Multiple Imputation for Cytokine/Metabolite Assay Data with Missing Data and Values Below Detectable Limits of Quantification

Saijuan Zhang* Ying Guo

Abstract

Many clinical trials data are complicated by the existence of fully missing values or left-censored values known to lie below detection limits, due to biological reasons or assay technical limitations. A conventional practice is to use the actual Lower Limit of Quantification (LLOQ) into the missing value. In this article, we describe a multiple imputation (MI) method for multiply imputing the missing and left-censored values of cytokines and metabolites to understand the potential treatment effects on these biomarkers. A key advantage of multiple imputation is that, once multiple imputed data sets are created, standard analysis methods for complete data can be applied, with imputation uncertainty being addressed by applying MI combining rules. It also provides a convenient approach to limit of quantification issues. We compare the proposed MI method with the existing methods including the conventional substitution analysis and the simple imputation techniques, and demonstrate its superiority in terms of relative bias, efficiency and coverage probability for the 95% confidence interval through simulation studies and a real study which evaluated active treatments on biomarkers to allergen in asthmatics.

Key Words: Assay; Biomarker; Limit of Quantification; Missing data; Multiple Imputation; Censoring.

1. Introduction

In randomized clinical trials, assessing the effect of treatments on the biomarker of disease are often complicated by the existence of fully missing values and left-censored values. Missing data arise naturally due to missed patient visits and premature discontinuations of treatment.

The left-censored data can result when values fall below a lower limit of quantification (LLOQ) of a instrument such as assay. Values below LLOQ are not provided or reported as non-quantitative, that is, $< LLOQ$, as they are considered to have high coefficients of variation. The simplest approach for dealing with such data is the compete-case analysis by removing or deleting all observations falling below the LLOQ. This approach is unappealing as it potentially discards useful information in the data. Several ad-hoc methods are proposed in literature, which replace values falling below the LLOQ with some fraction of LLOQ, such as the value of LLOQ itself, LLOQ/2 or LLOQ/$\sqrt{2}$. This simple substitution approach can lead to bias and invalid estimates of standard error, particularly when a large proportion of values fall below the LLOQ. Singh *et al.* (2002) analyzed the substitution method on censored response values in environmental studies, concluding that highly biased estimates result even in cases with a small percent of censored values and only a single detection limit. Richardson *et al.* (2003) found that substitution led to biased parameter estimates in the regression analysis with covariate alone subject to LLOQ. They show that the direction and magnitude of bias depends on substituted values, measurement error variance, and the underling true distribution of the covariate. Helse *et al.* (2002) reviews several of these substitution procedures, concluding that the substitution method leads to biased estimates and has no theoretical basis. These results provide strong evidence against using ad hoc substitution techniques. Multiple imputation (MI) is less common for dealing

*Merck Research Laboratories; 126 E. Lincoln Avenue, RY34-A316, Rahway NJ, 07065, U.S.A.

with the left-censored values than for missing data, but it represents an appealing option. A key advantage of MI is that, once multiple imputed data sets are created, standard analysis methods for complete data can be applied, with imputation uncertainty being addressed by applying MI combining rules. Some recent applications include Hopke *et al.* (2001), Lubin *et al.*(2004), Chu *et al.*(2008), Uh *et al.*(2008), Guo *et al.* (2010), Lee *et al.*(2012) and Liu *et al.* (submitted to Pharmaceutical Statistics).

In this article, we investigate the MI method that is based on an imputation model condition on additional observed variables (for example, treatment, age and gender, etc.). We show this method compares favorably to the conventional simple substitution method and the existing imputation techniques, specially when a large proportion of observations are below the limit of quantification and the sample size is small.

For the rest part of this article, the outline of the MI method is given in Section 2. A set of simulation studies is presented in Section 3 to investigate the performance of the MI methods compared with the conventional method using simple substitution. In Section 4, we describe the results for the incomplete cytokine/metabolite data based on the proposed MI approach. Finally, Section 5 offers a few concluding remarks.

2. Model and Methods

In clinical trials, censoring and missing may occur in many lab values due to biological reasons or assay technical limitations. Left-censored (or called censored from below) takes place when observations with a value at or below some threshold, all set to the value of that threshold, so that the true value might be equal to or smaller than the threshold. The threshold is often called the LLOQ.

