive assay was performed in approximately 48% of plasma samples.

The first set of parameters in Table 2 ($\Delta = 0.40$ and $\log\sigma_1^* = -0.50$) shows the performance of the various approaches when all parameter values are similar to those estimated in the vaccine group. As in the example, standard deviations of the parameter estimates of β_0 and β_2 are similar across the different methods and designs. However, there are differences in estimating β_1 across many of the model/designs. A comparison of A versus D shows the efficiency advantages of incorporating the second more sensitive assay into analysis. Although there is little efficiency advantage in using the joint model for estimating β_0 or β_2, there is a large pay-off in incorporating the second assay for estimating β_1. The lack of an efficiency gain for estimating β_0 or β_2 is due to the fact that there were few initial assays below detectable limits during the chronic phase or during the release period. For the first set of parameter values, the simulation results suggest a large pay-off in efficiency for estimating β_1 with the relative efficiency of the joint model over the single assay of $(0.206/0.132)^2 = 2.44$ for estimating β_1. In general, the efficiency for estimating β with the joint relative to the single assay model increases as ρ^* increase, as the ratio of σ_1^* to σ_2^* increases, and as the

detectable limit of the second assay decreases relative to the initial assay.

A comparison of A and B provides a comparison of the design in which the second assay is only performed when the initial assay is below its lower detection limit and the design in which the second assay is performed at random without regard to the initial assay values. For the first set of parameter values, this comparison shows a moderate efficiency gain in estimating β_1 (relative efficiency of 1.10) under the latter type of design. The efficiency is much larger when the variance parameters are increased. Specifically, when the variances σ_1^* and σ_2^* were increased to $\exp(0)$ and $\exp(-0.3)$ (third set of parameter values in Table 2), respectively, the relative efficiency for estimating β_1 with B versus A was increased to $(0.154/0.119)^2$=1.67. In addition to the efficiency gain in estimating β_1, Δ and ρ^* are estimated more efficiently under scenario B versus scenario A in all reasonable situations.

An alternative design in which the second assay is performed on all initial assays below detectable limits and performed on a random fraction of initial assays above detectable limits is evaluated through simulation. The simulation was conducted as in Table 2 with the third set of parameter values ($\Delta = 0$ and $\log \sigma_1^* = 0$). There are large efficiency gains in performing the second assay on even a small percentage of assays above detectable limits (data not shown). For example, when the second assay is performed on a random set of 0%, 5%, and 50% of serum samples with initial assays above detectable limits, the standard errors for $\widehat{\beta}_1$ are 0.154, 0.130, and 0.106, the standard errors for $\widehat{\Delta}$ are 0.071, 0.049, and 0.029, and the standard errors for $\text{logit}\widehat{\rho}^*$ are 0.98, 0.36, and 0.16, respectively. Thus, there are sizable efficiency gains in simply performing the second assay on a small fraction of samples for which the first assay is above detectable limits.

Table 2 shows bias in estimating β_0, β_1 and β_2 using the replacement analysis (E). For the first parameter configuration (with $\Delta = -0.40$), there is substantial negative bias in estimating β_1 (-3.0 versus -2.4) and substantial positive bias for estimating β_2 (-0.85 versus -1.00). Even when there is no systematic difference between the two assays ($\Delta = 0$), there is substantial bias in estimating β_1 when σ_1^2 is large (third set of parameters in Table 2). This

bias was substantially reduced when σ_1^2 was reduced (second set of parameters in Table 2).

I also conducted a simulation study focusing on estimating a simple linear regression, $Y_{ij} = \beta + 0 + \beta_1 t_j + \epsilon_{ij}$ (Table 3). Specifically, I considered a data structure with $I = 100$ with plasma samples measured at six equally spaced time points ($t_j = j$, for $j = 1, 2, ..., 6$). I assumed that $\beta_0 = 5.5$, $\beta_1 = -0.50$, and the lower detection limits for the initial and second assays are $C_1 = 4$ and $C_2 = 2$, respectively. For different values of $\log\sigma_1^*$, $\log\sigma_2^*$, and Δ, I computed simulation means and standard deviations of key parameters for the joint modeling approach, replacement approach, and single assay approach for the same models and designs presented in Table 2. As in Table 2, for computation simplicity, data were simulated without between-subject variation, and estimation did not account for this variation. The results show that scenarios A thru D provide nearly unbiased estimates of β_0, β_1, Δ and ρ^*. However, the replacement analysis (E) can be highly biased even in the situation in which $\Delta = 0$. In all cases, the parameters β_0 and β_1 are biased with the replacement analysis.

