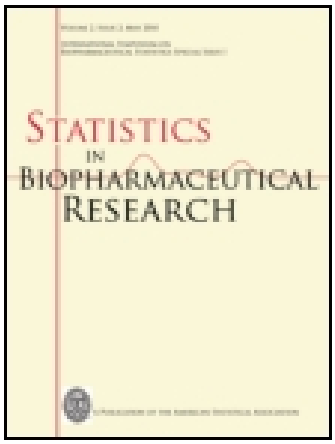
Adjusting for Baseline on the Analysis of Repeated Binary Responses With Missing Data

Peer-reviewed author version

Jiang, Honghua; Kulkarni, Pandurang M.; Mallinckrodt, Craig H.; Shurzinske, Linda; MOLENBERGHS, Geert & Lipkovich, Ilya (2015) Adjusting for Baseline on the Analysis of Repeated Binary Responses With Missing Data. In: STATISTICS IN BIOPHARMACEUTICAL RESEARCH, 7 (3), p. 238-250.

DOI: 10.1080/19466315.2015.1067251
Handle: http://hdl.handle.net/1942/19800



Statistics in Biopharmaceutical Research
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/usbr20

Adjusting for Baseline on the Analysis of Repeated Binary Responses with Missing Data

Honghua Jiang[a], Pandurang M Kulkarni[a], Craig H Mallinckrodt[a], Linda Shurzinske[a], Geert Molenberghs[b] & Ilya Lipkovich[c]

[a] Lilly Research Labs. Eli Lilly and Co. Indianapolis, IN. 46285.

[b] I-BioStat, Hasselt University, Diepenbeek, Belgium, and I-BioStat, Katholieke Universiteit Leuven, Leuven, Belgium

[c] Quintiles

Accepted author version posted online: 09 Jul 2015.

To cite this article: Honghua Jiang, Pandurang M Kulkarni, Craig H Mallinckrodt, Linda Shurzinske, Geert Molenberghs & Ilya Lipkovich (2015): Adjusting for Baseline on the Analysis of Repeated Binary Responses with Missing Data, Statistics in Biopharmaceutical Research, DOI: 10.1080/19466315.2015.1067251

To link to this article: http://dx.doi.org/10.1080/19466315.2015.1067251

Adjusting for Baseline on the Analysis of Repeated Binary Responses with Missing Data

Honghua Jiang[1], Pandurang M Kulkarni[1], Craig H Mallinckrodt[1], Linda Shurzinske[1], Geert Molenberghs[2] and Ilya Lipkovich[3]

1. Lilly Research Labs. Eli Lilly and Co. Indianapolis, IN. 46285.
2. I-BioStat, Hasselt University, Diepenbeek, Belgium, and I-BioStat, Katholieke Universiteit Leuven, Leuven, Belgium
3. Quintiles

Abstract

Little research has been done to evaluate the effect of adjusting for baseline in the analysis of repeated incomplete binary data through simulation study. In this article, covariate adjusted and unadjusted implementations of the following methods were compared in analyzing incomplete repeated binary data when the outcome at the study endpoint is of interest: logistic regression with the last observation carried forward (LOCF), generalized estimating equations (GEE), weighted GEE (WGEE), generalized linear mixed model (GLMM), and multiple imputation with analyses via GEE (MI). Incomplete data mimicking several clinical trial scenarios were generated using missing completely at random (MCAR), missing at random (MAR), and missing not at random (MNAR) mechanisms. Across the various analytic methods and scenarios covariate adjusted analyses generally yielded larger, less biased treatment effect estimates and larger standard errors compared with their unadjusted counterpart. The net result of these factors was increased power from the covariate adjusted analyses without increasing type I error rates. Although all methods were biased in at least some of the MNAR scenarios, the type I error rates from LOCF exceeded 20% whereas the highest rate from any other method in any scenario was less than 10%. LOCF also yielded biased results in MCAR and MAR data whereas the other methods were not biased or had smaller biases than LOCF. These results support longitudinal modeling of repeated binary data over LOCF logistic regression of the study endpoint only. These results also support covariate adjustment for baseline severity in these longitudinal models.

Key words: GEE, GLMM, multiple imputation, weighted analysis

1. **Introduction**

Binary outcomes derived from underlying continuous measures are commonly evaluated in randomized clinical trials to compare treatment effect. For example, in diabetes clinical trials glycated haemoglobin (HbA1c) is a continuous measure that reflects average plasma glucose for several months (Sacks et al., 2002). It is commonly used as the primary efficacy outcome measure. However, a clinically meaningful outcome is whether the endpoint HbA1c reaches the target of <7.0% (ADA, 2013). Therefore, comparing the proportion of patients reaching HbA1c target of <7.0% between treatments based on a binary outcome is an important objective in diabetes trials.

In longitudinal clinical trials patients are treated over a period of time and are evaluated at multiple time points. Usually, the primary efficacy evaluation is based on the measurement at the last scheduled time point. However, not all the patients complete the study. Some patients may withdraw before completing the study, and the measurement at the last scheduled time point will be missing. Missing data occur commonly in longitudinal studies for various reasons, including lack of efficacy, safety, re-location, etc. A historically common approach to handle missing data was to impute the missing observations with the last available observation of the patient, i.e., the last observation carried forward (LOCF) method. However, LOCF requires restrictive assumptions that are unlikely to hold in practice, and this approach is generally not acceptable (NRC, 2010)

Other analytic approaches for repeated binary data that do not require imputation include generalized estimating equations (GEE) (Liang and Zeger, 1986; Zeger and Liang, 1986). Under

a missing completely at random mechanism (MCAR) (Rubin, 1976), GEE provides unbiased and consistent parameter estimates even when the working correlation matrix is mis-specified. However, under the missing at random mechanism (MAR) (Rubin, 1976) parameter estimates based on GEE can be biased. A weighted generalized estimating equations approach (Robins et al. 1995; and Fitzmaurice et al. 1995) extends conventional GEE and provides consistent parameter estimates under MAR when the dropout model is correctly specified. In this approach, an individual's contribution is weighted by the inverse probability of dropout at the given time.

Likelihood-based generalized linear mixed model (GLMM) analyses of the available cases have also been widely used for the analysis of repeated binary data, generally under the assumption of MAR although some approximations to direct-likelihood require the more stringent MCAR assumption (Wolfinger and O'Connell, 1993).

Multiple imputation approaches (Rubin, 1987; Schafer, 1999; Shieh, 2003; Li et al., 2006; Carpenter and Kenward, 2013; and van Buuren, 2014) are commonly applied to continuous incomplete longitudinal data and can therefore be used to impute the continuous outcome from which the binary responses are derived without some of the restrictive assumptions that limit LOCF. After imputation, the resulting complete data sets can be analyzed with either GEE or likelihood-based methods.

Lipkovich et al. (2005) compared the performance of MI followed by GEE analysis with GLMM and GEE analyses of available cases in estimating treatment differences for binary outcomes derived from underlying continuous responses. The MI-based approach performed better than

GEE and GLMM in terms of precision, power, and type I error rate under MAR. However, under the missing not at random (MNAR) mechanism, all three methods yielded biased results. Liu and Zhan (2011) also conducted simulations to compare GLMM, GEE, and several MI approaches for the analysis of repeated binary responses with missing data in evaluating the treatment effect at study endpoint. Results indicated that GLMM performed better than GEE and MI approaches in terms of controlling type I error rate under MAR.

In many clinical settings the baseline severity is linked to the probability of achieving a target level of symptom severity. For example, in diabetes baseline HbA1c levels influence whether the patient can achieve the HbA1c target of <7.0%. Therefore, adjusting for baseline severity may improve analytic performance in these situations. However, little research has been done to evaluate the effect of adjusting for initial disease severity in the analysis of repeated binary data with missing values. The present research evaluates the performance of the covariate adjusted and unadjusted analyses of repeated binary outcomes in terms of type I error rate, power, precision, and bias across several common statistical approaches. Focus is on evaluating the treatment effect at the study endpoint. We note that by the treatment effect we aim at the "efficacy estimand" that is the treatment effect, assuming had all patients who discontinued the trial would have completed the study while complying with the protocol.