2.1 Conventional method

One of the conventional methods that deal with left-censoring is to use LLOQ itself or $LLOQ/2$ to replace the censored values. This method is very simple compared to MI method in imputation but it may destroy distribution properties. And to deal with missing data, one of the conventional methods is to naively exclude them from analysis or simply fill in with the mean of available non-censored and non-missing data. Here we define the conventional method as the following: replace left-censored data with $LLOQ/2$ and replace missing with the mean of available non-censored and non-missing data. Suppose the observed data are $y_1, \ldots, y_n$ and the LLOQ is denoted as τ. Then we predict y as

$$y_{i,conv} = \begin{cases} y_i & \text{if } y_i > \tau \\ \tau/2 & \text{if } y_i \leq \tau \\ \frac{\sum_i (y_i I(y_i > \tau))}{\sum_i I(y_i > \tau)} & \text{if } y_i = NA \end{cases} \quad (1)$$

This conventional method has many limitations. For example, it cannot be applied to scenarios that LLOQ is smaller than 0, because $LLOQ/2$ failed to jump into censoring area.

2.2 Tobit Regression

One of the existing multiple imputation methods that handle missing and left-censored values is developed using Tobit regression. Tobit Regression was designed to estimate linear relationships between variables when there is either left- or right-censoring in the dependent variable. Tobit regression supposes that for each observed y_i there exists a latent variable y_i^* which linearly depends on its corresponding covariate $X_i = (x_{i1}, \ldots, x_{ip})^T$ and the linear relationship between y_i^* and X_i is determined by $\beta = (\beta_1, \ldots, \beta_p)^T$. In addition, there

is a normally distributed error term which captures the random noise between the latent variable y_i^* and its covariate X_i.

$$y_i = \begin{cases} y_i^* & \text{if } y_i^* > \tau \\ \tau & \text{if } y_i^* \leq \tau \end{cases}$$

Here, we define the existing method as the following: suppose we have only two treatment arms A and B,

$$y_i^* = \begin{cases} \beta_A + \epsilon_{A,i}, \epsilon_{A,i} \sim N(0, \sigma_A^2) & \text{if treatment = A} \\ \beta_B + \epsilon_{B,i}, \epsilon_{B,i} \sim N(0, \sigma_B^2) & \text{if treatment = B} \end{cases} \quad (2)$$

Note that this method uses only information from available dependent variable y_i to predict censored data and missing data. This method does not incorporate all observed variables (for example, age) when generating imputed values. In addition, it performs imputation for censored or missing data on each treatment arm separately which allows for different means and variances in two treatment arms.

However, we argue that when we have a very small clinical trial with around 15 observations for each treatment and high percentage of left-censoring and missing, do we still benefit by doing imputation separately on each treatment arm or would it be better to combine data on treatment arms to increase the available data and perform imputation together? The trade-off is loss in the flexibility of allowing different variances in two treatment arms. When additional information like period, age and gender is available, can we benefit from incorporating the information in imputing censored data and missing data? Hence, our proposed model is:

$$y_i^* = \beta^T X_i + \epsilon_i, \epsilon_i \sim N(0, \sigma^2) \quad (3)$$

Thus, the mean for the censored data should be

$$E(Y_i) = \left\{ \Phi\left(\frac{\beta^T X_i - \tau}{\sigma} \right) [\beta^T X_i + \sigma \lambda_i] \right\} + \Phi(\frac{\tau - \beta^T X_i}{\sigma})\tau$$

where $\lambda_i = \frac{\phi\left(\frac{\tau - \beta^T X_i}{\sigma}\right)}{\left[1 - \Phi\left(\frac{\tau - \beta^T X_i}{\sigma}\right)\right]}$. λ_i is called the inverse Mills ratio (IMR) which measures the amount of censoring - the higher λ_i, the more censoring. $\phi(\cdot)$ and $\Phi(\cdot)$ represent the normal probability density and cumulative distribution function, respectively.

2.3 Imputation procedure

Practically, the following steps are used for sampling values which are below LLOQ. If an observation y_i from subject i is a censored value, we denote the "hidden" variable as W_i and assume it follows a truncated normal distribution. In order to infer a general formula for sampling a censored variable W_i, we now assume W_i is censored by an interval $[a, b]$. Then our case is actually a special case with $a = -\infty$ and $b = LLOQ$ because the "hidden" value is no larger than LLOQ. We first obtain $\hat{\beta}$ and $\hat{\sigma}$, estimates of β and σ, respectively, from Tobit model. Then introduce a latent variable Z (Damien, 2001) which has the joint distribution with W_i given by

$$f_{W_i,Z}(w, z) \propto I_{\left(0, \exp\left(-\frac{(w - \hat{\beta}^T X_i)^2}{2\hat{\sigma}^2}\right)\right)}(z) I(w \in (a, b))$$

