Modeling and nonparametric methodology for count data in drug studies

Tuan H. Nguyen, Rutgers University

Javier Cabrera, Rutgers University

José Pinheiro Novartis Pharmaceuticals

Abstract

Objectives: To evaluate and compare the statistical operational characteristics of alternative methods of analyzing longitudinal count data in the context of clinical trial data. Typical statistical models for these data are generalized linear mixed model (GLMM), generalized estimating equations (GEE), and nonparametric methodologies (Wilcoxon, Van Elteren tests, and Lehmann-Hodges estimators).

Study Design and Settings: We compare those approaches via a simulation study in term of power, bias, rooted mean squared error, and coverage probabilities. The simulated data sets try to resemble data from a typical trial where observations are only collected during a subset of weeks during a trial. The mean structure consists of patient's characteristics (age, gender), drug-response model, and random effects (week within patient, patient within center, and center). The non-informative missing data mechanism (MCAR, and exponential dropout) is also implemented. We vary the treatment effects and the rate of dropout.

Results: In term of testing hypothesis, GLMMs provide the most statistical powerful tests among all models. In term of estimation, GLMMs and GEEs perform similarly in term of bias and accuracy, while nonparametric method (Lehmann-Hodges estimators) does worse because of missing data due to dropout.

Conclusions: The GLMM (with sandwich estimator for covariance matrix) is the model of preference in our study.

1 Introduction

In this paper, we consider analysis of count longitudinal data, particularly from a drug development program to treat urinary incontinence. The clinical endpoint for this program is a count variable, the number of incontinence episodes experienced by a patient in the trial. A typical trial for this type of indication has a duration of twelve weeks, with dairy counts of incontinences being extracted from diary data. The daily counts are generally only collected during a subset of the weeks: baseline, week 2, week 6, and week 12 (end of the study). The "traditional" statistical analysis for this type of data is based on weekly counts, with the primary endpoint being the change from baseline at the end of the trial. Last observation carried forward (LOCF) is used for patients who drop out before the study ends. The most commonly used approach for this analysis is nonparametric methodology (Wilcoxon, Van Elteren, or rank ANOVA, etc)[5]. The reason for this approach is that the treatment effect, measured by change at the end of trial from baseline, typically does not fit into the gaussian distribution framework. Furthermore, there is a belief that baseline measurements have already summarized all patient's characteristics. Despite the strengths of these methods, they also have some important disadvantages (e.g., how to handle missing data, etc.), which leads to the question of whether other alternative approaches are more appropriate to analysis this type of indication.

In recent years, several approaches have been proposed to model this type of indication. Most of them can be classified into two classes, the "subject-specific" and the "population-averaged" approaches. Generalized linear mixed models (GLMMs) [7][2] have been used to model the correlation among observations via random effects, and estimate the subject-specific effects (e.g. treatment effect, etc.). Meanwhile, the generalized estimating equations (GEEs) method (Liang and Zeger[6]) is usually used to model correlation due to repeated measurement within each subject and provide population-averaged effects. However, the challenges arise when we face the combination of repeated measurements and random effects, for example, in above clinical trial situation. Typically, for gaussian distribution framework, this combination can be solved due to special properties of gaussian distribution, as discussed in Pinheiro and Bates [8]. But, for non-Gaussian distribution framework, one might end up with either generalized estimating equations or generalized

linear mixed models, depending on the choice made.

In this study, we evaluate and compare the statistical characteristics of some alternative analysis methods for longitudinal count data in this context of clinical trial data primarily via simulation. The standard analysis using nonparametric methodology (Wilcoxon, Van Elteren tests, and Lehmann-Hodges estimator) is included as the benchmark of comparison. Both generalized estimating equations and generalized linear mixed models are used as alternative approaches to the standard method. The evaluation and comparison mainly focus on their ability of detecting the treatment effect (hypothesis testing) and their performance of estimating the treatment effect (point estimation). Section 2 provides a detailed simulation design of this study, while section 3 gives a brief description of methods and models used in this simulation study. Finally, section 4 and 5 give and discuss the results of this study.

2 Design of Simulation study

In this study, we consider 45 scenarios of trials where the relative maximal reduction rate of treatment patients is varied between 0% and 50%, by steps of 5%, and the dropout rate is changed from 0% to 40%, by steps of 10%. Our simulation design mimics a typical clinical trial of a drug development program. There are 30 centers (hospitals), each of which has 5 treatment patients and 5 placebo patients. For each patient, the daily incontinence counts are simulated from Poisson distribution with the rate changing over time (the parametric form of this rate will be described below), then combined into weekly means (number of incontinence episodes per day) by taking average of all counts within week. Patients' gender, age, center, time points and treatment indicators are treated as exploratory variables. Gender is generated from Bernoulli distribution with 50% chance of being male. Age is the integer part of a number, generated from normal distribution with mean of 45 and standard deviation of 10. Center ID is labeled from 1 to 30, patient ID was label from 1 to 3000 (each distinct patient corresponds to a distinct ID). Time point was 0, 2, 6, or 12 for baseline, week 2, week 6, and week 12 respectively. Treatment indicator is either 0 or 1, for placebo and treatment group, respectively.