2. **Statistical Methods**

For the analysis of the single outcome at study endpoint without repeated measures, logistic regression with LOCF was used. For repeated binary outcome analyses GEE, WGEE, GLMM,

and MI approaches were used, and the study endpoint contrast was derived from the repeated measures analysis. Details of each analysis are described below.

Logistic regression:

For the analysis of a single binary outcome at study endpoint, let Y_i=1 represent that the *i*th patient achieved the HbA1c target of <7.0% at the study endpoint, and Y_i=0 otherwise. Let $\boldsymbol{\beta}$ be a vector of regression coefficients; $\boldsymbol{X}_i$ be a vector of covariates such as treatment indicator and baseline HbA1c; and p_i = pr(Y_i=1| $\boldsymbol{X}_i$, β)=E (Y_i | $\boldsymbol{X}_i$, $\boldsymbol{\beta}$). Then the logistic regression model can be expressed as:

$$\text{logit}\,(p_i) = \log\left(\frac{p_i}{1-p_i}\right) = \boldsymbol{X_i'}\,\boldsymbol{\beta} \qquad (1)$$

With LOCF there is no missing data and the key assumption is that patients' observations would not have changed had they stayed in the trial. Logistic regression was implemented in the present study using SAS PROC GENMOD (SAS, 2008).

Generalized Estimating Equations (GEE)

For the analysis of a repeated binary outcome, let Y_{it}=1 represent that the *i*th patient achieved HbA1c target of <7.0% at time *t*, and Y_{it} =0 otherwise; $\boldsymbol{X}_{it}$ be the vector of covariates for the *i*th patient at time *t*; let $\boldsymbol{Y_i} = (Y_{i1}, Y_{i2}, \ldots, Y_{iT})'$, $\boldsymbol{X}_i = (\boldsymbol{X}_{i1}, \boldsymbol{X}_{i2}, \ldots, \boldsymbol{X}_{iT})$, and T be the number of scheduled study visits at which the data are collected. Let $\boldsymbol{\beta}$ be the vector of regression coefficients; and $\boldsymbol{p}_i(\boldsymbol{\beta})$= pr($\boldsymbol{Y}_i$=1| $\boldsymbol{X}_i$, $\boldsymbol{\beta}$)=E ($\boldsymbol{Y}_i$ | $\boldsymbol{X}_i$, $\boldsymbol{\beta}$). Then, the GEE proposed by Liang and Zeger (1986) and Zeger and Liang (1986) takes the form

$$\mathrm{U}(\boldsymbol{\beta}) = \sum_{i=1}^{N} \boldsymbol{D'}_i \boldsymbol{V}_i^{-1}\left(\boldsymbol{Y_i} - \boldsymbol{p_i}(\boldsymbol{\beta})\right) = \boldsymbol{0} \qquad (2)$$

where $\boldsymbol{D}_i = \partial \boldsymbol{p_i}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}$ and $\boldsymbol{V_i}$ is a "working" covariance matrix. SAS PROC GENMOD (SAS, 2008) was used to implement GEE in the present study. The key missing data assumption for GEE is that the missing data arise from a MCAR mechanism (Molenberghs and Kenward, 2007).

Weighted Generalized Estimating Equations (WGEE)

The WGEE method proposed by Robins et al. (1995) and Fitzmaurice et al. (1995) is less restrictive than standard GEE in that the key missing data assumption is that the missing data arise from an MAR mechanism - given that the probabilities of dropout for each subject are correctly specified (Molenberghs and Kenward, 2007). Fitzmaurice et al. (1995) provided a formulation of the WGEE method based on the article by Robins et al. (1995). Under this approach, equation (2) is modified as:

$$\mathrm{U}(\boldsymbol{\beta}) = \sum_{\boldsymbol{i=1}}^{\boldsymbol{N}} \frac{\mathbf{1}}{\boldsymbol{\pi_i}} \boldsymbol{D'}_{\boldsymbol{i}} \boldsymbol{V_i^{-1}} \big(\boldsymbol{Y_i} - \boldsymbol{p_i}(\boldsymbol{\beta})\big) = \mathbf{0} \qquad (3)$$

where $\boldsymbol{\pi_i}$ is the estimated probability of *i*th patient following the dropout pattern that was observed for that patients. That is, for a patient who discontinued at visit *d*, $\boldsymbol{\pi_i}$ is the probability of patient's remaining in the study through visit (*d*-1) and discontinuing at visit *d*; for a patient who had completed the trial, $\boldsymbol{\pi_i}$ is the probability of that patient remaining in the study through the last scheduled visit. Therefore an individual's contribution in estimating equation is weighted by the inverse probability of observing patient's discontinuation pattern. In the present study WGEE was implemented using PROC GENMOD (SAS, 2008) to estimate the dropout probabilities and incorporate the weightings and conduct analysis.

Generalized Linear Mixed Model (GLMM)

The GLMM extends the generalized linear model by incorporating normally distributed random parameters for individual subjects (Breslow and Clayton, 1993) under the assumption of MAR. The fixed-effect inference is conditional on random parameters and has a subject-specific interpretation. The form of GLMM based on the logit link function to fit the response probability, p_{it}, at time point t is as

$$\text{logit}\,(p_{it}) = \boldsymbol{X}'_{\boldsymbol{it}}\,\boldsymbol{\beta} + \boldsymbol{Z}'_{\boldsymbol{it}}\,\boldsymbol{b}_{\boldsymbol{i}} \qquad (4)$$

where $\boldsymbol{\beta}$ is the vector of the fixed-effect parameters, $\boldsymbol{b}_{\boldsymbol{i}}$ is the vector of random subject parameters and $\boldsymbol{b}_{\boldsymbol{i}} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{V})$, and $\boldsymbol{X}_{\boldsymbol{it}}$ and $\boldsymbol{Z}_{\boldsymbol{it}}$ and the vectors of known covariates. In the present study, GLMM was implemented using PROC GLIMMIX in SAS (SAS, 2008).

Multiple Imputation (MI):

MI is an extension of single imputation where the missing data are imputed several times, say m times. Then each of the m complete data sets is analyzed with standard methods, and the results of the m analyses are combined according to Rubin's rule [Rubin, 1987]. MI usually assumes an MAR mechanism. The following process was implemented for the multiple imputation method.

1. Bayesian regression, which included earlier values as predictors, was used to impute the missing data with a separate predictive distribution for each treatment group [Rubin, 1987]. The missing continuous HbA1c values were imputed first with the SAS MI procedure (SAS, 2008). Next, the continuous outcomes were dichotomized into binary response data according to whether they were <7.0% or not. We generated m=30 complete data sets.

2. The GEE model as previously described was used to analyze each of the 30 imputed complete data sets for the repeated binary response, and 30 sets of parameter estimates were obtained.

3. The 30 estimates and associated standard errors were combined into the final estimates with SAS PROC MIANALYZE (SAS, 2008) according to Rubin's rule.

3. **Simulation**

3.1 Simulation setting

The continuous HbA1c values (%) at each visit were simulated based on a multivariate normal distribution with mean profiles (Table 1) and variance-covariance matrix with elements $\sigma_{i,j}=\rho\ \sigma_i\ \sigma_j$, where ρ was the correlation between the repeated outcomes. The binary outcome was constructed based on whether the HbA1c value was <7.0% or not. Data were simulated based on inputs obtained from actual diabetes clinical trials. A compound-symmetry correlation matrix was used, with $\rho = 0.5$, and the variance (σ_i^2) increasing over time from visit 1 (at baseline) to visit 4 (1.0, 1.0, 1.2, 1.4). The sample size of 50 and 200 per treatment group were used to mimic phase 2 and phase 3 trial settings, respectively.