Therefore the full conditional distributions $f_{Z|W_i}(z|w)$ and $f_{W_i|Z}(w|z)$ are

$$Z|(W_i = w) \sim U\left(0, \exp\left(-\frac{(w - \hat{\beta}^T X_i)^2}{2\hat{\sigma}^2}\right)\right)$$

$$W_i|(Z=z) \sim U(\max(a, \hat{\beta}^T X_i - \hat{\sigma}\sqrt{-2\ln(z)}), \min(b, \hat{\beta}^T X_i + \hat{\sigma}\sqrt{-2\ln(z)}))$$

Then we can perform Gibbs sampling based on the above conditional distributions.

For Missing data, if an observation y_i from subject i is missing, we denote the "hidden" variable as M_i and assume it follows $N(\hat{\beta}^T X_i, \hat{\sigma}^2)$.

The above steps for imputing values that are no larger than LLOQ and values that are missing are repeated for m times to obtain m-imputed datasets.

3. Simulation

3.1 Simulation settings

The performances of the conventional method, the existing method and our proposed method were investigated by simulations in this section. Five criteria were used for assessment. Suppose we perform N simulations and within each simulation we perform M imputations:

1. Bias of treatment effect estimate:

$$\frac{1}{N}\sum_{i=1}^{N}(\hat{\beta}_{trt,i} - \beta_{trt})$$

   $\hat{\beta}_{trt,i}$ is the average of M multiple imputation (MI) estimates of β_{trt} in the i^{th} simulation.

2. Variance of treatment effect estimate:

$$\frac{1}{N-1}\sum_{i=1}^{N}(\hat{\beta}_{trt,i} - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_{trt,i})^2$$

3. Mean squared error of treatment effect estimate

$$\frac{1}{N-1}\sum_{i=1}^{N}(\hat{\beta}_{trt,i} - \beta_{trt})^2$$

4. 95% confidence interval (CI) coverage probability: it is estimated by the proportions of simulations with CI containing the true treatment effect. A good method should have the coverage probability close to the nominal coverage level, in our case, 95%. The CI of the i^{th}simulation is calculated as

$$\left(\hat{\beta}_{trt,i} - z_\alpha\sqrt{var(\hat{\beta}_{trt,i})}, \hat{\beta}_{trt,i} + z_\alpha\sqrt{var(\hat{\beta}_{trt,i})}\right)$$

   where α is the significance level 0.05, so $z_\alpha = 1.96$. For the i^{th}simulation, the variance of treatment effect estimate is calculated as (Little and Rubin, 2002)

$$var(\hat{\beta}_{trt,i}) = \bar{T}_i + B_i \times (M+1)/M$$

   where $\bar{T}_i$ is the average of the within imputation variance of the i^{th}simulation, calculated as

$$\bar{T}_i = \frac{1}{M}\sum_{j=1}^{M}(se(\hat{\beta}^j_{trt,i}))^2$$

$\hat{\beta}^j_{trt,i}$ is the MI estimate of treatment effect of the i^{th}simulation and the j^{th} imputation. B is the between-imputation variance, calculated as

$$B_i = \frac{1}{M-1}\sum_{j=1}^{M}(\hat{\beta}^j_{trt,i} - \hat{\beta}_{trt,i})^2$$

5. 95% CI bandwidth: a method with a shorter CI bandwidth is preferred. It is estimated as the mean of the width of N simulated CIs

$$\frac{1}{N}\sum_{i=1}^{N}\left(2z_\alpha\sqrt{var(\hat{\beta}_{trt,i})}\right)$$

We performed $N = 1000$ simulations, $M = 15$ multiple imputation within each simulation and 100 Gibbs sampling within each multiple imputation. In terms of computing time, our proposed method is not excessively labor intensive. Two types of study designs are simulated: a parallel study and a 2×2 crossover study. For the parallel study, each treatment arm was assumed to have only 15 subjects and treatment effect varies from smaller effect to larger effect ($\beta = 0.5, 2.5, 5, 7.5$). The censoring percentage varies from 10% up to 60% and the missing percentage is around $6\% \sim 8\%$. For the crossover study, each sequence was assumed to have has 8 subjects and two periods with 64 observations in total. The censoring percentage and missing percentage are similar to the settings in parallel study. The subjects' random effect is assumed to be 94% of total random effects.

3.2 Simulation results

3.2.1 Parallel study

Simulation results for the parallel study are summarized in Table 1. When the treatment effect is small, all three methods demonstrate small bias of the estimates of treatment effect. And their estimates are very close. However, when treatment effect becomes larger, our proposed method outperforms the other two methods, especially when the censoring percentage becomes larger.