The parametric model for Poisson rate parameter has two parts: linear part modeling patient's characteristics and nonlinear part modeling time-response model for the drug. Linear part includes fixed effects (age, gender), and random effects (center (σ^{center}), patient($\sigma^{patient}$), and week(σ^{week})). All three random effects (week within patient, patient within center) are generated from normal distribution with mean of 0, and standard deviation of 0.3, 0.15, and 0.075 for center, patient, and week effect respectively. Nonlinear part is the Emax model (three parameters – baseline (e_0), maximum reduction (e_{max}), and time to achieve half of maximum reduction (e_{50})), representing the effect of treatment over time (denoted t, in days),

$$E_{max}(t) = e_0 + \frac{e_{max} * t}{e_{50} + t}$$

For both placebo and treatment patient, baseline parameter, e_0, is chosen to be zero. Parameter e_{50} is set at 10.5 for placebo patient and at 35 for treatment patient. Parameter e_{max} is calculated so that at the end of trial (84^{th} day), the absolute reduction rate for placebo patient is 1.4%, i.e., the ratio of mean at the end of trial over that at baseline is 98.6%. For treatment patient, this parameter is adjusted to fit the designed scenarios (reduction rates are varied between 5% to 40% relative to the placebo reduction rate). Overall, the Poisson rate parameter μ_{ijt}, for patient with ID j, in the i^{th} center, at t^{th} day of the trial is set at

$$\log(\mu_{ijt}) = 0.025 * A_j + \sigma_i^{center} + \sigma_j^{patient} + \sigma_{jt}^{week} - E_{max}(t),$$

where A_j is age of patient with ID j, σ_i^{center} is random effect for the i^{th} center, $\sigma_j^{patient}$ is random effect for patient j, and σ_{jt}^{week} is random effect for week corresponding to t^{th} day of patient j.

After generating all daily incontinence counts, we create some missing data by two models: missing completely at random (MCAR) and exponential dropout model. During the simulated trial, each daily observation has 3% chance of being missing (for each observation, a Bernoulli random variable is generated to determine whether it will be missing or not). The dropout model is carried out by first generating a number X from exponential distribution with parameter λ, then erasing all observation whose time point (in day) bigger than X. The parameter λ is set such that probability of dropout, i.e., Prob(X<84), is equal to the desired dropout rates (varying between 0% to 40%, by steps of 10%).

The output data sets from simulation consist of following variables: response (weekly means obtained by taking average of daily counts), age, gender, patient ID, center ID, week (time indicator), and treatment (indicator whether patient received treatment or placebo). For each of those scenarios, 1000 simulated data sets are generated.

3 Analysis Methods

Nonparametric methods

Nonparametric methods (Wilcoxon, Van Elteren test, and Lehman-Hodges estimation) are applied to change from baseline at the end of trial (week 12) on the log scale. To deal with missing data, the Last Observation Carried Forward (LOCF) is employed for all of these methods, i.e, the latest non-missing weekly means is used in place of missing observations. The following is a brief description of those methods. For full details, one can refer back to Lehmann[5].

The Wilcoxon rank sum test is performed on two samples of size 150, corresponding to treatment and placebo group. The Wilcoxon statistic, with continuity correction, is approximated by a normal distribution. The null hypothesis is rejected when p-value is less than .05. This method is implemented by calling PROC NPAR1WAY in SAS[9].

The Van Elteren test, a stratified version of Wilcoxon test, uses baseline severity (average number of continence episodes per day) to stratify population into 3 groups: mild (average less than 2), moderate (average between 2 and 4), and severe (average greater than 4). These thresholds are chosen, without any scientific knowledge, for simulation purpose only, because they approximately divide the whole population into 3 equal strata. The Van Elteren statistic is the weighted sum of Wilcoxon statistics, obtained from each of strata, where all weights are disproportional to the stratum's sample size. It can be approximated by normal distribution. However, as shown by Koch et al.[4], the van Elteren test is a member of a general family of Mantel-Haenszel mean score tests; we implement this test using SAS procedure PROC FREQ [9] for Cochran-Mantel-Haenszel Statistic with modified Ridit scores.

Hodges-Lehmann (HL) estimator for location shift (denoted Δ_{HL}) is a nonparametric method to estimate the treatment effect of the drug. Technically, it is a median of all difference of change on log scale of any arbitrary pair of treatment and placebo patient. Since the estimated change, the result of this estimator, is calculated on the log scale, the reduction rate of treatment could be estimated (interpreted) as $1 - \exp(\Delta_{HL})$. Furthermore, the asymptotic standard errors can be estimated and used as a useful tool for comparison. This method is implemented by procedure PROC NPAR1WAY [9] in SAS.

Generalized estimating equations

GEEs (Generalized estimating equations) for the Poisson family (log link) are conducted on the response (weekly means) with age, gender, treatment indicator (as a factor), and the interaction term between treatment and week (time point, as a factor) as covariates. The working correlation structure is set to be exchangeable within each patient, i.e., correlations of observations between any two weeks are equal. No imputation is implemented for fitting GEE models; the parameters of correlation structure are estimated from non-missing data only.