Incomplete data sets were then generated from the complete data sets using 3 rates of missing data and 3 missing data mechanisms. Rates of missing data were either 45% in both treatments groups or 25% in one treatment group and 45% in the other. Missingness mechanisms were either MCAR, MAR, or MNAR. In MCAR the outcomes did not differ for patients that completed compared with those who dropped out. In MAR the probability of a value (y_i) being missing depended on the observed value at the previous visit (y_{i-1}), expressed as logit (p (y_i

missing | y_{i-1})= $a + b * y_{i-1}$. The value b =1.5 was used in the dropout model, and the value a was chosen for each treatment group to achieve the desired rates of missing data. In MNAR the probability of a value (y_i) being missing depended on the value itself (y_i), expressed as

logit (p (y_i missing | y_i)= $a + b * y_i$. The value b =0.4 was used in the dropout model, and the value a was chosen for each treatment group to achieve the desired rates of missing data. For simplicity, only monotone missingness was considered. For each scenario 2000 data sets were simulated.

The performance of different analysis methods was evaluated based on bias ($\hat{\beta}$ -β_T) in scenarios when there was no difference between treatments at endpoint, relative bias ($\frac{\hat{\beta} - \beta_T}{\beta_T} \times 100$) in scenarios where treatments did differ at endpoint. Methods were also compared based on 95% confidence interval (CI) coverage (using normal theory approximation), standard errors (SE) (average of SEs from the 2000 simulations), and type I error rate for scenarios with no difference between treatments and power for scenarios where treatments differed. $\hat{\beta}$ is the estimate of log odds ratio for unadjusted or adjusted analysis. β_T is the "true" log odds ratio for the unadjusted or adjusted analysis which is based on estimates from the values obtained by averaging results from the corresponding complete data sets.

3.2. **Simulation Results**

Tables 2 and 3 show the simulation results for different methods: covariate adjusted/ unadjusted logistic regression (Adj.logit/Unadj.logit) with LOCF method, covariate adjusted/unadjusted generalized estimating equations (Adj.GEE/Unadj.GEE), covariate adjusted/unadjusted weighted

generalized estimating equations (Adj.WGEE/Unadj.WGEE), covariate adjusted/unadjusted generalized linear mixed model (Adj.GLMM/Unadj.GLMM), and covariate adjusted/unadjusted generalized estimating equations on the basis of multiple imputation (Adj.MI/ Unadj.MI).

Under MCAR, LOCF appreciably inflated type I error when rates of missing data differed between treatments. All analyses yielded the anticipated nominal type I error rate when rates of missing data were equal for the two treatments. MI analyses yielded the lowest type I error rate (<3%).

All analyses yielded relatively unbiased estimates with the absolute bias <0.06 for the no treatment effect case, and relative bias ≤16% for the strong treatment effect case, except for LOCF. LOCF had biases 7-18 times greater than other methods for the no treatment effect case with unequal missing proportions and relative bias up to 35% for the strong treatment effect case. All analyses produced CIs with coverage close to their nominal level except for LOCF analysis which was associated with coverage lower than the nominal level, with the MI analysis exceeding the nominal level. For the strong treatment effect case, adjusted analyses yielded greater, less biased treatment effect estimates but also larger SEs compared to the unadjusted analyses, resulting in increased power with the large sample size. However, slightly lower power of the adjusted analyses was observed with the small sample size.

Under MAR, Unadj.GEE, Unadj.GLMM, WGEE, and LOCF methods inflated type I error with unequal missing data rates. All analyses preserved type I error rate with the equal missing proportion case . MI analyses had the lowest type I error rate, <4%. All analyses yielded relatively unbiased estimates with the absolute bias ≤0.07 for the no treatment effect case, except

for LOCF, Unadj.GLMM and WGEE. However, for the strong treatment effect case LOCF, GLMM, and WGEE yielded significantly biased estimates with relative bias up to 66.1%. In general, adjusted analyses had less bias than their unadjusted counterparts. All analyses produced CIs with coverage close to the nominal level except for LOCF, Unadj.GLMM, and WGEE, which yielded coverage lower than the nominal level, while MI analysis exceeded their nominal level. For the strong treatment effect case, adjusted analyses yielded greater, less biased treatment effect estimates and SEs compared to the unadjusted analyses, but resulted in increased power in general.

Under MNAR, all analyses inflated type I error rate except for MI and WGEE. For the strong treatment effect case with equal missing data rates all analyses except for LOCF and Unadj.GLMM produced relatively unbiased estimates (relative bias <=11.7%). However, for the strong treatment effect case with unequal missing data rates all analyses except for WGEE yielded significantly biased estimates with relative bias up to 37.9%.

5. Clinical Trial Examples

The methods tested in the simulation study were also applied to data from five diabetes clinical trials in which an active drug (treatment) was compared with different comparators over 52 weeks. The missing data proportions for the treatment and comparators were 15.1%, 23.9%; 12.8%, 12.2%; 22.7%, 26.9%; 23.4%, 20.3%; and 19.4%, 20.0% for study 1 to 5, respectively. The proportions of patients who achieved HbA1c target of <7.0% at the final visit for the treatment and comparators were 73%, 55%; 54%, 31%; 64%, 53%; 61%, 48%; and 64%, 38% for study 1 to 5, respectively. The analyses were conducted on proportion of patients whose 52-

week endpoint HbA1c was <7.0%. The LOCF analyses were implemented using logistic models with a factor for treatment only (Unadj.Logit) and with baseline HbA1c (Adj.Logit). The models for the GEE, WGEE, GLMM, and MI analyses included treatment, visit, and treatment-by-visit interaction, with or without baseline HbA1c, and baseline HbA1c-by-visit interaction. An unstructured covariance matrix or working correlation matrix was used except for WGEE where an independent working correlation matrix was used which was considered as a best fit (Fitzmaurice et al. 1995). Results are displayed in Table 4 and Figure 1. As in the simulated data, covariate adjusted methods yielded larger treatment effects compared to their unadjusted counterparts; and standard errors were larger and p values smaller from adjusted analyses for all studies except for study 4, where adjusted and unadjusted methods yielded comparable effects and standard errors. The possible reason is that that the treatment group had a slightly lower mean baseline HbA1c value compared to the comparator group in study 4, while the baseline HbA1c was slightly higher in the treatment group or close to the comparator in other studies. The differences (treatment-comparator) in baseline HbA1c were -0.02%, 0.07%, 0.04%, -0.04%, and 0.12% for study 1 to 5, respectively.

6. Summary and Discussion

In longitudinal clinical trials it is often reasonable to assume missing data arise from an MAR mechanism (Verbeke and Molenberghs, 2000; NRC, 2010). Dichotomized continuous outcomes or other binary measures are often of interest. Therefore, appropriate modeling of incomplete longitudinal binary clinical trial data is important.

Longitudinal binary data analyses with incomplete measures have been studied extensively in the literature (Albert, 1999; Cnaan et al. 1997; Fitzmaurice et al. 1994). However, little research has been done to evaluate the effect of adjusting for baseline in the analysis of repeated incomplete binary data through simulation study. Simulations were conducted over a variety of scenarios in this article to examine how including or not including baseline severity as a covariate influenced results across several methods for analyzing incomplete longitudinal binary responses. The covariate adjusted analyses generally yielded larger, less biased treatment effect estimates and larger standard errors compared with their unadjusted counterpart. The net result of these factors was increased power from the adjusted analyses without increasing type I error rate. Results from five phase 3 diabetes trials were consistent with the simulation findings.