For the variance of the treatment effect estimate, although the conventional method has smaller variance than the two multiple imputation methods, our proposed method has the smallest mean squared error of treatment effect estimate, especially when treatment effect becomes larger and censoring becomes heavier.

Based on the definitions, for the 95% CI coverage rate it is the higher the better while for the 95% CI bandwidth it is the shorter the better. The existing method has the highest coverage rate but it also has the longest 95% CI bandwidth. Our proposed method has a little bit lower coverage rate than the existing method but its 95% CI bandwidth is much shorter than the existing method.

In summary, our proposed method is more accurate and stable than the other two methods, especially when censoring is heavier.

3.2.2 Crossover study

Simulation results of crossover study are summarized in Table 2. For illustration purpose, the corresponding graphs were in appendix. Our proposed method has the smallest bias of treatment effect estimate, especially when treatment effect is larger and percentage of censoring is higher.

For variance of treatment effect estimate, the conventional method has competitive small variance with our proposed method. But our proposed method has the smallest mean squared error, especially when censoring becomes heavier.

Although conventional method has the shortest 95% CI bandwidth, it has the worst 95% CI coverage probability. The other two MI methods have higher coverage of true treatment effect. But only our proposed method has consistently high coverage, especially when censoring is heavy.

Again, overall our proposed method outweighs the other two methods in terms of accuracy and stability, especially as the censoring becomes heavier.

4. Example

The following example was used to test the simulation results. This study was a two-part randomized, placebo-controlled, crossover trial in asthmatics, to evaluate the differential effects of three treatments, MT, ML and ND, on these markers: sputum cysteinyl leukotrienes (LTC4, LTD4, LTE4), and sputum cyotokines (IL5, IL 13, TARC and Eotaxin). The endpoints were the concentrations of these markers at 7 hours after allergen challenges. Some values of these biomarkers were either fully missing or below the limit of quantification. Table 3 lists the percentages of censoring and missing for each biomarker. Note that the censoring could be high up to 60%.

In using LTC4 as an example, to investigate whether we could find any difference in the three methods in terms of treatment comparison with placebo. The results were summarized in Table 4. For treatment ML, it was found that the three methods gave different results. Both the conventional and existing methods failed to have significant p-values ($>$ 0.1) while our proposed method was significant (p-value = 0.0668).

5. Discussion

Biomarkers are now a key component of many clinical studies. The application of biomarkers will improve decision making, accelerate drug development and reduce development costs. However, biomarker data are commonly subject to missing and left censoring. The emerging use of biomarkers in clinical studies suggests the need to address these issues, otherwise their application may be compromised. Motivated by cytokine/metabolite assay data, we proposed a multiple imputation for the analysis of biomarker data with missing and left-censored values. Simulation results show that the proposed MI method outperforms the conventional simple substitution method and the existing imputation method that does not incorporate available information, specifically in the cases of a large proportion of observations below the limit of quantification or a very small sample size.

The MI method presented in this article focused on a single measurement of biomarker. To assess the robustness of the proposed method, we applied it for the imputation of censored and missing values in a crossover study setting, and surprisingly found the MI method still performed well at least in the simulation scenarios investigated. However, further development of the MI method for handling repeated measurements in the crossover study (e.g., incorporating the correlations between repeated measurements into the imputation model by using mixed models) is worthwhile in our future studies.

Table 1: Comparison of three methods (conventional, existing and proposed) in terms of simulation results of bias, variance and mean squared error (MSE) for parallel study. β_{trt} is the true treatment effect.