A comparison of B with A show the increase in efficiency of performing the second assay at random rather than only when the initial assay is below the limits of detection. For example, when $\log\sigma_1^* = 0$, $\log\sigma_2^* = -0.5$, $\Delta = 0$, and $\rho^* = 0.95$, there is a 54% gain in relative efficiency $((0.0196/0.0158)^2)$ for estimation β_1 by performing the second measurement at random without regard to the initial assay value as compared with performing the second assay only when the initial assay is below its lower detection limit. This magnitude is comparable for other parameter configurations. A comparison of scenario A-C with D shows that there may be substantial efficiency advantages to incorporating the second assay into the analysis rather than just using the single initial assay.

Table 4 shows the results of simulations which incorporate between-subject variation. I fit linear models similar to the model presented in Table 3, but with the addition of a random intercept term. I present results for $\sigma_b = 0, 1$, and 1.5. The results are similar to those presented in Table 3, when I did not incorporate between-subject variation. Namely,

(i) there is efficiency gain in choosing a random set of time points for performing the second assay as compared with only performing the second assay when the first assay is below its detection limit (i.e., B is more efficient than A), (ii) there is efficiency gain in jointly modeling both assays as compared with only modeling the single assay (i.e., A is more efficient than D), and (iii) simply replacing the initial assay with the second more sensitive assay results in biased estimation. (i.e., E results in biased estimation).

Inference about $\underline{\boldsymbol{\beta}}$ is robust to departures in the measurement error distribution when Y_{ij} is not subject to detection limits. Simulations were conducted to examine the robustness of the various models and designs to lower detection limits in Y_{ij} when the measurement error distributions are misspecified. Table 5 shows the results of simulations when data are simulated as in Table 2 with the exception that the measurement errors are generated as mixtures of normals as opposed to correctly specified normal distributions. I present simulations with an increasing separation between the two normal distributions, reflecting increasing departure from normality in the measurement error distributions. The results in Table 5 suggest that the joint and single assay models are robust to moderate departure from normality in the measurement error distributions. For large departures such as the third mixture distribution in Table 5, resulting in a distinct bimodal measurement error distribution, there was sizable bias for many of the parameters in all of the models and designs.

5 Discussion

This paper presented an approach for modeling longitudinal biomarker data with multiple assays, each of which had different known lower detection limits and different measurement errors. With continuing advances in the development of new biomarkers for assessing disease outcome, the problem of incorporating different assays into a longitudinal data analysis is an increasingly important problem. As an example, the motivation for this research comes from a prospective vaccination study where the outcome is a measure of RNA viral load which can

be measured by a number of well established assays. Due to cost and/or laborious laboratory work, it may not be feasible to perform the more sensitive assay for each longitudinal plasma sample. Performing a "crude" initial assay (with a high lower detection limit), followed by a more sensitive assay on a fraction of available plasma samples provides a good alternative to performing the more sensitive assay on all samples.

I proposed a joint modeling approach and, with analysis and simulations, compared the approach with a simple replacement approach where I replaced the initial assay measurement with a more sensitive assay measurement when the initial assay value was below its lower detection limit. I also compared the joint approach to a simpler model which only used the initial assay measurements for inference. In general, the joint modeling approach had better statistical properties than the other approaches. Specifically, estimates from the joint model were unbiased and more efficient than estimates which only incorporated the initial assay. Further, the replacement analysis could lead to substantial bias when the measurement errors are moderate or large.

In addition to the comparison of different modeling approaches, the choice of a design for measuring the second assay was important for efficient estimation. The simulation study results showed that there can be large efficiency gains in performing the second assay without regard to the value of the initial assay as compared with performing the second assay only when the initial assay value is below its lower detection limit. Further, the efficiency of the design in which the second assay was performed only when the initial assay value is below detectable limits is greatly improved with only a small fraction of second assays be performed when the initial assay value is above detectable limits.