In regards to how the various methods handled missing data, with MNAR data all methods except for WGEE yielded biased results in at least some scenarios and all methods except MI and WGEE inflated type I error rates in some scenarios. With MCAR data LOCF yielded biased results, inflated type I error rates and had poor CI coverage, whereas results from other methods were not biased. With MAR data, LOCF again yielded biased results. GLMM also yielded biased results. MI yielded unbiased results, as expected, because it assumes MAR. Results from GEE were relatively unbiased. Counter to expectation, results from WGEE were biased in the MAR scenarios where dropout rates differed but were not biased when dropout rates were equal. Recall the description in section 3.1 of the dropout mechanisms applied to delete data.

> *In MAR the probability a value (y_i) being missing depended on the observed value at the previous visit (y_{i-1}), expressed as logit (p (y_i missing | y_{i-1})= $a + b * y_{i-1}$. The value b =1.5 was used in the dropout model and the value a was chosen for each treatment group to achieve the desired rates of missing data.*

Thus, when the dropout rate was equal in the two treatments they shared a common intercept (a), but when dropout rates differed the intercepts differed. That is, there were separate dropout models for each treatment. However, in the WGEE analyses a single model was used to estimate the weights for both treatment groups. Therefore, in scenarios where the dropout rates differed, the model used to estimate the weights was not the same as the model used to generate the missing data. In addition, for MAR logit (p (y_i missing | y_{i-1})= $a + b * y_{i-1}$.

However, the weights were estimated using logit (p (y_i missing | y_{i-1})= $a + b * y_i$.

The issue of having the same or separate models by treatment group applies to the imputation model in MI. In the present study MI was implemented with separate imputation models for each treatment, thereby accommodating different dropout models for each treatment in the scenarios where dropout rates differed.

These results illustrate the potential importance of modeling considerations in the handling of missing data. For methods that explicitly impute missing values or model dropout, it may be useful to consider separate models for each treatment or for groups of treatments (e.g., all doses of a drug in one group, placebo in the other), especially when rates, timing, and/or reasons for dropout differ.

Although these modeling considerations are important, they should not be taken as motivation to return to the use of ad hoc methods. For example, in the MAR scenarios where WGEE yielded type I error rates of 6.1% – 6.9%, the corresponding rates from LOCF were 20.6% - 23.5%.

In this article, we only allowed missingness to depend on outcomes prior to (MAR) or at the current visit (MNAR) and monotone missingness. Evaluation of model performance under the scenarios where missingness depends on baseline covariates, and/or with non-monotone missingness is a future research topic.

The present investigation focused on only one of the many modeling considerations, fitting a single covariate that describes baseline severity. Results support including baseline severity as a covariate in analyses of incomplete longitudinal binary outcomes.

References

ADA, (2013), "Standards of medical care in diabetes," *Diabetes Care*, 36(Suppl 1):s11–66.

Albert PS. (1999), "Longitudinal data analysis (repeated measures) in clinical trials," *Statistics in Medicine*, 18: 1707 – 1732.

Breslow, N. E., and Clayton, D. G. (1993), "Approximate inference generalized linear mixed model," *Journal of the American Statistical Association*, 88:9–25.

Carpenter, J. R. and Kenward, M. G. (2013), "Multiple Imputation and its Application," John Wiley & Sons, Chichester

Cnaan A, Laird N. M, Slasor P. (1997), "Tutorial in biostatistics: using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data," *Statistics in Medicine,* 16:2349–80.

Fitzmaurice, G. M., Laird, N. M., and Lipsitz, S. R. (1994), "Analysing incomplete longitudinal binary responses: a likelihood-based approach," *Biometrics* 50, 601-612.

Fitzmaurice, G. M., Molenberghs, G., and Lipsitz, S. R. (1995), "Regression models for longitudinal binary responses with informative dropouts," *Journal of the Royal Statistical Society Series B,* 57:691–704.

Liang, K.Y., and Zeger, S. L. (1986), "Longitudinal data analysis using generalized linear models," *Biometrika*, 73, 13-22.

Li, X. M., Mehrotra, D. V., and Barnard, J. (2006), "Analysis of incomplete longitudinal binary data using multiple imputation," *Statistics in Medicine,* 25:2107–2124.

Lipkovich, I., Duan. Y., and Ahmed, S. (2005), "Multiple imputation compared with restricted pseudo-likelihood and generalized estimating equations for analysis of binary repeated measures in clinical studies," *Pharmaceutical Statistics*, 4:267-285

Liu, G.F., and Zhan, X.J. (2011), "Comparisons of methods for analysis of repeated binary responses with missing data," *Journal of Biopharmaceutical Statistics*; 21:371–392.

Molenberghs, G., and Kenward, M.G. (2007), "Missing Data in Clinical Studies," Chichester: John Wiley & Sons.

National Research Council, (2010), "The prevention and Treatment of Missing Data in Clinical Trials," Panel on Handling Missing Data in Clinical Trials. Committee on National Statistics, Division of Behavioral and Social Sciences and Education. Washington, DC: The National Academies Press.

Robins, J. M., Rotznitzky, A., and Zhao, L. P. (1995), "Analysis of semiparametric regression models for repeated outcomes in the presence of missing data," *Journal of the American Statistical Association,* 90:106–121.

Rubin, D. B. (1976), "Inference and Missing Data," *Biometrika*, 63, 581–592.

Rubin, D. B. (1987), "Multiple Imputation for Nonresponse in Surveys," New York: Wiley.

Sacks, D.B., Bruns, D.E., Goldstein, D.E., Maclaren, N.K., McDonald, J.M., and Parrott, M. (2002), "Guidelines and recommendations for laboratory analysis in the diagnosis and management of diabetes mellitus," *Clinical Chemistry*, 48:436–472

SAS Institute Inc. 2008. SAS/STAT® 9.2 User's Guide. Cary, NC: SAS Institute Inc.

Schafer, J. L. (1999), "Multiple imputation: A primer," *Statistical Methods in Medical Research* 8(1):3–15.

Shieh, Y.-Y. (2003), "Imputation Methods on General Linear Mixed Models of Longitudinal Studies," Federal Committee on Statistical Methodology Research Conference, Arlington, VA.

van Buuren, S. (2014), "Fully Conditional Specification," In Molenberghs. G., Fitzmaurice G.M., Kenward, M.G., Tsiatis, A.A., Verbeke, G. (Eds), *Handbook of Missing Data Methodology*, Chapter 14. Chapman & Hall/CRC Press, Boca Raton, FL

Verbeke, G., and Molenberghs, G. (2000), "Linear mixed models for longitudinal data," Springer: New York.

Wolfinger, R., and O'Connell, M. (1993), "Generalized linear mixed models: A pseudo-likelihood approach," *Journal of Statistical Computation and Simulation* 48:233–243.

Zeger, S. L. and Liang, K.-Y. (1986), "Longitudinal data analysis for discrete and continuous outcomes," *Biometrics* 42, 121-130.