β_{trt}	censor.%	Bias			Variance			MSE			95% CI Coverage			95% CI Bandwidth		
		conv.	exist.	proposed	conv.	exist.	proposed	conv.	exist.	proposed	conv.	exist.	proposed	conv.	exist.	proposed
0.5	12 %	-0.08	0.03	0.03	0.59	0.92	0.66	0.59	0.92	0.66	94.1 %	97.6 %	92.0 %	2.98	4.25	3.00
0.5	16 %	-0.09	0.00	0.03	0.56	0.97	0.65	0.57	0.97	0.65	94.5 %	97.9 %	92.2 %	3.00	4.43	3.05
0.5	21 %	-0.06	0.05	0.04	0.63	1.18	0.69	0.63	1.18	0.70	93.3 %	96.8 %	92.8 %	3.03	4.61	3.07
0.5	29 %	-0.12	-0.02	-0.02	0.67	1.49	0.79	0.68	1.49	0.79	93.4 %	95.9 %	90.4 %	3.06	4.95	3.09
0.5	36 %	-0.11	0.04	0.04	0.64	1.96	0.82	0.65	1.96	0.82	94.3 %	96.0 %	90.3 %	3.20	5.38	3.15
2.5	9 %	-0.33	0.07	0.05	0.64	0.95	0.67	0.75	0.95	0.67	91.0 %	96.5 %	91.9 %	3.08	4.22	3.01
2.5	12 %	-0.41	0.00	-0.02	0.61	0.92	0.67	0.78	0.92	0.67	91.8 %	96.7 %	92.8 %	3.06	4.27	3.00
2.5	17 %	-0.41	0.08	0.04	0.64	1.12	0.72	0.80	1.12	0.72	91.2 %	96.9 %	91.7 %	3.09	4.51	3.04
2.5	23 %	-0.47	0.06	0.00	0.64	1.31	0.70	0.86	1.31	0.70	91.2 %	97.4 %	92.7 %	3.14	4.76	3.08
2.5	29 %	-0.47	0.12	0.06	0.68	1.60	0.76	0.90	1.61	0.76	90.1 %	96.7 %	91.1 %	3.23	5.04	3.11
2.5	43 %	-0.68	-0.22	-0.08	0.76	2.18	0.99	1.23	2.23	1.00	88.8 %	94.8 %	87.5 %	3.51	5.59	3.15
5	8 %	-0.63	0.03	-0.03	0.62	0.91	0.64	1.02	0.91	0.64	88.2 %	97.2 %	93.9 %	3.16	4.18	3.01
5	9 %	-0.62	0.05	0.01	0.63	0.93	0.64	1.01	0.94	0.64	89.3 %	96.4 %	92.3 %	3.17	4.22	2.98
5	14 %	-0.65	0.11	0.05	0.63	1.05	0.65	1.04	1.07	0.66	88.8 %	96.9 %	91.5 %	3.20	4.41	3.03
5	18 %	-0.75	0.03	-0.02	0.68	1.21	0.72	1.24	1.21	0.72	85.0 %	96.9 %	91.5 %	3.23	4.58	3.05
5	29 %	-0.86	0.00	-0.04	0.81	1.75	0.88	1.56	1.75	0.88	82.8 %	96.2 %	88.3 %	3.45	5.08	3.11
5	40 %	-1.16	-0.56	-0.21	0.96	2.09	1.01	2.32	2.40	1.05	77.6 %	90.2 %	84.4 %	3.79	5.46	3.14

Table 2: Comparison of three methods (conventional, existing and proposed) in terms of simulation results of bias, variance and mean squared error (MSE) for 2 × 2 crossover study. β_{trt} is the true treatment effect.

β_{trt}	censor.%	Bias			Variance			MSE			95% CI Coverage			95% CI Bandwidth		
		conv.	exist.	proposed	conv.	exist.	proposed	conv.	exist.	proposed	conv.	exist.	proposed	conv.	exist.	proposed
2	10 %	-0.25	0.04	-0.02	0.07	0.09	0.06	0.13	0.1	0.06	87.6 %	99.9 %	99.8 %	1.08	1.72	1.31
2	16 %	-0.29	0.01	-0.01	0.09	0.15	0.10	0.17	0.15	0.10	83.1 %	99.8 %	98.8 %	1.14	1.96	1.53
2	24 %	-0.33	-0.10	-0.05	0.10	0.18	0.11	0.21	0.19	0.11	83.8 %	99.1 %	99.1 %	1.24	2.2	1.78
2	33 %	-0.39	-0.31	-0.08	0.15	0.29	0.18	0.30	0.39	0.19	81.1 %	97.0 %	98.4 %	1.47	2.55	2.15
2	41 %	-0.46	-0.56	-0.15	0.19	0.47	0.26	0.41	0.78	0.29	80.9 %	90.8 %	97.3 %	1.73	2.88	2.49
2	54 %	-0.68	-1.17	-0.40	0.27	0.79	0.32	0.73	2.15	0.48	83.5 %	77.1 %	96.5 %	2.2	3.61	3.23
3	9 %	-0.29	0.04	0.01	0.07	0.08	0.06	0.15	0.09	0.06	84.8 %	99.4 %	99.1 %	1.05	1.52	1.31
3	15 %	-0.37	0.00	-0.02	0.08	0.12	0.09	0.22	0.12	0.09	75.6 %	99.3 %	99.3 %	1.12	1.74	1.50
3	23 %	-0.42	-0.07	-0.04	0.09	0.21	0.12	0.27	0.21	0.12	73.8 %	98.4 %	98.8 %	1.21	1.98	1.77
3	33 %	-0.49	-0.29	-0.07	0.11	0.22	0.17	0.35	0.31	0.17	71.1 %	95.4 %	98.8 %	1.37	2.23	2.10
3	42 %	-0.60	-0.65	-0.16	0.17	0.41	0.21	0.53	0.83	0.24	65.4 %	81.8 %	97.9 %	1.57	2.56	2.49
3	51 %	-0.82	-1.28	-0.36	0.22	0.54	0.22	0.89	2.17	0.35	56.9 %	53.9 %	97.0 %	1.83	2.91	2.89
4	10 %	-0.36	0.02	0	0.06	0.08	0.07	0.19	0.08	0.07	78.6 %	99.1 %	98.8 %	1.07	1.42	1.31
4	16 %	-0.44	-0.01	-0.02	0.08	0.12	0.09	0.27	0.12	0.09	68.9 %	99.0 %	99.0 %	1.11	1.63	1.51
4	23 %	-0.50	-0.05	-0.03	0.09	0.19	0.12	0.33	0.19	0.12	68.0 %	97.8 %	99.0 %	1.24	1.86	1.76
4	33 %	-0.57	-0.28	-0.1	0.12	0.31	0.14	0.44	0.39	0.15	69.9 %	89.6 %	98.5 %	1.44	2.12	2.11
4	41 %	-0.77	-0.80	-0.28	0.19	0.42	0.17	0.78	1.05	0.25	61.6 %	67.1 %	97.6 %	1.73	2.38	2.50