From result of fitting GEE model on data, the estimated coefficient of treatment indicator, denoted β, could be interpreted as the population-wise effect of treatment on log scale, i.e., an population average of difference between change from baseline at the end of trial among treatment and placebo patients, regardless all random effects. We also interpret and derive an estimate of reduction rate by following

$$\text{reduction rate} = 1 - \exp(-\beta).$$

Two estimates of standard error are obtained and considered, the model-based estimate (GEE equivalent of the inverse of the Fisher information matrix), and the empirical or data-based estimate (an sandwich estimate of the variance). The model-based estimate has been showed to be consistent when the model and working correlation structure are correctly specified, while empirical estimate are also consistent even if the correlation structure is misspecified. The scale parameter is estimated by Pearson score (Pearson chi-square statistic), then used to adjust the estimated covariance matrix. We also employ the Wald-type test for coefficient of treatment indicator, and derive p-value for each simulation. The GEE with model-based estimated covariance matrix is re-

ferred as model-based GEE, while the empirical estimated covariance matrix GEE model is called empirical GEE. This method is carried out by using procedure PROC GENMOD [9] in SAS.

Generalized linear mixed models

GLMMs (Generalized linear mixed models) for the Poisson family (log link) are conducted on the response (weekly means) with age, gender, treatment indicator (as a factor), and the interaction term between treatment and week (time point, as a factor) as covariate. A multilevel model is used with center ID (as Subject) and patient ID within center ID as random effects. A multiplicative over-dispersion parameter is included in this model to account for non-estimable weekly effect. Same as GEEs, no imputation is applied here when we fit the GLMMs.

Our main focus on result of fitting GLMMs on data is estimate of treatment indicator, its standard error and its p-values in nonzero hypothesis testing. As in GEE estimation, the reduction rate of drug could be estimated (or interpreted) using the following formula

$$\text{reduction rate} = 1 - \exp(-\beta),$$

where β denotes the coefficient of treatment indicator. The standard errors are estimated, one from a model-based GLMMs, and the other from empirical model (empirical covariance "sandwich" estimator). In hypothesis testing of nonzero coefficient, a Wald-type test is derived from estimated coefficient and standard error. This method is implemented by procedure PROC GLIMMIX [9] in SAS.

4 Results of Simulation

Evaluation Criteria

When describing the results from various methods, we focus on the four following quantities:

1. *Statistical power.* The empirical power function is defined as proportion of times when we reject the null hypothesis, i.e., p-value less than .05. The empirical type I error is the value of power function when there is no treatment effect.

2. *Bias.* The bias is the difference of an estimate from the true value of parameters; however, raw bias values are typically hard to evaluate. Therefore, we adopt the notion of standardized bias, discussed in Burton et al [1], as proportion of bias in term of uncertainty in the parameter estimate.

$$\text{standardized bias} = \frac{\text{estimated} - \text{true value}}{\text{SD of estimated}}$$

As a rule of thumb, any standardized bias exceeding 50 % is considered troublesome.

3. *Root mean squared error.* Mean square error (MSE) provides information on the accuracy of an estimate, as it is a sum of variance and square of bias. We adopt root mean squared error, which is the transformed MSE back onto the same scale of the parameter.

4. *Coverage.* The coverage of confidence interval is the proportion of times that the true parameter value lies in the obtained confidence interval. The nominal coverage rate used in this study is 95%. As discussed in Burton et al [1], over-coverage suggests that the results are too conservative which leads to the loss of statistical power. While, under-coverage is usually unacceptable because it is over-confident in the estimation and leads to higher than expected type I error. The average length of the 95% confidence interval is also considered in this study as an evaluation tool.

Results

The comparison of hypothesis testing performance among different methods is summarized by figure 1, consisting of 5 graphs of power curves in 5 different dropout rate settings (dropout rate varies from 0% to 40% by steps of 10%). The nonparametric tests perform better than tests provided by GEE method but worse than that of GLMMs for reduction rates less than 30%. Van Elteren tests perform slightly better than Wilcoxon tests in some scenarios (usually in the mid-range of relative reduction rate 5% – 15%). For tests provided by GEEs, the empirical and model-based version have very similar power curve, but they have the least power compared to other methods. Their empirical type I errors are considerably below the nominal significant level of .05. It is noticed that tests in GEE models gain comparable power with respective to those of nonparametric methods in cases of large dropout rates and high reduction rates. Meanwhile, the GLMMs provide the most

statistical powerful tests in all our cases, and there is no significant difference between empirical and model-based GLMMs.

For comparison of point estimation performance among all methods, see figure 2 for standardized bias and figure 3 for root mean squared error (RMSE). GLMMs and GEE models worked well in estimating the reduction rates, and are not affected by changing dropout rates. There are no significant difference between GLMMs or GEE model estimation in terms of standardized bias. The accuracy (RMSE) of GEEs and GLMMs estimates are also similar, but GLMMs estimates have better accuracy than that of GEEs for dropout rate less than 30%. While nonparametric method (Hodges-Lehmann estimator) only performs well when dropout rate was low, best in term of standardized bias when there is no dropout. As showed in figure 2, HL estimator has very large bias compared to other methods when there is a significant amount of dropout in data.