Appendix

Examples of SAS code

```
**************************************************************************
Adjusted and Unadjusted GEE models
**************************************************************************;
proc genmod data= simu_data;
class subjid  visit trt ;
model target_a1c=   trt  <baselinehba1c> visit visit*trt <baselinehba1c*visit> /dist=bin ;
repeated subject=subjid/type=un;
lsmeans visit*trt /diff;
run;

**************************************************************************
Adjusted and Unadjusted GLMM models
**************************************************************************;
proc glimmix data= simu_data empirical ;
nloptions maxiter=100 tech=NRRIDG;
class subjid visit trt;
model target_a1c = trt  < baselinehba1c > visit visit*trt <baselinehba1c*visit>/dist=bin ;
random int / subject=subjid;
lsmeans visit*trt/diff;
```



```
run;

*************************************************************************

Adjusted and Unadjusted MI

*************************************************************************;

proc mi data=one seed=1305417 nimpute=30 out=outone;

by trt;

var baselinea1c v1 v2 v3;

monotone reg(v2= baselinea1c v1);

monotone reg(v3= baselinea1c v1 v2 );

run;

ods output diffs=MIdiffs (where= (visit in (3) and _visit in (3))rename=(

stderr=MIse estimate=MIest));

proc genmod data=outone descending;

class subjid   visit trt;by _imputation_;

model target_a1c= <baselinea1c> trt  visit <baselinea1c*visit> visit*trt/dist=bin ;

repeated subject=subjid/type=un;

lsmeans visit*trt/pdiff;

proc mianalyze data=MIdiffs;
```

```
modeleffects Miest;
stderr mise;
run;

**************************************************************************
Adjusted and Unadjusted WGEE models
**************************************************************************;
%MACRO WGEE(
      INPUTDS =,
      VISVAR  =,
      STRVIS  =,
      ENDVIS  =,
      TRTVAR  =,
      SUBJVAR =,
      Y       =,
      CLASVAR =,
      DMODL   =,
      DCOVTYPE=,
      MCOVTYPE=,
      TRIM    =);

PROC SORT DATA=&INPUTDS OUT=ONE1;
```

```
 BY &SUBJVAR &VISVAR;
 WHERE &STRVIS<=&VISVAR<=&ENDVIS;
RUN;

DATA _LASTVIS;
 SET ONE1;
 BY &SUBJVAR &VISVAR;
 IF LAST.&SUBJVAR;
 LASTVIS=&VISVAR;
 KEEP &SUBJVAR LASTVIS;
RUN;

DATA TWO1;
 MERGE ONE1 _LASTVIS;
 BY &SUBJVAR;
 IF &VISVAR<LASTVIS THEN DROP=0;
 IF &VISVAR=LASTVIS AND LASTVIS<&ENDVIS THEN DROP=1;
 IF &VISVAR=LASTVIS AND LASTVIS=&ENDVIS THEN DROP=0;
RUN;

%*************************************************************************;
%*Visitwise logistic regression analysis of dropout to determine weights *;
```



```
%*************************************************************************;
proc sort data=TWO1; by &TRTVAR &SUBJVAR; run;
ODS LISTING CLOSE;
 PROC GENMOD DATA=TWO1 DESCENDING;
   CLASS &CLASVAR;
   MODEL DROP=&DMODL/DIST=BIN PRED TYPE3;

   REPEATED SUBJECT=&SUBJVAR/WITHINSUBJECT=&VISVAR TYPE=&DCOVTYPE
         CORRW ECOVB ECORRB MCOVB MCORRB MODELSE;

   ODS OUTPUT OBSTATS=_PRED GEEEMPPEST=EMPEST GEEMODPEST=MODEST
          GEENCORR=MODCORR GEENCOV=MODCOV GEEWCORR=WORKING
          GEERCORR=EMPCORR GEERCOV=EMPCOV
          TYPE3=TYPE3;
RUN;

ODS LISTING;

%*************************************************;
%* Merge probability of dropout with main data  *;
%*************************************************;
```



```
DATA _PRED2;
  SET _PRED;
  RENAME DROP=PDROP;
RUN;

DATA THREE1;
  MERGE TWO1 _PRED2;
  IF PRED NE . ;
RUN;

%*************************************************************************
***;
%* Accumulate inverse probability weights over visits and output new data set  *;
%*************************************************************************
***;
proc sort data=THREE1; by &SUBJVAR; run;
DATA _WGT (KEEP=&SUBJVAR WI);
  SET THREE1;
  BY &SUBJVAR;
  RETAIN WI;
  IF FIRST.&SUBJVAR THEN WI=1;
  IF NOT LAST.&SUBJVAR THEN WI=WI*(1-PRED);
```



```
 IF LAST.&SUBJVAR THEN DO;
  IF &VISVAR<&ENDVIS THEN WI=WI*PRED;
    ELSE WI=WI*(1-PRED);
    WI=1/WI;
    OUTPUT;
 END;
RUN;


%*************************************************;
%* Add inverse probability weight to data set and*;
%* trim weight to eliminate unstable weights     *;
%*************************************************;
DATA FOUR;
 MERGE THREE1 _WGT;
 WI_TR=MIN(WI, &TRIM);
 BY &SUBJVAR;
RUN;


PROC GENMOD DATA=FOUR DESCENDING;
  SCWGT WI;
  CLASS &CLASVAR;
  MODEL &Y =<baselinea1c> VISIT trt <baselinea1c*visit> VISIT*trt/DIST=bin  TYPE3;
```



```
  REPEATED SUBJECT=&SUBJVAR/ TYPE=&MCOVTYPE CORRW;
  LSMEANS &TRTVAR*&VISVAR/CL DIFF ILINK;

RUN;

%MEND WGEE;
%WGEE(
  INPUTDS   =dataset,
  VISVAR    =VISIT,
  STRVIS    =1,
  ENDVIS    =3,
  TRTVAR    =trt,
  SUBJVAR   =subjid,
  Y         =target_a1c,
  CLASVAR   =subjid VISIT trt,
  DMODL     =baselinea1c a1c trt baselinea1c*trt a1c*trt,
  DCOVTYPE  =ind,
  MCOVTYPE  =ind,
  TRIM      =20);
```