Table 3: The percentages of censoring and missing of each biomarker in a real study.

Cytokine	%censoring (#censoring/N)	%missing (#missing/N)
LTC4	61%(39/64)	8%(5/64)
LTD4	44%(28/64)	8%(5/64)
LTE4	5%(3/64)	8%(5/64)
IL5	8%(5/64)	5%(3/64)
IL13	34%(22/64)	6%(4/64)
TARC	28%(18/64)	5%(3/64)
Eotaxin	25%(16/64)	6%(4/64)

Table 4: Comparison of three methods in a real study, in terms of fold change of LTC4 concentration over placebo.

Treatment	Method	Geometric Mean of fold change over Placebo	90% CI for Geometric mean of fold change over Placebo	P-Value 2-sided[@]	Effect Size
MT	Conventional	0.87	(0.60,1.27)	0.5400	-0.22
MT	Existing	1.00	(0.44,2.29)	0.9973	-0.06
MT	Proposed	0.64	(0.34,1.21)	0.2520	-0.53
ML	Conventional	0.75	(0.40,1.42)	0.4448	-0.41
ML	Existing	0.75	(0.28,2.00)	0.6283	-0.32
ML	Proposed	0.44	(0.21,0.92)	0.0668	-0.71
ND	Conventional	1.01	(0.64,1.61)	0.9571	-0.02
ND	Existing	1.23	(0.50,3.05)	0.6991	-0.29
ND	Proposed	0.90	(0.44,1.84)	0.8000	-0.11
@Not adjusted for simultaneous multiple comparisons					

REFERENCES

Chu, Haitao and Nie, Lei and Kensler, Thomas W. (2008), "A Bayesian approach estimating treatment effects on biomarkers containing zeros with detection limits," *Statistics in Medicine* , 27, 2497-2508.

Cooper, G. S., Savitz, D. A., Millikan, R. and Chiu Kit, T. (2002), "Organochlorine exposure and age at natural menopause," *Epidemiology* , 13, 729-1733.

Guo, Y., Harel, O. and Little, R.J. (2010). "How Well Quantified is the Limit of Quantification?" *Epidemiology*, 21(4), S10-S16.

Helsel, Dennis R. (2004), *Nondetects and Data Analysis: Statistics for Censored Environmental Data*, New York, USA: Wiley.

Hopke, P. K., Liu, C. and Rubin, D. B. (2001), "Multiple imputation for multivariate data with missing and below-threshold measurements," *Biometrics* , 57, 22–33.

Lee, M., Kongb, L. and Weissfeld, L. (2012), "Multiple imputation for left-censored biomarker data based on Gibbs sampling method," *Statistics in Medicine* , 31, 1838-1848.

Little, R.J. and Rubin, D.B. (2002) *Statistical Analysis with Missing Data*, Wiley: Hoboken, NJ, Second Edition.