The estimated standard errors for all methods are also summarized in figure 4. GLMMs do well in estimating the "true" standard errors; estimated standard errors from empirical GLMMs spread more than that of model-based GLMMs. Furthermore, the empirical standard error, the standard deviation of all estimated, is closed to mean of the estimated standard errors, which indicates the unbiasedness of these estimation. HL estimator does relatively good, as showed in figure 4; empirical standard error is closed to mean of estimated standard errors. Meanwhile, standard errors, estimated in GEE models, consistently over-estimate, compared to the empirical standard error.

Comparison of interval estimation among all methods is summarized in table 1. Hodges-Lehmann estimator has the worse coverage probability when missing data due to dropout becomes significant. However, it gives a great estimate (i.e. short CI length with coverage probability coincides nominal probability) when there is no dropout or no treatment effect. The coverage of estimate in GEE models, both empirical and model-based version, are consistently conservative for all cases, and very similar to each other. The coverage in model-based GLMMs are the most conservative with coverage probability of one in almost all cases, but their CI lengths are very large, almost double length of empirical GLMMs' estimate. The empirical GLMMs are relatively permissive compared to other methods, but they have the shortest CI length.

5 Discussion and Conclusion

Overall, the GEE method in this study fails to be a good alternative approach due to the fact that there is correlation among patient with center, while GEE model assumes that all subjects are independent. Also, there is no significant improvement between model-based and empirical GEE models indicates that empirical estimator of variance, while it is robust against misspecification of covariance structure, is not robust against the correlation among subjects. It is also noted that the only reason for bad behavior in hypothesis testing due to the overestimate of standard errors. For GEE method to be applicable in this situation, a more sophisticated mean structure is needed to overcome the over-estimate of standard errors. For example, the categorical variable center should be included to compensate for the center random effect. This issue will be a topic for our future study.

Nonparametric tests perform quite well compared to GEE method, but not as good as that in GLMMs. The stratified Wilcoxon performs slightly better than the Wilcoxon, which indicates that the baseline has somewhat effect on the power of the test. Meanwhile, the nonparametric estimator performs worse, and is effected strongly by missing data due to dropout. It suggests that LOCF is not an appropriate imputation when dropout is MCAR type, and more sophisticated techniques are required to handle missing in this type of data.

In this simulation study, GLMMs appear to be the best model. Tests in both model-based and empirical GLMMs have similar statistical power, but model-based GLMMs have smaller variance of estimated standard error than that of empirical GLMMs. This phenomenon has been studied rigorously under simple linear regression case in Kauermann and Carroll (2001) [3]. In that paper, they show that the larger variance of estimated standard errors is the price for the robustness of empirical GLMMs; it leads to under-coverage of empirical GLMMs. This issue has been confirmed in our study. However, with 95% nominal probability, the coverage probabilities are all above 90%, which seems to be acceptable. Therefore, we suggest to use empirical GLMMs for analysis of count longitudinal data in context of clinical trial for drug development program.

References

[1] Andrea Burton, Douglas G. Altman, Patrick Royston, and Roger L. Holder. The design of simulation studies in medical statistics. *Statistics in Medicine*, 25:4279–4292, 2006.

[2] Jiming Jiang. *Linear and generalized linear mixed models and their applications*. Springer, New York, 2007.

[3] Goran Kauermann and Raymond J. Carroll. A note on the efficiency of sandwich covariance matrix estimation. *Journal of the American Statistical Association*, 96:1387–1396, 2001.

[4] Gary G. Koch, Ingrid A. Amara, Gordon W. Davis, and Dennis B. Gillings. A review of some statistical methods for covariance analysis of categorical data. *Biometrics*, 38:563–595, 1982.

[5] Erich L. Lehmann. *Nonparametrics: Statistical Methods based on Ranks*. Holden-Day, San Francisco, 1975.

[6] Kung-Yee Liang and Scott L. Zeger. Longitudinal data analysis using generalized linear model. *Biometrika*, 73:13–22, 1986.

[7] Charles E. McCulloch, Shayle R. Searle, and John M. Neuhaus. *Generalized, linear, and mixed models*. Wiley, New York, 2008.

[8] José C. Pinheiro and Douglas M. Bates. *Mixed-effects models in S and S-PLUS*. Springer, New York, 2000.

[9] SAS. *SAS/STAT 9.2 Users Guide*. SAS Institute Inc., Cary, NC., 2008.