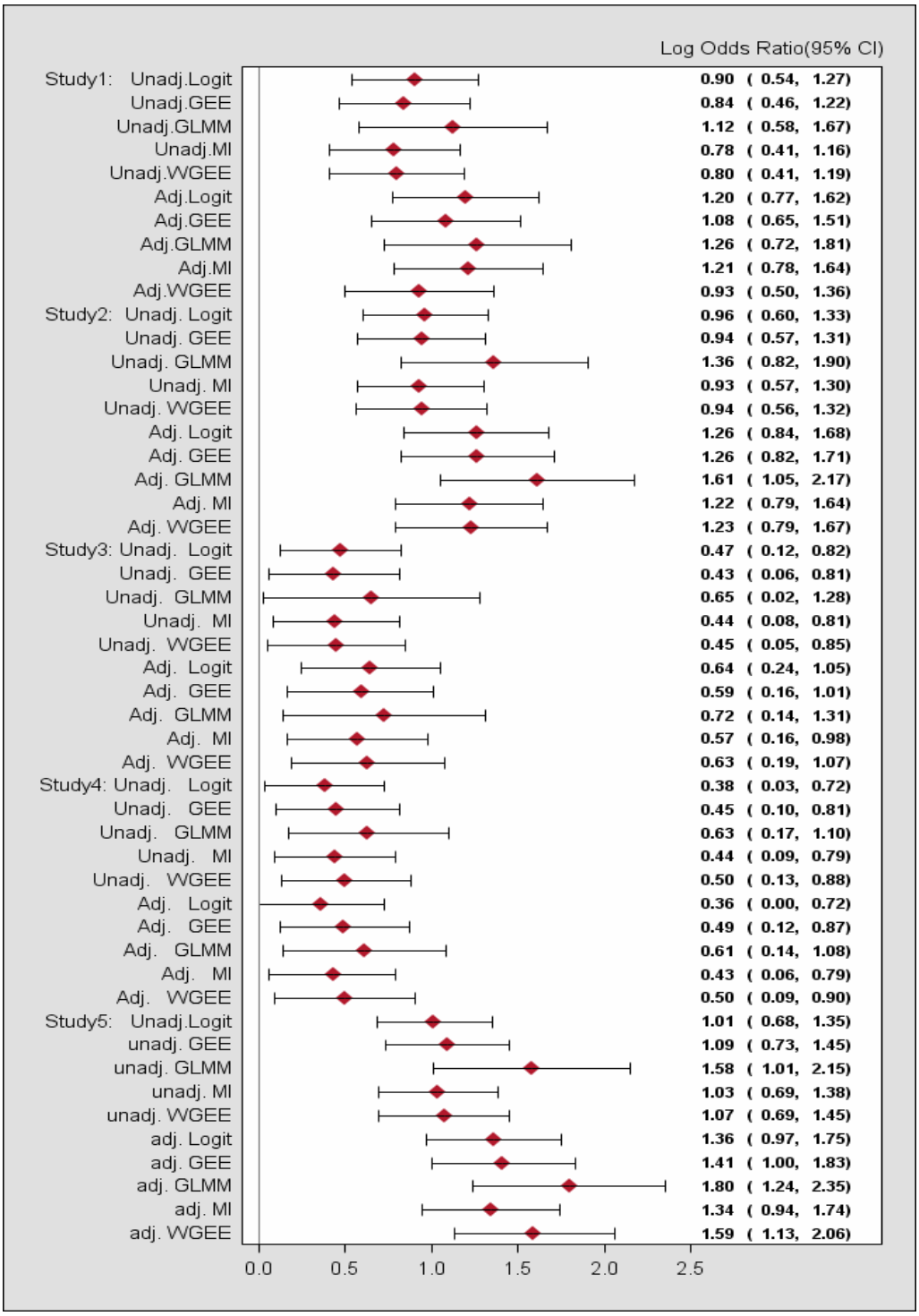



Figure 1. Log odds ratio and 95% CI by study and method

Table 1. Mean treatment profiles in simulation model

Hypothesis		visit1 HbA1c (%) (% achieved target)*	visit2 HbA1c (%)(% achieved target)	visit3 HbA1c (%)(% achieved target)	visit4 HbA1c (%)(% achieved target)
No treatment effect	Treatment	8.5 (0%)	7.6 (27%)	7.3 (39%)	7.0 (50%)
	Comparator	8.5 (0%)	7.6 (27%)	7.3 (39%)	7.0 (50%)
Strong treatment effect	Treatment	8.5 (0%)	7.6 (27%)	7.3 (39%)	7.0 (50%)
	Comparator	8.5 (0%)	7.8 (21%)	7.6 (29%)	7.4 (37%)

* HbA1c (%) is a continuous variable from which a binary response of whether the patient achieved the target HbA1c of 7.0% is derived. We let all baseline (visit 1) HbA1c values be >7.0% because this usually is one of the inclusion criteria in diabetes trials.

Table 2: Summary of analysis results from 2000 simulations with no treatment effect

			(45%, 45%)*				(25%, 45%)*			
	N		Bias	SE	Cov (%)	Type I error rate (%)	Bias	SE	Cov (%)	Type I error rate (%)
MCAR	50	Unadj.logit	0.01	0.41	95	5.1	0.15	0.41	92	**7.9**
		Adj.logit	0.00	0.45	95	4.9	0.18	0.45	93	**7.2**
		Unadj.GEE	0.01	0.53	95	5.2	0.02	0.50	95	5.4
		Adj.GEE	0.01	0.62	95	5.2	0.03	0.57	95	4.8
		Unadj.GLMM	0.01	0.60	95	5.1	0.02	0.56	95	4.9
		Adj.GLMM	0.01	0.65	95	4.6	0.03	0.60	95	4.8
		Unadj.MI	0.00	0.50	98	2.1	0.00	0.48	97	3.2
		Adj.MI	-0.02	0.57	98	1.9	0.01	0.54	97	2.6
		Unadj.WGEE	0.01	0.57	97	3.2	0.05	0.52	96	3.9
		Adj.WGEE	0.00	0.63	96	4.0	0.06	0.58	96	4.1
	200	Unadj.logit	0.01	0.20	95	5.1	0.15	0.20	89	**11.0**
		Adj.logit	0.01	0.22	95	5.1	0.17	0.22	88	**11.9**
		Unadj.GEE	0.01	0.26	96	4.4	0.01	0.25	95	5.1
		Adj.GEE	0.01	0.30	95	4.8	0.01	0.27	94	5.7
		Unadj.GLMM	0.01	0.29	96	4.5	0.01	0.27	95	5.0
		Adj.GLMM	0.01	0.31	95	5.0	0.01	0.29	94	5.7
		Unadj.MI	0.00	0.25	98	2.0	0.02	0.24	97	3.3
		Adj.MI	0.01	0.28	98	2.1	0.02	0.26	97	3.5
		Unadj.WGEE	0.01	0.27	97	2.7	0.04	0.25	96	4.2
		Adj.WGEE	0.01	0.30	97	3.4	0.05	0.28	95	5.2
MAR	50	Unadj.logit	0.01	0.42	95	5.2	0.24	0.41	90	**10.1**
		Adj.logit	0.00	0.46	94	5.5	0.28	0.46	90	**10.5**
		Unadj.GEE	0.02	0.51	94	5.7	-0.04	0.48	94	**6.1**
		Adj.GEE	0.01	0.58	95	5.3	-0.03	0.55	95	5.2
		Unadj.GLMM	0.02	0.58	95	4.9	-0.11	0.55	94	5.6
		Adj.GLMM	0.01	0.62	95	4.4	-0.07	0.58	95	4.7

ACCEPTED MANUSCRIPT

		Unadj.MI	0.00	0.51	97	3.0	-0.01	0.48	97	3.2
		Adj.MI	-0.01	0.58	97	2.6	-0.02	0.55	97	2.5
		Unadj.WGEE	0.03	0.66	94	5.5	0.13	0.59	95	5.5
		Adj.WGEE	0.03	0.67	94	5.6	0.13	0.61	94	**6.3**
	200	Unadj.logit	0.01	0.21	95	5.1	0.23	0.20	79	**20.6**
		Adj.logit	0.01	0.23	94	6.0	0.28	0.23	77	**23.5**
		Unadj.GEE	0.01	0.25	94	6.0	-0.05	0.24	94	**6.1**
		Adj.GEE	0.01	0.28	94	5.7	-0.03	0.27	94	5.9
		Unadj.GLMM	0.01	0.28	94	5.6	-0.11	0.27	93	**7.3**
		Adj.GLMM	0.01	0.29	94	5.7	-0.07	0.28	94	5.9
		Unadj.MI	0.01	0.25	96	3.8	0.02	0.24	96	3.7
		Adj.MI	0.01	0.28	96	3.7	0.02	0.26	97	3.4
		Unadj.WGEE	0.00	0.38	96	3.8	0.19	0.33	94	**6.1**
		Adj.WGEE	0.00	0.35	96	4.4	0.17	0.31	93	**6.9**
MNAR	50	Unadj.logit	0.00	0.41	95	5.1	0.13	0.41	93	**6.5**
		Adj.logit	0.00	0.45	96	4.2	0.16	0.45	94	6.0
		Unadj.GEE	0.02	0.53	95	5.2	-0.10	0.49	94	**6.2**
		Adj.GEE	0.02	0.61	95	5.3	-0.10	0.56	93	**6.5**
		Unadj.GLMM	0.02	0.59	95	4.6	-0.14	0.55	94	**6.2**
		Adj.GLMM	0.02	0.64	95	4.5	-0.12	0.59	93	**6.3**
		Unadj.MI	-0.01	0.50	98	2.4	-0.08	0.47	96	3.4
		Adj.MI	-0.02	0.57	98	2.1	-0.10	0.53	97	3.2
		Unadj.WGEE	0.02	0.57	97	3.3	-0.01	0.52	97	3.3
		Adj.WGEE	0.01	0.63	96	4.4	-0.04	0.57	96	4.3
	200	Unadj.logit	0.00	0.20	94	5.8	0.13	0.20	90	**10.3**
		Adj.logit	0.00	0.22	95	5.2	0.15	0.22	89	**11.3**
		Unadj.GEE	0.01	0.26	94	6.0	-0.10	0.24	92	**7.8**
		Adj.GEE	0.01	0.29	94	5.6	-0.10	0.27	92	**8.4**
		Unadj.GLMM	0.01	0.29	94	5.7	-0.13	0.27	91	**8.6**
		Adj.GLMM	0.01	0.30	95	5.3	-0.11	0.28	92	**8.4**
		Unadj.MI	0.01	0.25	97	2.8	-0.06	0.23	96	4.2

ACCEPTED MANUSCRIPT

		Adj.MI	0.00	0.28	98	2.5	-0.07	0.26	96	3.8
		Unadj.WGEE	0.01	0.28	97	3.5	-0.01	0.25	96	4.0
		Adj.WGEE	0.01	0.30	96	4.2	-0.03	0.27	95	4.7

* Percentage of missing data for comparator and treatment, respectively; Cov: coverage of 95% CI; Est: - log odds ratio;

Boldface font indicates that the type I error rate is beyond 2 SEs of simulations.