The results of simulation studies (Table 5) suggested that the joint and single assay models are robust to moderate departure from normality of the measurement error distributions. However, there was sizable bias when this distribution was highly non-normal. It is therefore important for the practitioner to examine departures from normality in the measurement error distribution when using these models. The Q-Q plots presented in the SIV/vaccine

analysis provide a demonstration of this.

All methods discussed in this paper showed a large decrease in viral load during the ART period with a differential effect on viral load during the release period (follow-up after vaccine and ART is terminated) between vaccinated and unvaccinated animals. Specifically, unvaccinated animals showed a quick rebound in viral load, while vaccinated animals showed a partially sustained reduction in viral load after ART was terminated. These are consistent with results presented in (von Gegerfelt et al., 2007) which demonstrated the therapeutic effectiveness of the vaccine by using a more crude analysis in which the initial assay values were replaced with more sensitive assay values when they were available. Further, in this analysis, measurements below lower detection limits were replaced by the value of one-half the detection limit, a strategy which can induce bias relative to a model based approach (Hughes, 1999).

This paper considers only two assays. In fact, in the SIV/vaccine study there were a small group of plasma samples which were measured with additional more sensitive assays (I did not use this added information in the data analysis in this paper). Extensions to include more than two assays is an area for future research. The joint modeling approach can be extended to include an excess proportion of assay measurements below lower detection limits using mixture models (Moulton and Halsey, 1995; Moulton et al., 2002). There was no evidence of such an excess of low values in the SIV/vaccine study (recall that there was a lower proportion of second assay measurements that were below detectable limits than would be expected under the model), but such an extension may be very appropriate in other applications. Such an extension is the subject of future research. The focus of this research is on modeling data with lower detection limits since this was of particular concern in the SIV/vaccine study. The approach could be extended to incorporate upper detection limits for applications where this is appropriate.

The joint model assumes that the difference between the two assays is a constant (equation (2)), and therefore a slope of 1 is assumed between the two measurements. Unfortu-

nately, it is very difficult to estimate a more complex relationship between the two assays when as in the SIV/vaccine study, second assays are only performed when the initial assay is below detectable limits. When a sizable proportion of both assay values are above detectable limits, a more flexible relationship between the two measurements that incorporates an arbitrary slope in a simple linear model or a cubic spline representation is possible. This is an area for future research.

Acknowledgments

I thank Dr. George Pavlakis of the Human Retrovirus Section of the Vaccine Branch of NCI for providing the data to illustrate the methodology developed in this paper. This study utilized the high-performance computational capabilities of the Biowulf Linux cluster at the National Institutes of Health, Bethesda, MD (http://biowulf.nih.gov). I thank the associate editor and referee who made constructive comments which lead to an improved paper.

References

Abramowitz, M. and Stegun, I.A. (1972). *Handbook of Mathematical Functions*. Dover, New York.

Aptech Systems (1992). *Gauss Systems*, Version 3.0. Kent, Washington: Aptech systems.

Hughes, J.P. (1999). Mixed effects models with censored data with application to HIV RNA levels. *Biometrics* **55**, 625-629.

Laird, N.M. and Ware, J.H. (1982). Random-effects models for longitudinal data. *Biometrics* **38**, 963-974.

McCulloch, C.E. (1997). Maximum likelihood algorithms for generalized linear mixed models. *The Journal of the American Statistical Association* **92**, 162-197.

Moulton, L.H. and Halsey, N.A. (1995). A mixture model with detection limits for regression analysis of antibody response to vaccine. *Biometrics* **51**, 1570-1578.

Moulton, L.H., Curriero, F.C., and Barroso, P.F. (2002). Mixture models for quantitative HIV RNA data. *Statistical Methods in Medical Research* **11**, 317-325.

Thiebaut, R., Jacqmin-Gadda, H., Babiker, A., Commenges, D., and the CASCADE collaboration. (2005). Joint modelling of bivariate longitudinal data with informative dropout and left censoring, with application to the evolution of CD4+ cell counts and HIV RNA viral load in response to treatment of HIV infection. *Statistics in Medicine* **24**, 65-82.

von Gegerfelt, A.S., Rosati, M, Alicea, C., Valentine A., Roth, P., Bear, J., Franchini, G., Albert, P.S., Bischofberger, N., Boyer, J.D., Weinter, D.B., Markham, P., Israel, Z.R., Eldridge, J.H., Pavlakis, G.N., and Felber, B.K. (2007). Long-lasting decrease of ivremia in macaques chronically infected with simian immunodeficiency SIVmac251 after therapeutic DNA immunization. *Journal of Virology* **81**, 1972-1979.