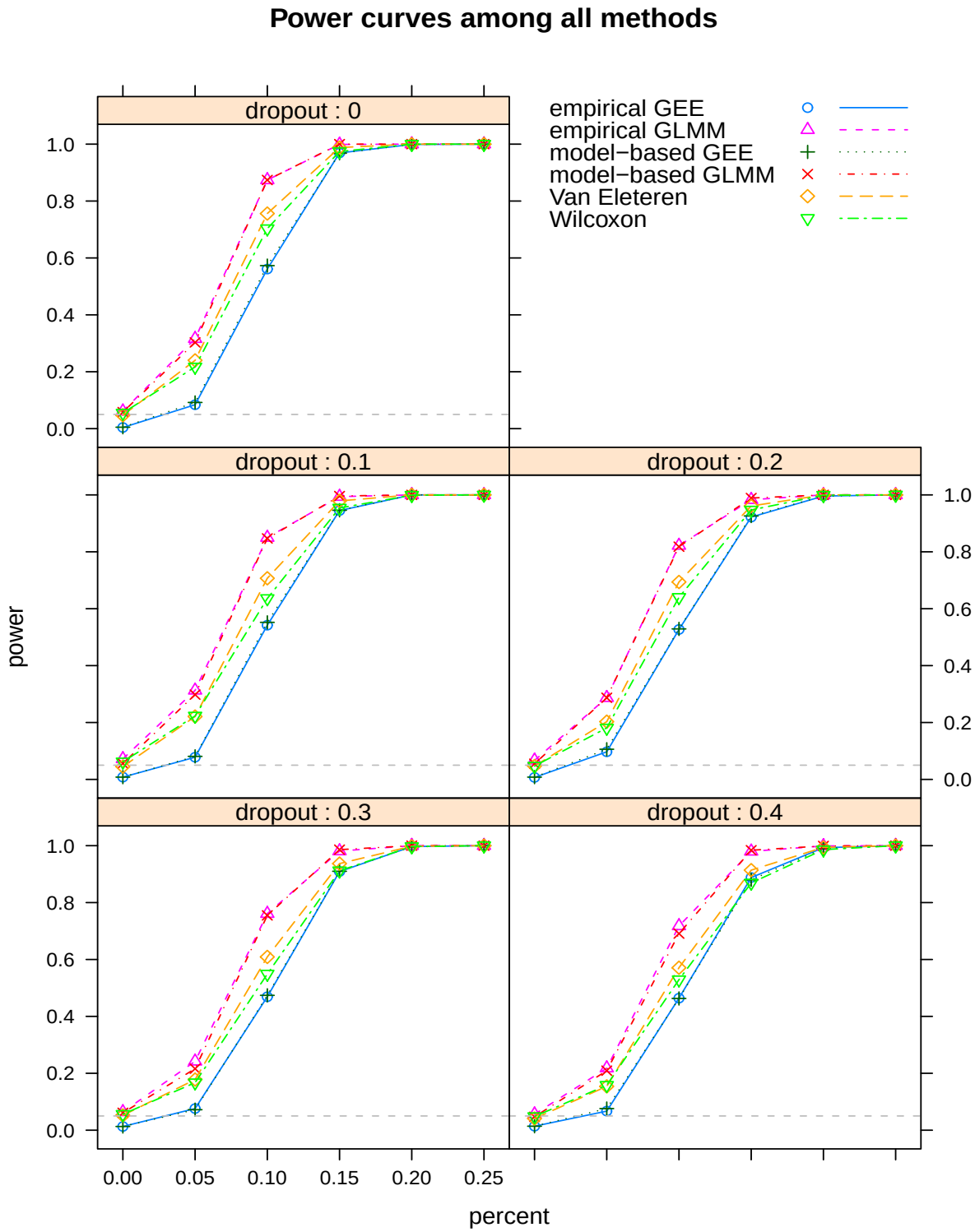

Figure 1: Power curves among different methods for different dropout rates. GLMMs work the best; while GEE models have the least power, due to some violation of their assumptions. The nonparametric testing work quite well among all methods.