Table 3: Summary of analysis results from 2000 simulations with strong treatment effect

			(45%, 45%)*					(25%, 45%)*					(45%, 25%)*				
	N		Est	RelBias (%)	SE	Cov (%)	Power (%)	Est	RelBias (%)	SE	Cov (%)	Power (%)	Est	Rel Bias (%)	SE	Cov (%)	Power (%)
MCAR	50	Unadj.logit	-0.48	-10.3	0.42	95	19.7	-0.36	-31.7	0.42	94	14.8	-0.63	18.5	0.42	95	30.3
		Adj.logit	-0.56	-11.7	0.47	94	22.4	-0.43	-32.8	0.46	92	15.5	-0.75	17.0	0.47	95	35.0
		Unadj.GEE	-0.55	3.6	0.55	95	17.1	-0.55	4.7	0.51	95	19.5	-0.56	6.6	0.51	96	19.4
		Adj.GEE	-0.68	6.0	0.63	95	18.0	-0.68	5.9	0.59	95	21.4	-0.69	7.0	0.59	95	20.3
		Unadj.GLMM	-0.60	13.0	0.61	95	16.1	-0.61	16.0	0.57	95	18.8	-0.61	15.1	0.57	95	17.9
		Adj.GLMM	-0.71	11.2	0.66	95	17.5	-0.71	11.3	0.61	95	20.6	-0.71	11.4	0.62	95	19.2
		Unadj.MI	-0.54	2.7	0.51	98	14.2	-0.55	4.0	0.49	97	17.6	-0.56	5.5	0.49	97	16.0
		Adj.MI	-0.65	2.1	0.59	98	14.1	-0.66	3.3	0.56	98	17.5	-0.67	4.6	0.56	97	16.5
		Unadj.WGEE	-0.53	-0.3	0.58	97	12.5	-0.51	-4.4	0.53	96	14.8	-0.57	7.2	0.54	97	16.4
		Adj.WGEE	-0.66	2.8	0.65	96	14.5	-0.63	-2.2	0.60	96	17.4	-0.69	8.6	0.60	96	19.0
	200	Unadj.logit	-0.47	-11.7	0.21	95	62.3	-0.35	-33.1	0.21	85	39.8	-0.60	12.8	0.21	94	82.1
		Adj.logit	-0.55	-14.5	0.23	94	67.9	-0.42	-35.1	0.23	82	44.5	-0.71	10.7	0.23	95	89.4
		Unadj.GEE	-0.55	2.9	0.27	95	53.5	-0.54	2.8	0.25	94	58.6	-0.53	0.2	0.25	95	55.3
		Adj.GEE	-0.65	1.2	0.30	95	57.5	-0.65	1.2	0.28	94	62.7	-0.64	-0.7	0.28	95	63.0
		Unadj.GLMM	-0.59	11.7	0.30	95	51.5	-0.60	12.8	0.28	93	58.4	-0.58	8.9	0.28	95	53.9
		Adj.GLMM	-0.67	4.9	0.31	94	57.5	-0.67	5.3	0.29	94	62.7	-0.66	3.1	0.29	96	62.2
		Unadj.MI	-0.54	2.3	0.25	97	58.9	-0.53	-0.7	0.24	97	60.8	-0.54	1.6	0.24	97	62.5
		Adj.MI	-0.64	0.4	0.29	97	62.2	-0.62	-2.4	0.27	97	65.2	-0.64	0.5	0.27	97	69.4
		Unadj.WGEE	-0.52	-1.3	0.28	97	47.2	-0.50	-6.1	0.26	96	49.6	-0.54	2.7	0.26	96	55.4
		Adj.WGEE	-0.62	-2.4	0.31	96	53.5	-0.59	-7.1	0.29	95	55.0	-0.65	1.9	0.29	96	64.0
MAR	50	Unadj.logit	-0.38	-28.2	0.44	94	12.4	-0.19	-63.7	0.43	88	7.6	-0.70	31.7	0.43	94	35.5
		Adj.logit	-0.47	-26.5	0.49	94	15.8	-0.24	-62.6	0.48	85	7.6	-0.85	33.4	0.48	93	41.4
		Unadj.GEE	-0.54	2.6	0.54	94	17.0	-0.61	14.3	0.51	94	22.5	-0.49	-8.0	0.50	95	14.9
		Adj.GEE	-0.66	2.8	0.61	95	18.6	-0.72	12.2	0.57	94	23.3	-0.62	-3.9	0.57	95	17.7
		Unadj.GLMM	-0.63	18.2	0.60	95	16.9	-0.76	43.2	0.57	94	25.1	-0.47	-12.2	0.55	95	12.5
		Adj.GLMM	-0.71	10.8	0.64	95	19.1	-0.80	25.4	0.61	95	24.5	-0.61	-4.5	0.59	95	16.9