Liu, G., Hu, P. and Mehrotra, D. V. "A two-step multiple imputation for analysis of repeated measures with missing data and left truncation". Submitted to Pharmaceutical Statistics.

Lubin, J. H., Colt, J. S., Camann, D., Davis, S., Cerhan, J. R., Severson, R. K., Bernstein, L. and Hartge P. (2004), "Epidemiologic evaluation of measurement data in the presence of detection limits," *Environmental Health Perspectives* , 112(17), 1691–1696.

Damien, P. and Walker, G. (2001), "Sampling Truncated Normal, Beta, and Gamma Densities," *Journal of Computational and Graphical Statistics*, Volume 10, Number 2, 206-215.

Singh, A. and Nocerino, J. (2002), "Robust estimation of mean and variance using environmental data sets with below detection limit observations," *Chemometrics and Intelligent Laboratory Systems* , 60, 69–86.

Uh, H., Hartgers, H.F., Yazdanbakhsh, M. and Houwing-Duistermaat, J. (2008) Evaluation of regression methods when immunological measurements are constrained by detection limits. *BMC Immunology*, 9:59.

6. Appendix

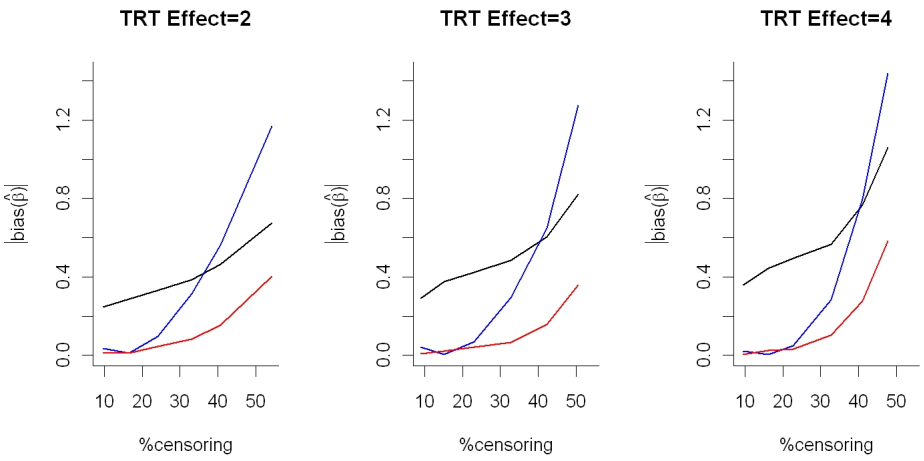

Figure 1: Comparison of three methods in terms of bias of treatment effect estimate for the crossover study. Black line - conventional method; blue line - existing method; red line - proposed method.

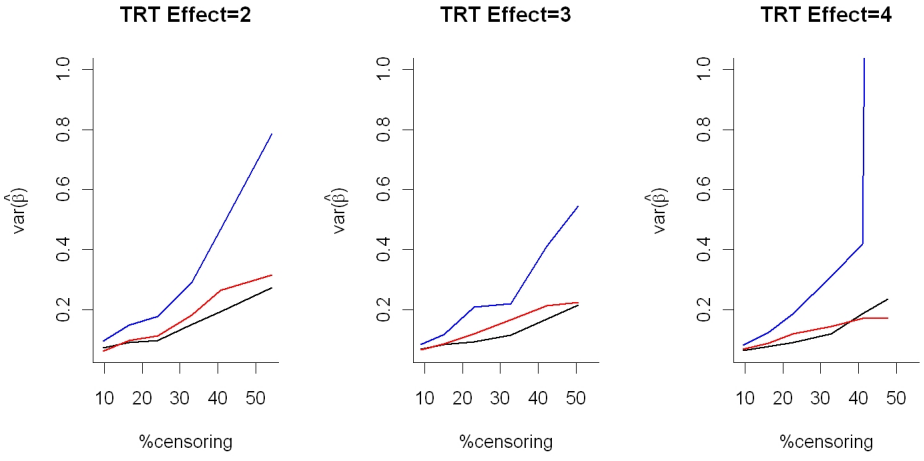

Figure 2: Comparison of three methods in terms of variance of treatment effect estimate for the crossover study. Black line - conventional method; blue line - existing method; red line - proposed method.