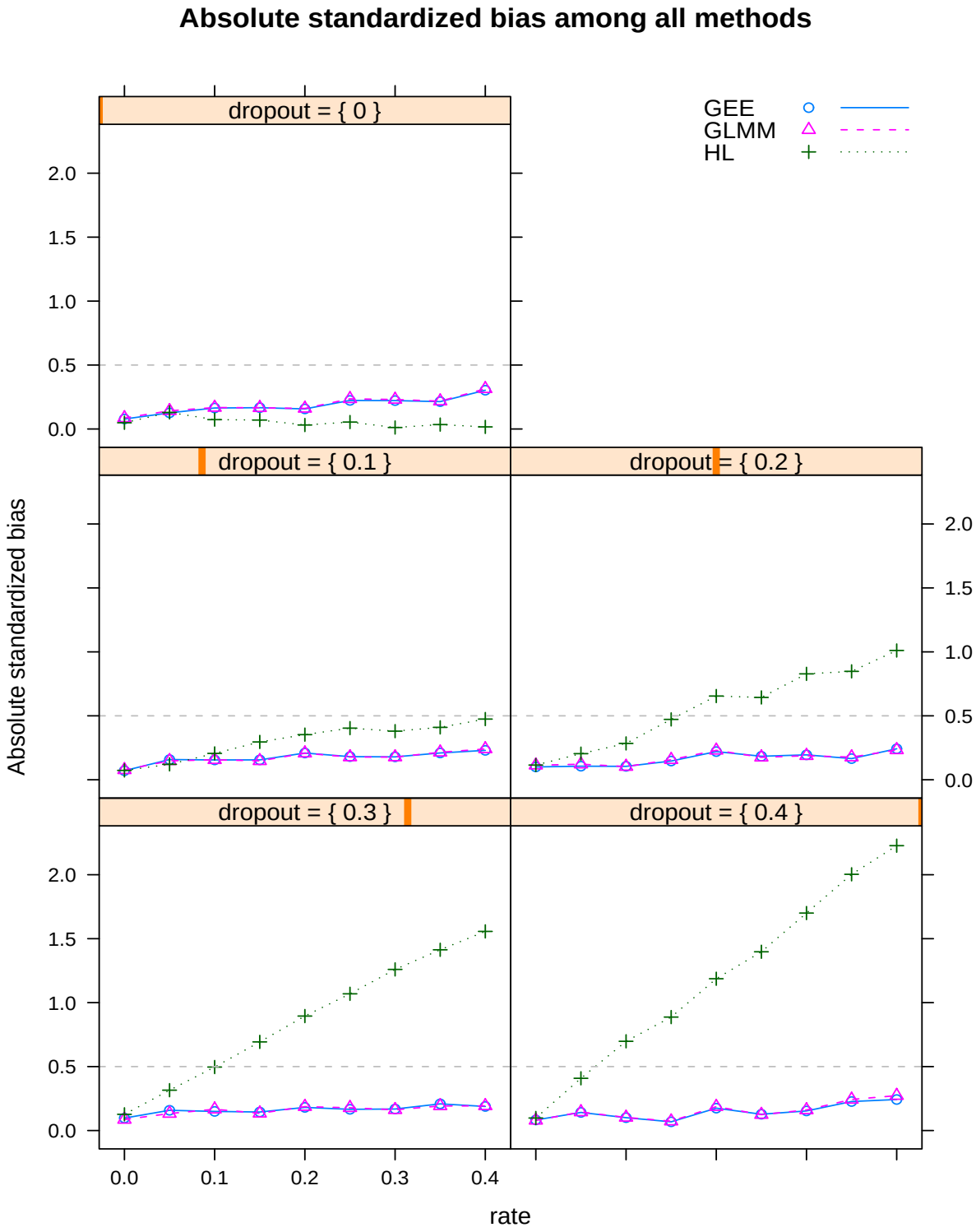

Figure 2: Standardized bias among different methods for different dropout rates. GLMMs and GEE models have similar bias. The Lehmann-Hodges estimator works worse when dropout rate increases.