		Unadj.MI	-0.54	1.5	0.53	97	13.2	-0.58	9.2	0.51	97	17.0	-0.52	-2.0	0.49	97	15.7
		Adj.MI	-0.65	1.0	0.61	97	13.4	-0.69	8.5	0.58	97	19.0	-0.63	-1.6	0.56	97	16.9
		Unadj.WGEE	-0.51	-3.0	0.69	94	11.9	-0.41	-23.3	0.62	94	10.9	-0.70	32.3	0.59	95	20.1
		Adj.WGEE	-0.63	-0.8	0.70	94	15.4	-0.54	-16.1	0.64	94	13.6	-0.81	27.0	0.62	94	25.3
	200	Unadj.logit	-0.37	-29.3	0.22	89	41.9	-0.18	-66.1	0.21	61	13.6	-0.68	27.8	0.21	89	88.9
		Adj.logit	-0.45	-29.5	0.24	87	48.4	-0.22	-65.9	0.23	55	15.4	-0.82	27.4	0.24	88	93.7
		Unadj.GEE	-0.53	0.3	0.26	96	52.7	-0.59	11.8	0.25	94	66.4	-0.46	-12.4	0.25	93	46.0
		Adj.GEE	-0.63	-1.1	0.29	95	57.5	-0.68	6.3	0.28	94	69.4	-0.58	-9.7	0.27	94	54.2
		Unadj.GLMM	-0.60	13.5	0.29	95	53.6	-0.72	36.6	0.28	90	74.5	-0.45	-15.9	0.27	93	37.1
		Adj.GLMM	-0.67	4.6	0.31	95	58.7	-0.75	16.8	0.29	93	73.6	-0.57	-11.3	0.28	94	50.0
		Unadj.MI	-0.53	0.4	0.26	97	53.5	-0.52	-2.3	0.25	97	55.8	-0.53	0.5	0.24	96	61.1
		Adj.MI	-0.64	-0.5	0.29	97	58.3	-0.62	-3.2	0.28	96	61.5	-0.64	-0.4	0.27	96	66.2
		Unadj.WGEE	-0.46	-13.7	0.40	96	24.2	-0.27	-48.5	0.35	91	14.9	-0.73	37.1	0.32	93	64.3
		Adj.WGEE	-0.57	-10.7	0.37	96	35.4	-0.41	-35.3	0.33	90	24.7	-0.81	25.9	0.31	92	74.7
MNAR	50	Unadj.logit	-0.46	-13.7	0.42	94	19.1	-0.38	-29.2	0.42	93	15.3	-0.60	12.3	0.42	96	30.5
		Adj.logit	-0.55	-13.5	0.46	94	21.7	-0.46	-28.3	0.46	92	17.1	-0.72	12.9	0.47	95	32.9
		Unadj.GEE	-0.55	4.2	0.54	94	17.8	-0.65	21.8	0.50	94	25.3	-0.46	-13.4	0.50	95	14.1
		Adj.GEE	-0.68	6.7	0.62	95	19.1	-0.77	20.3	0.57	95	26.2	-0.59	-8.2	0.57	95	16.9
		Unadj.GLMM	-0.61	14.2	0.60	94	16.4	-0.73	37.9	0.56	94	24.9	-0.48	-9.6	0.56	95	12.4
		Adj.GLMM	-0.71	11.7	0.65	95	18.7	-0.81	27.3	0.60	94	26.2	-0.61	-5.2	0.60	95	16.1
		Unadj.MI	-0.54	2.4	0.50	97	14.8	-0.62	16.3	0.48	97	22.1	-0.48	-9.4	0.48	97	13.9
		Adj.MI	-0.65	2.2	0.58	98	15.1	-0.75	16.5	0.55	97	23.8	-0.58	-9.1	0.55	97	13.3
		Unadj.WGEE	-0.53	-0.1	0.57	96	12.0	-0.55	2.8	0.53	96	16.0	-0.53	-0.8	0.53	96	14.9
		Adj.WGEE	-0.66	3.2	0.64	95	14.9	-0.70	9.0	0.59	96	21.4	-0.64	-0.7	0.58	95	17.6
	200	Unadj.logit	-0.47	-11.9	0.21	94	60.5	-0.37	-29.7	0.21	87	44.2	-0.58	9.5	0.21	95	79.1
		Adj.logit	-0.56	-13.1	0.23	92	67.9	-0.45	-30.4	0.23	86	50.4	-0.69	8.1	0.23	95	86.0
		Unadj.GEE	-0.55	2.9	0.26	95	54.8	-0.64	19.8	0.25	93	73.1	-0.43	-18.1	0.25	93	41.6
		Adj.GEE	-0.65	2.1	0.30	95	59.9	-0.73	14.7	0.28	94	76.0	-0.54	-15.3	0.27	93	50.1
		Unadj.GLMM	-0.59	11.6	0.29	95	53.5	-0.71	34.2	0.27	90	74.0	-0.46	-14.1	0.27	95	38.3
		Adj.GLMM	-0.68	5.7	0.31	95	58.9	-0.77	20.0	0.29	93	75.7	-0.56	-13.0	0.29	94	48.8
		Unadj.MI	-0.54	2.0	0.25	97	59.1	-0.59	10.8	0.24	96	70.8	-0.47	-11.8	0.24	97	50.8
		Adj.MI	-0.64	0.7	0.28	97	64.4	-0.70	9.4	0.27	97	76.8	-0.56	-12.9	0.27	96	56.8

ACCEPTED MANUSCRIPT

		Unadj.WGEE	-0.53	-0.2	0.28	97	48.8	-0.53	0.8	0.26	97	56.2	-0.51	-3.5	0.25	97	53.2
		Adj.WGEE	-0.64	-0.7	0.30	96	55.4	-0.66	2.9	0.28	96	65.7	-0.60	-6.7	0.28	96	58.0

* Percentage of missing data for comparator and treatment, respectively; Cov: coverage of 95% CI; Rel Bias: relative Bias;

Est: - log odds ratio; Boldface font indicates that the type I error rate is beyond 2 SEs of simulations.

Table 4. Analysis results from five diabetes clinical studies

		Log odds ratio	SE	LL	UL	p-value
Study 1	Unadj.logit	0.90	0.18	0.54	1.27	<0.001
	Adj.logit	1.20	0.21	0.77	1.62	<0.001
	Unadj.GEE	0.84	0.19	0.46	1.22	<0.001
	Adj.GEE	1.08	0.22	0.65	1.51	<0.001
	Unadj.GLMM	1.12	0.27	0.58	1.67	<0.001
	Adj.GLMM	1.26	0.27	0.72	1.81	<0.001
	Unadj.MI	0.78	0.19	0.41	1.16	<0.001
	Adj.MI	1.21	0.22	0.78	1.64	<0.001
	Unadj.WGEE	0.80	0.20	0.41	1.19	<0.001
	Adj.WGEE	0.93	0.21	0.50	1.36	<0.001
Study 2	Unadj.logit	0.96	0.18	0.60	1.33	<0.001
	Adj.logit	1.26	0.21	0.84	1.68	<0.001
	Unadj.GEE	0.94	0.18	0.57	1.31	<0.001
	Adj.GEE	1.26	0.22	0.82	1.71	<0.001
	Unadj.GLMM	1.36	0.27	0.82	1.90	<0.001
	Adj.GLMM	1.61	0.28	1.05	2.17	<0.001
	Unadj.MI	0.93	0.18	0.57	1.30	<0.001
	Adj.MI	1.22	0.21	0.79	1.64	<0.001
	Unadj.WGEE	0.94	0.19	0.56	1.32	<0.001
	Adj.WGEE	1.23	0.22	0.79	1.67	<0.001
Study 3	Unadj.logit	0.47	0.18	0.12	0.82	0.007
	Adj.logit	0.64	0.20	0.24	1.05	0.002
	Unadj.GEE	0.43	0.19	0.06	0.81	0.021
	Adj.GEE	0.59	0.21	0.16	1.01	0.006
	Unadj.GLMM	0.65	0.31	0.02	1.28	0.038
	Adj.GLMM	0.72	0.29	0.14	1.31	0.014
	Unadj.MI	0.44	0.18	0.08	0.81	0.016
	Adj.MI	0.57	0.21	0.16	0.98	0.005

	Unadj.WGEE	0.45	0.20	0.05	0.85	0.026
	Adj.WGEE	0.63	0.22	0.19	1.07	0.004
Study 4	Unadj.logit	0.38	0.17	0.03	0.72	0.028
	Adj.logit	0.36	0.18	0.00	0.72	0.047
	Unadj.GEE	0.45	0.18	0.10	0.81	0.011
	Adj.GEE	0.49	0.19	0.12	0.87	0.009
	Unadj.GLMM	0.63	0.23	0.17	1.10	0.007
	Adj.GLMM	0.61	0.23	0.14	1.08	0.009
	Unadj.MI	0.44	0.17	0.09	0.79	0.012
	Adj.MI	0.43	0.18	0.06	0.79	0.019
	Unadj.WGEE	0.50	0.19	0.13	0.88	0.008
	Adj.WGEE	0.50	0.20	0.09	0.90	0.014
Study 5	Unadj.logit	1.01	0.17	0.68	1.35	<0.001
	Adj.logit	1.36	0.20	0.97	1.75	<0.001
	Unadj.GEE	1.09	0.18	0.73	1.45	<0.001
	Adj.GEE	1.41	0.21	1.00	1.83	<0.001
	Unadj.GLMM	1.58	0.28	1.01	2.15	<0.001
	Adj.GLMM	1.80	0.28	1.24	2.35	<0.001
	Unadj.MI	1.03	0.17	0.69	1.38	<0.001
	Adj.MI	1.34	0.20	0.94	1.74	<0.001
	Unadj.WGEE	1.07	0.19	0.69	1.45	<0.001
	Adj.WGEE	1.59	0.23	1.13	2.06	<0.001

SE: standard error; LL: lower limit of 95% confidence interval; UL: upper limit of 95% confidence interval