Appendix: Bias in the replacement analysis

Asymptotic bias is shown for the replacement analysis in a simple illustrative example. Consider the simple case in which there is a single measurement on each individual. Denote A_{1i} and A_{2i} as the initial and second assay values, respectively, with both having mean μ ($\Delta = 0$). Also, let C_1 be the lower detection limit for the initial assay and suppose that there is no lower limit of detection and no measurement error for the second assay. Also denote ϵ_i as the residual error and δ_{1i} as the measurement error for the first assay, with the variance of these two quantities being denoted as σ^2 and σ_1^2, respectively. The two assay values can be expressed as $A_{1i} = \mu + \delta_{i1} + \epsilon_i$ and $A_{2i} = \mu + \epsilon_i$, respectively. The asymptotic bias was evaluated by generating a sample of 20,000 individuals and estimating μ by maximizing a likelihood similar to (6). When $\mu = 3$, $C_1 = 3$, and $\sigma_1^2 = 1$, the asymptotic bias of μ ($\widehat{\mu}$-μ) is 0.54, 0.55, and 0.81, for $\sigma_1^2 = 0.1$, 0.2, and 1, respectively. In all cases tried. μ was positively biased in the replacement analysis.

Figure Legends:

Figure 1: Profiles for the first four patients in the vaccination cohort. 1=initial assay, while 2=second assay, The dotted line shows the lower detection limit for the initial assay, while the -.-. line shows the lower detection limit for the second assay. The solid line shows the SIV chronic phase, while the dashed line - - - shows the release period after ART is terminated. The area without a line shows the viral load during the ART period.

Figure 2: Q-Q plots for (A) initial assay for vaccinated group, (B) second assay for vaccinated group, (C) initial assay for control group.

Table 1: Models fit to vaccine treated Macaques ($I = 12$).

Models	Parameter Estimates (SE)							
	$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\Delta}$	$\log \widehat{\sigma}_b$	$\log \widehat{\sigma}_1^*$	$\log \widehat{\sigma}_2^*$	$\text{logit}\widehat{\rho}^*$
Joint	5.56	-2.42	-0.97	-0.40	-0.78	-0.45	-0.27	2.88
	(0.15)	(0.14)	(0.09)	(0.09)	(0.25)	(0.05)	(0.12)	(0.75)
Single Assay	5.56	-2.44	-0.97	–	-0.76	-0.45	–	–
	(0.15)	(0.15)	(0.09)	–	(0.22)	(0.05)	–	–
Replacement	5.56	-2.37	-0.85	–	-0.85	-0.51	0.67	–
	(0.14)	(0.14)	(0.09)	–	(0.22)	(0.05)	(0.21)	–

Table 2: Simulation: Data simulated according to (3) with $\beta_0 = 5.6$, $\beta_1 = -2.4$, $\beta_2 = -1.0$, $\rho^* = 0.95$, and $\log \sigma_2^* = -0.30$ with the same data structure as the vaccine group in SIV/vaccine study (i.e., $I = 12$ and number/time of measurements identical to the example). The models/designs are: (A) Joint model with second assay only observed when the initial assay is below its lower-limit of detection. (B) Joint model with second assay observed at random, independent from the results of the initial assay, with the same proportion of second assay measurements as in A. (C) Joint model with second assay observed at all time points. (D) Model for single initial assay. (E) Replacement analysis. Simulations are based on 2000 simulated datasets. Standard errors (SE) is the estimated standard deviation in parameter estimates over the 2000 simulations.