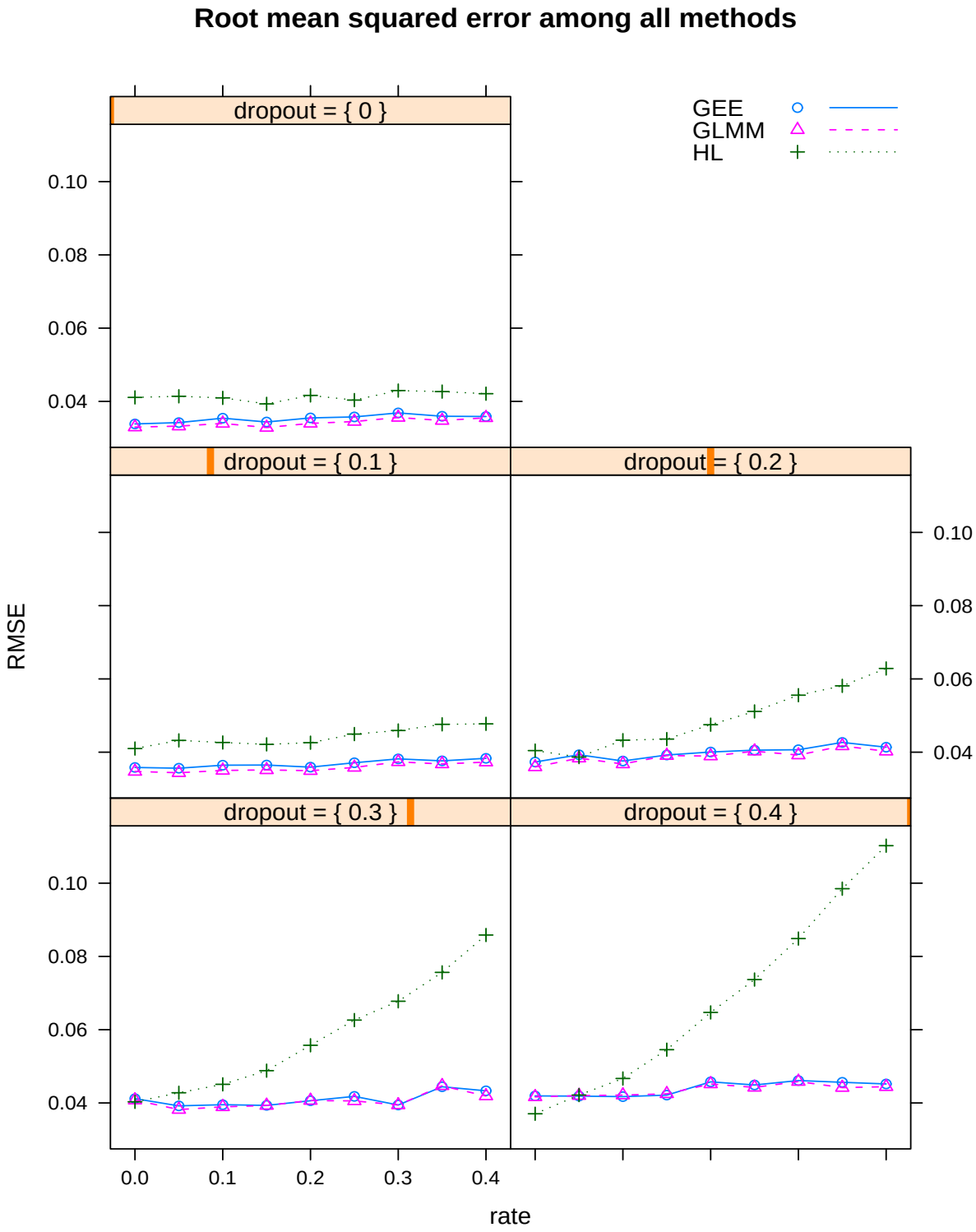

Figure 3: Root mean squared error among different methods for different dropout rates. GLMMs and GEE models have similar root mean squared error, but GLMMs have little better RMSE. The Lehmann-Hodges estimator works worse when dropout rate increases.

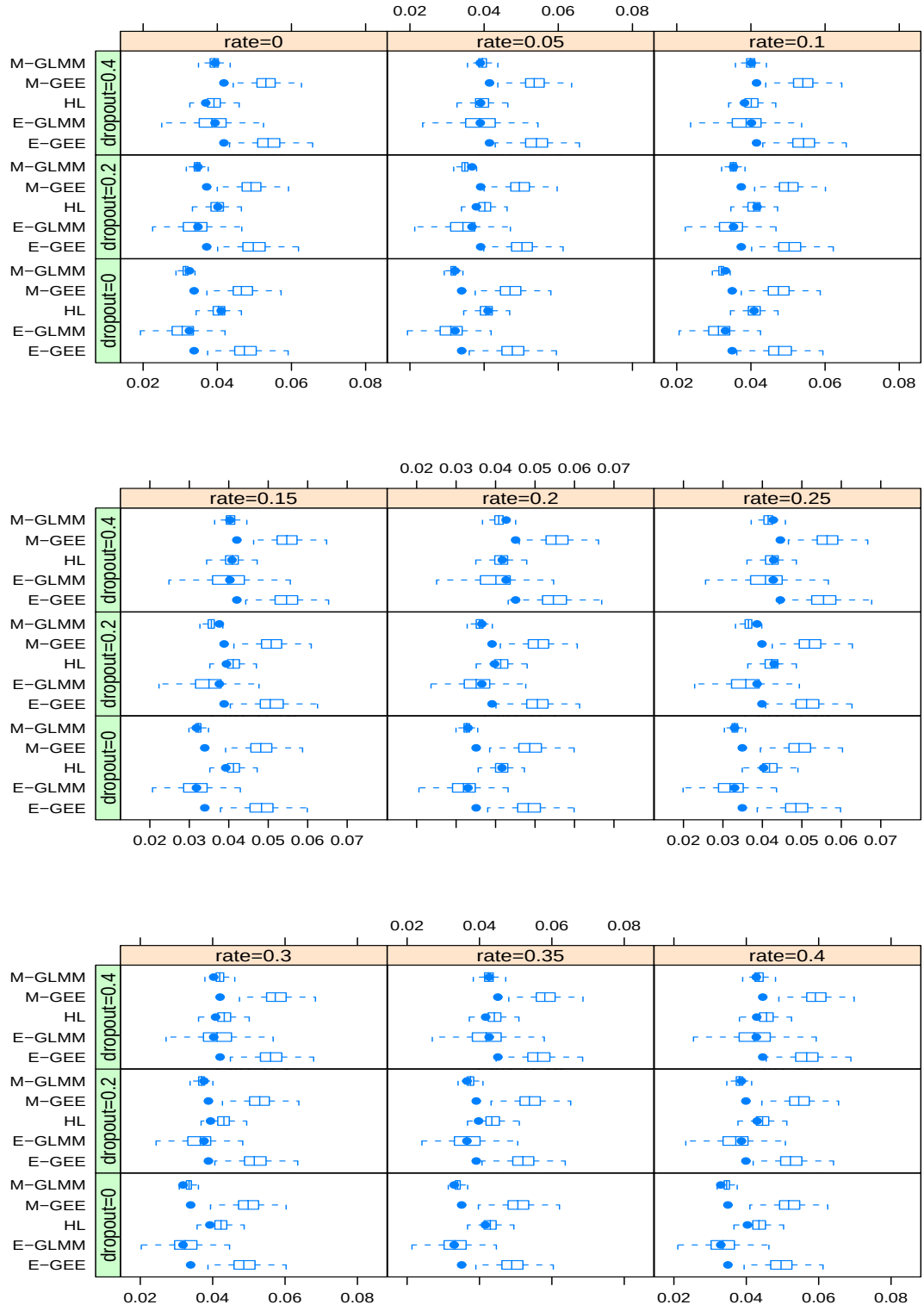

Figure 4: Box plots of all estimated standard errors. The dots in this panel represent sample standard deviation, obtained from estimated coefficients of TREATMENT indicator. Box plots on the same row have the same dropout rate, and same column indicates of the same rate.