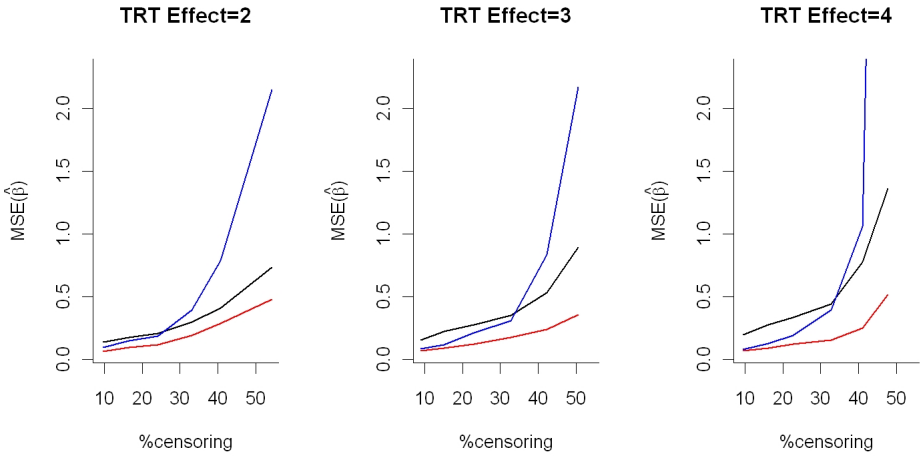

Figure 3: Comparison of three methods in terms of mean squared error (MSE) of treatment effect estimate for the crossover study. Black line - conventional method; blue line - existing method; red line - proposed method.

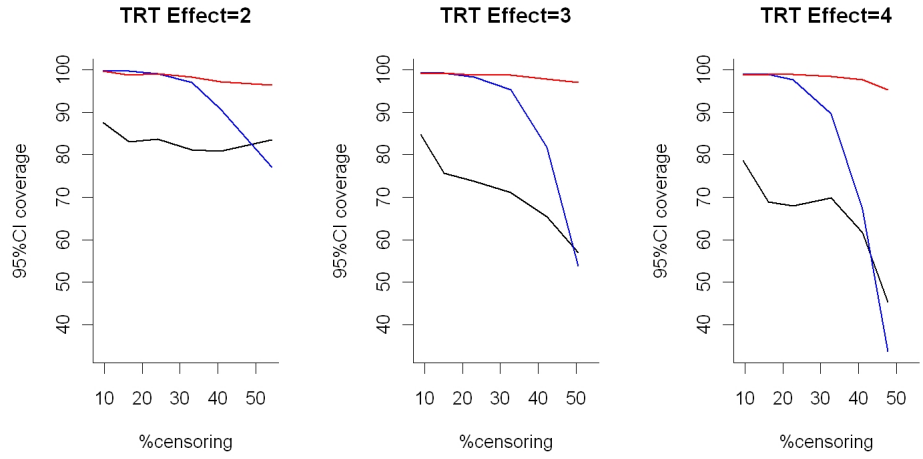

Figure 4: Comparison of three methods in terms of 95% confidence interval coverage for the crossover study. Black line - conventional method; blue line - existing method; red line - proposed method.

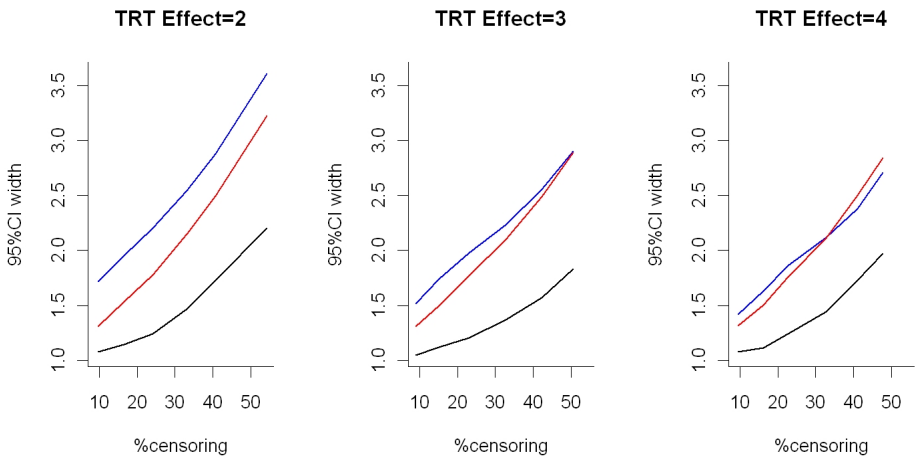

Figure 5: Comparison of three methods in terms of 95% confidence interval bandwidth for the crossover study. Black line - conventional method; blue line - existing method; red line - proposed method.