Parameters		Design-Model	Average (SE)				
Δ	$\log \sigma_1^*$		$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\Delta}$	logit $\widehat{\rho}^*$
-0.40	-0.50	A	5.60	-2.41	-1.00	-0.40	3.20
			(0.059)	(0.132)	(0.078)	(0.066)	(0.96)
		B	5.60	-2.41	-1.00	-0.40	2.96
			(0.057)	(0.126)	(0.075)	(0.023)	(0.18)
		C	5.60	-2.41	-1.00	-0.40	2.95
			(0.056)	(0.112)	(0.072)	(0.017)	(0.14)
		D	5.60	-2.42	-1.00	–	–
			(0.059)	(0.206)	(0.078)	–	–
		E	5.62	-2.99	-0.85	–	–
			(0.057)	(0.280)	(0.078)	–	–
0	-0.50	A	5.60	-2.40	-1.00	-0.00	3.17
			(0.059)	(0.113)	(0.075)	(0.063)	(0.84)
		B	5.60	-2.40	-1.00	-0.00	2.96
			(0.057)	(0.110)	(0.075)	(0.023)	(0.17)
		C	5.60	-2.40	-1.00	-0.00	2.95
			(0.056)	(0.095)	(0.072)	(0.017)	(0.13)
		D	5.60	-2.42	-1.00	–	–
			(0.060)	(0.200)	(0.078)	–	–
		E	5.61	-2.44	-0.88	–	–
			(0.057)	(0.125)	(0.081)	–	–
0	0	A	5.60	-2.40	-1.00	-0.00	3.11
			(0.095)	(0.154)	(0.119)	(0.071)	(0.98)
		B	5.60	-2.40	-1.00	0.00	2.96
			(0.086)	(0.119)	(0.099)	(0.031)	(0.17)
		C	5.60	-2.40	-1.00	0.00	2.96
			(0.079)	(0.096)	(0.081)	(0.025)	(0.13)
		D	5.60	-2.40	-1.00	–	–
			(0.098)	(0.168)	(0.130)	–	–
		E	5.40	-2.22	-1.04	–	–
			(0.202)	(0.247)	(0.151)	–	–

Table 3: Simulation: Comparison of models/designs for estimating a simple linear regression model. The design has $I = 100$ subjects and $n_i = 6$ (for all i) equally spaced time points ($t_j = j$) with $Y_{ij} = \beta_0 + \beta_1 t_j + \epsilon_{ij}$, where $\beta_0 = 5.5$, $\beta_1 = -0.50$, $\rho^* = 0.95$, and the lower limit of detection for the two assays are $C_1 = 4$ and $C_2 = 2$. The models/designs are: (A) Joint model with second assay only observed when the initial assay is below its lower-limit of detection. (B) Joint model with second assay observed at random, independent from the results of the initial assay, with the same proportion of second assay measurements as in A. (C) Joint model with second assay observed at all time points. (D) Model for single initial assay. (E) Replacement analysis. Simulations are based on 2000 simulated datasets.

Parameters			Design-Model	Average (SE)			
Δ	$\log \sigma_1^*$	$\log \sigma_2^*$		$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\Delta}$	logit $\widehat{\rho}^*$
0	0	-0.50	A	5.50	-0.500	0.00	3.03
				(0.083)	(0.020)	(0.049)	(0.44)
			B	5.50	-0.500	0.00	2.95
				(0.076)	(0.016)	(0.030)	(0.14)
			C	5.50	-0.500	0.00	2.95
				(0.070)	(0.013)	(0.027)	(0.11)
			D	5.50	-0.500	–	–
				(0.104)	(0.034)	–	–
			E	5.02	-0.429	–	–
				(0.104)	(0.022)	–	–
-0.40	0	-0.50	A	5.50	-0.500	-0.40	3.03
				(0.086)	(0.021)	(0.050)	(0.46)
			B	5.50	-0.500	0.00	2.95
				(0.075)	(0.016)	(0.030)	(0.14)
			C	5.50	-0.500	0.00	2.95
				(0.070)	(0.013)	(0.027)	(0.11)
			D	5.50	-0.500	–	–
				(0.104)	(0.034)	–	–
			E	4.52	-0.409	–	–
				(0.103)	(0.022)	–	–
0	-0.50	-1.0	A	5.50	-0.50	0.00	3.05
				(0.050)	(0.013)	(0.035)	(1.12)
			B	5.50	-0.500	0.00	2.96
				(0.047)	(0.010)	(0.018)	(0.14)
			C	5.50	-0.500	0.00	2.95
				(0.044)	(0.008)	(0.016)	(0.11)
			D	5.50	-0.501	–	–
				(0.069)	(0.026)	–	–
			E	5.35	-0.476	–	–
				(0.067)	(0.014)	–	–