		Rate=0%		Rate=5%		Rate=10%		Rate=15%		Rate=20%		Rate=25%		Rate=30%		Rate=35%		Rate=40%	
DROPOUT		CI len	CP	CI len	CP	CI len	CP	CI len	CP	CI len	CP	CI len	CP	CI len	CP	CI len	CP	CI len	CP
0%	HL	0.159	0.954	0.159	0.957	0.160	0.947	0.162	0.958	0.163	0.961	0.164	0.958	0.166	0.950	0.169	0.948	0.171	0.959
	E-GEE	0.187	0.996	0.188	0.995	0.188	0.986	0.190	0.994	0.191	0.992	0.192	0.988	0.192	0.986	0.194	0.991	0.195	0.996
	M-GEE	0.184	0.995	0.186	0.994	0.187	0.989	0.190	0.994	0.192	0.994	0.195	0.990	0.196	0.987	0.200	0.992	0.204	0.997
	E-GLMM	0.123	0.942	0.124	0.917	0.125	0.933	0.126	0.935	0.127	0.920	0.127	0.928	0.129	0.912	0.132	0.931	0.133	0.916
	M-GLMM	0.251	1.000	0.255	1.000	0.258	0.999	0.262	1.000	0.266	1.000	0.271	1.000	0.276	1.000	0.283	1.000	0.290	1.000
10%	HL	0.158	0.954	0.159	0.931	0.161	0.947	0.162	0.941	0.163	0.936	0.165	0.947	0.167	0.933	0.170	0.920	0.173	0.912
	E-GEE	0.191	0.992	0.193	0.994	0.193	0.990	0.196	0.994	0.196	0.992	0.197	0.993	0.198	0.989	0.199	0.989	0.201	0.994
	M-GEE	0.188	0.992	0.191	0.994	0.192	0.990	0.196	0.994	0.197	0.993	0.200	0.994	0.203	0.989	0.205	0.991	0.210	0.994
	E-GLMM	0.130	0.929	0.131	0.939	0.131	0.935	0.133	0.936	0.134	0.937	0.135	0.920	0.137	0.918	0.139	0.934	0.141	0.940
	M-GLMM	0.265	1.000	0.268	1.000	0.272	1.000	0.276	1.000	0.280	1.000	0.285	1.000	0.291	1.000	0.297	0.999	0.304	1.000
20%	HL	0.157	0.968	0.158	0.955	0.160	0.935	0.162	0.931	0.163	0.907	0.166	0.902	0.169	0.871	0.171	0.857	0.175	0.820
	E-GEE	0.197	0.993	0.198	0.986	0.199	0.992	0.200	0.990	0.199	0.983	0.202	0.983	0.204	0.989	0.205	0.981	0.206	0.989
	M-GEE	0.194	0.992	0.196	0.988	0.198	0.994	0.200	0.990	0.200	0.984	0.205	0.986	0.209	0.989	0.212	0.983	0.216	0.993
	E-GLMM	0.138	0.932	0.138	0.914	0.140	0.938	0.141	0.932	0.142	0.926	0.144	0.920	0.146	0.929	0.147	0.900	0.149	0.929
	M-GLMM	0.279	1.000	0.283	1.000	0.287	1.000	0.291	0.999	0.296	0.999	0.302	1.000	0.308	0.999	0.314	0.999	0.322	1.000
30%	HL	0.156	0.960	0.157	0.937	0.159	0.924	0.161	0.896	0.163	0.843	0.165	0.830	0.169	0.778	0.173	0.713	0.176	0.630
	E-GEE	0.204	0.987	0.204	0.986	0.205	0.986	0.206	0.985	0.208	0.989	0.209	0.989	0.210	0.992	0.213	0.979	0.215	0.987
	M-GEE	0.201	0.987	0.203	0.986	0.204	0.989	0.206	0.988	0.209	0.992	0.212	0.989	0.215	0.991	0.220	0.984	0.224	0.988
	E-GLMM	0.148	0.924	0.148	0.943	0.150	0.925	0.151	0.935	0.152	0.930	0.153	0.928	0.156	0.944	0.158	0.913	0.161	0.943
	M-GLMM	0.298	0.999	0.302	1.000	0.307	1.000	0.311	1.000	0.315	1.000	0.322	1.000	0.328	1.000	0.335	1.000	0.342	1.000
40%	HL	0.153	0.976	0.155	0.934	0.157	0.911	0.160	0.846	0.163	0.760	0.166	0.722	0.169	0.602	0.173	0.496	0.179	0.367
	E-GEE	0.212	0.986	0.213	0.992	0.213	0.993	0.215	0.991	0.216	0.975	0.219	0.989	0.221	0.980	0.221	0.983	0.223	0.984
	M-GEE	0.209	0.986	0.211	0.990	0.213	0.992	0.216	0.990	0.218	0.974	0.222	0.991	0.226	0.981	0.229	0.989	0.233	0.985
	E-GLMM	0.158	0.936	0.161	0.934	0.159	0.935	0.163	0.944	0.164	0.918	0.166	0.932	0.169	0.923	0.171	0.931	0.172	0.940
	M-GLMM	0.321	1.000	0.324	1.000	0.329	1.000	0.335	1.000	0.340	1.000	0.346	1.000	0.352	1.000	0.360	1.000	0.369	1.000

Table 1: Confidence interval lengths and coverage probability among all estimates