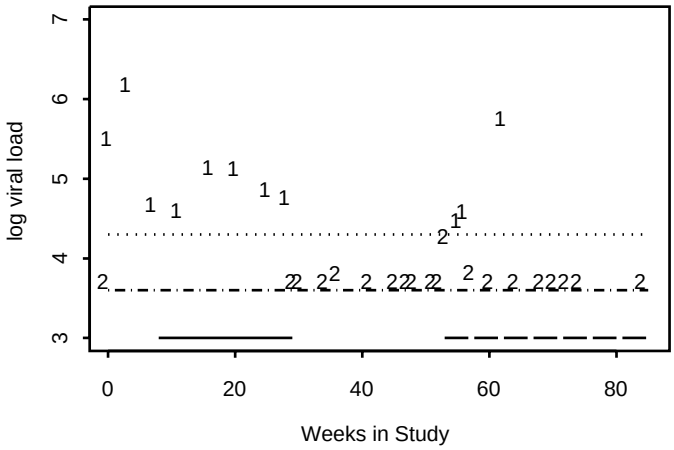
Table 4: Simulation: Comparison of models/designs for estimating a linear mixed model of the form $Y_{ij} = \beta_0 + \beta_1 t_j + b_i + \epsilon_{ij}$, where $\sigma_b = \text{var}(b_i)$, $j = 1, 2, .., 6$ $n_i = 6$ (for all i) and $I = 100$. Further, $\beta_0 = 5.5$, $\beta_1 = -0.50$, $\Delta = 0$, $\log \sigma_1^* = 0$, $\log \sigma_2^* = -0.5$, $\rho^* = 0.95$, and the lower limit of detection for the two assays are $C_1 = 4$ and $C_2 = 2$. The models/designs are: (A) Joint model with second assay only observed when the initial assay is below its lower-limit of detection. (B) Joint model with second assay observed at random, independent from the results of the initial assay, with the same proportion of second assay measurements as in A. (C) Joint model with second assay observed at all time points. (D) Model for single initial assay. (E) Replacement analysis. Simulations are based on 2000 simulated datasets.

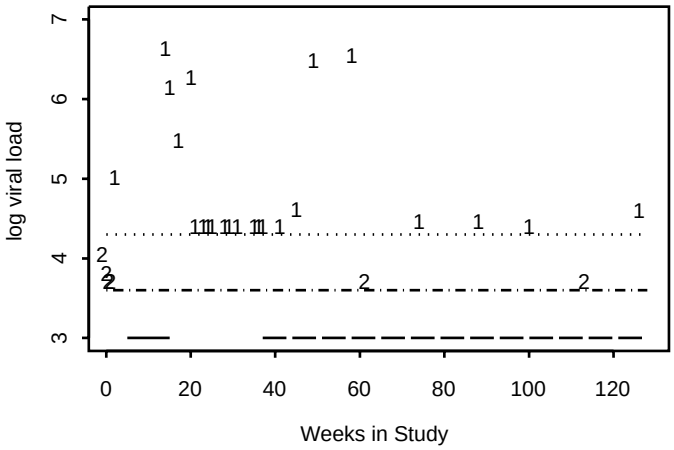
Between-subject variance	Design-Model	Average (SE)			
σ_b		$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\Delta}$	logit $\widehat{\rho}^*$
0	A	5.50	-0.500	0.00	3.04
		(0.083)	(0.020)	(0.050)	(0.46)
	B	5.50	-0.499	0.00	2.97
		(0.075)	(0.016)	(0.031)	(0.14)
	C	5.50	-0.499	0.00	2.95
		(0.070)	(0.013)	(0.028)	(0.11)
	D	5.50	-0.501	–	–
		(0.106)	(0.035)	–	–
	E	5.02	-0.428	–	–
		(0.100)	(0.021)	–	–
1	A	5.50	-0.499	0.00	3.13
		(0.130)	(0.021)	(0.059)	(0.69)
	B	5.50	-0.499	0.00	2.96
		(0.124)	(0.016)	(0.030)	(0.17)
	C	5.50	-0.499	0.00	2.95
		(0.120)	(0.013)	(0.027)	(0.13)
	D	5.50	-0.500	–	–
		(0.149)	(0.035)	–	–
	E	5.18	-0.455	–	–
		(0.177)	(0.025)	–	–
1.5	A	5.50	-0.500	0.00	3.17
		(0.172)	(0.021)	(0.065)	(0.81)
	B	5.50	-0.500	0.00	2.96
		(0.168)	(0.017)	(0.030)	(0.17)
	C	5.50	-0.499	0.00	2.95
		(0.163)	(0.013)	(0.026)	(0.12)
	D	5.50	-0.500	–	–
		(0.190)	(0.035)	–	–
	E	5.29	-0.473	–	–
		(0.208)	(0.024)	–	–

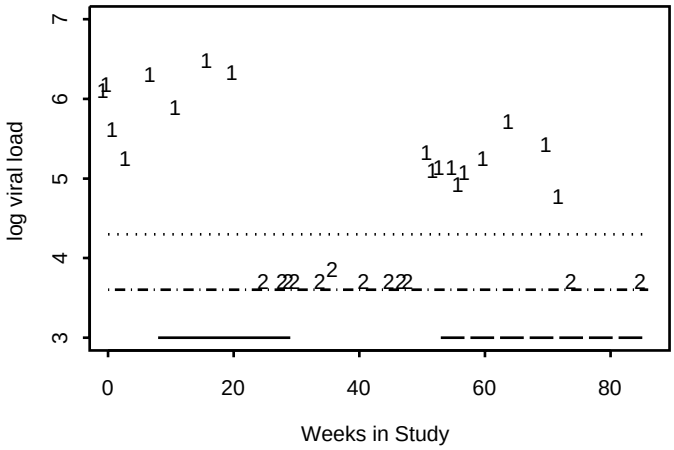
Table 5: Simulation to examine the effect of model misspecification of the measurement error distribution on estimation. Data are simulated with the same data structure in Table 2 with measurement error generated as a two-group mixture of normals as compared with the normal distribution in Table 2. As in Table 2, $\beta_0 = 5.6$, $\beta_1 = -2.4$, $\beta_2 = -1.0$, and $\rho^* = 0.95$. Further, as in the third set of parameters in Table 2, $\Delta = 0$, $\log \sigma_2^* = 0$ and $\log \sigma_2 = -0.30$. The models/designs are the same as defined in Table 2. Simulations are based on 2000 simulated datasets. Standard errors (SE) is the estimated standard deviation in parameter estimates over the 2000 simulations.

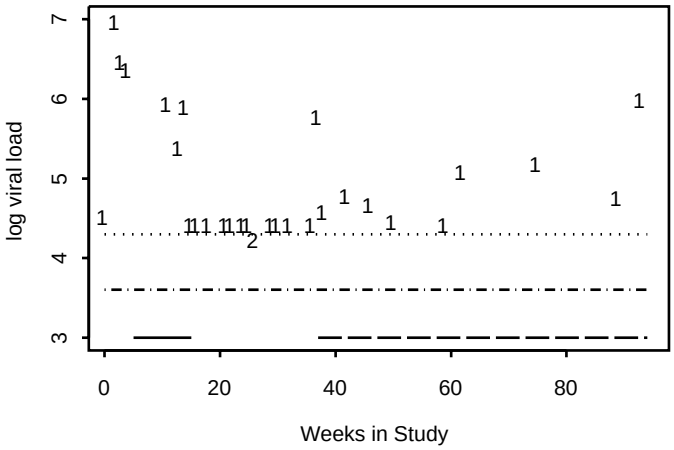
Mixture of normals[1]	Design-Model	Average (SE)				
		$\widehat{\beta}_0$	$\widehat{\beta}_1$	$\widehat{\beta}_2$	$\widehat{\Delta}$	logit $\widehat{\rho}^*$
$\frac{1}{2}N(-1, 0.50) + \frac{1}{2}N(1, 0.50)$	A	5.60	-2.37	-1.01	-0.00	3.05
		(0.097)	(0.154)	(0.118)	(0.074)	(0.86)
	B	5.60	-2.38	-1.00	0.00	2.97
		(0.086)	(0.116)	(0.095)	(0.030)	(0.17)
	C	5.60	-2.38	-1.01	0.00	2.95
		(0.079	(0.095)	(0.078)	(0.025)	(0.12)
	D	5.59	-2.37	-0.99	–	–
		(0.101)	(0.168)	(0.132)	–	–
	E	5.35	-2.13	-1.05	–	–
		(0.221)	(0.265)	(0.150)	–	–
$\frac{1}{2}N(-1.5, 0.31) + \frac{1}{2}N(1.5, 0.31)$	A	5.60	-2.32	-1.04	0.00	3.09
		(0.096)	(0.147)	(0.115)	(0.069)	(0.59)
	B	5.59	-2.34	-1.00	0.00	2.97
		(0.088)	(0.114)	(0.097)	(0.033)	(0.17)
	C	5.60	-2.34	-1.01	0.00	2.95
		(0.081)	(0.094)	(0.079)	(0.027)	(0.12)
	D	5.58	-2.36	-0.99	–	–
		(0.101)	(0.162)	(0.132)	–	–
	E	5.26	-1.98	-1.06	–	–
		(0.219)	(0.257)	(0.152)	–	–
$\frac{1}{2}N(-2, 0.20) + \frac{1}{2}N(2, 0.20)$	A	4.28	-3.52	-1.00	-0.38	2.31
		(0.074)	(0.921)	(0.086)	(0.074)	(1.05)
	B	4.28	-3.69	-1.00	0.38	2.11
		(0.073)	(0.749)	(0.081)	(0.057)	(0.31)
	C	4.28	-3.51	-1.00	0.38	2.07
		(0.073)	(0.83)	(0.067)	(0.047)	(0.22)
	D	4.28	-3.64	-1.03	–	–
		(0.078)	(0.503)	(0.150)	–	–
	E	4.48	-2.91	-0.81	–	–
		(0.057)	(0.764)	(0.068)	–	–

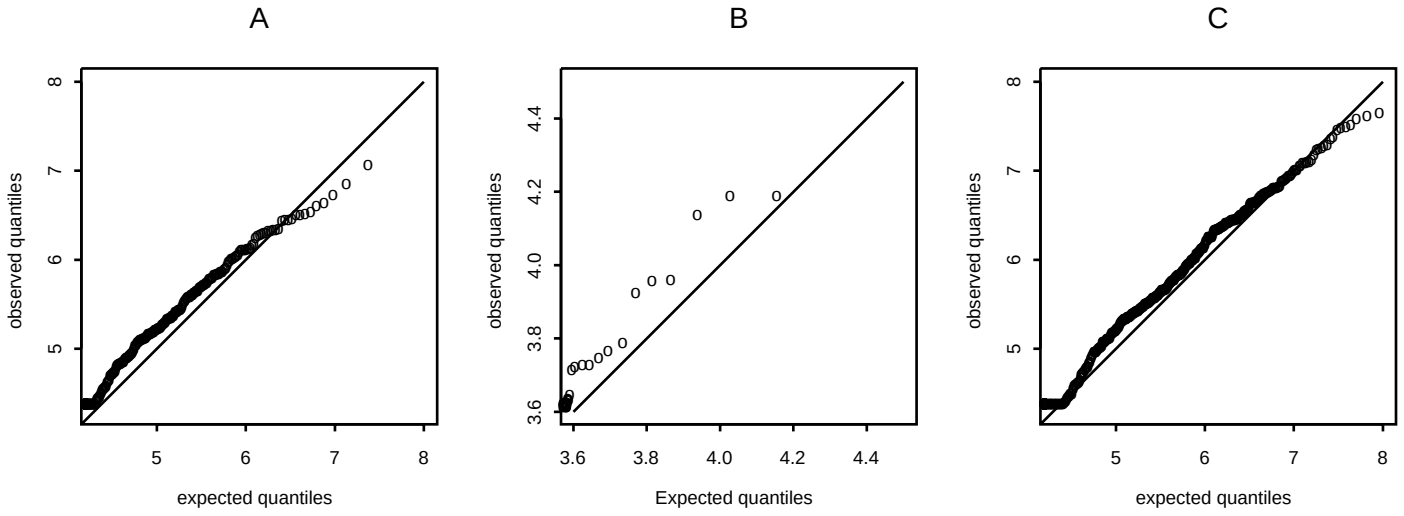
[1] Measurement error for the first and second assay is generated by multiplying the mixture distribution by σ_1^* and σ_2^*, respectively.


log viral load
Weeks in Study


log viral load
Weeks in Study


log viral load
Weeks in Study


log viral load
Weeks in Study


A
B
C
observed quantiles
expected quantiles
Expected quantiles
5
6
7
8
3.6
3.8
4.0
4.2
4.4