

STAT/BIOST 571: Homework 3 Solutions

Question 1 (25 points)

Part (1)

We fitted the following model:

$$\text{logit}(\mu_{ij}) = \beta_0 + \beta_1 \text{Age}_{ij} + \beta_2 \text{MatSmoking}_{ij} + \beta_3 \text{Age}_{ij} \times \text{MatSmoking}_{ij}$$

where $\mu_{ij} = E[Y_{ij}] = P[Y_{ij} = 1]$, Y_{ij} is the indicator variable or the presence of wheezing, and $\text{var}(Y_{ij}) = \text{diag} \left\{ \frac{\mu_{ij}(1-\mu_{ij})}{\phi} \right\}$.

The parameter estimates with their corresponding naive SEs are provided in Table 1.

Table 1: Parameter estimates using a GLM

	Estimate	Std. Error
(Intercept)	-1.9008	0.0887
Age	-0.1413	0.0695
Maternal_smoking	0.3140	0.1394
Age:Maternal_smoking	0.0708	0.1107

Part (2)

We fit a similar model than in part (a), but using working independence and exchangeable correlation matrices, and computing robust SEs. The parameter estimates are presented in Table 2.

Table 2: Parameter estimates with GLM and GEE1 approaches

Approach	GEE1				GLM	
Working covariance:	Independence		Exchangeable		Naive logistic regression	
	Estimate	Std.err	Estimate	Std.err	Estimate	Std. Error
(Intercept)	-1.9008	0.1191	-1.9005	0.1191	-1.9008	0.0887
Age	-0.1413	0.0582	-0.1412	0.0582	-0.1413	0.0695
Maternal_smoking	0.3140	0.1878	0.3138	0.1878	0.3140	0.1394
Age:Maternal_smoking	0.0708	0.0883	0.0708	0.0883	0.0708	0.1107

We appreciate that:

- the point estimates for all the coefficients are practically the same with the three methods.
- the robust SEs obtained using a GEE1 with working independence and exchangeable correlation matrices are practically the same.
- comparing the SEs calculated with the naive logistic regression against the robust SEs obtained with the GEE1 approaches, the SEs for the age and the interaction coefficients are slightly larger while the SEs for the intercept and the maternal smoking coefficients are slightly smaller.

Parameter interpretation:

- β_1 (age's coefficient): corresponds to the average increase in the log odds of asthma (wheezing) for a one year increase in age among children with a non-smoker mother.
- β_2 (maternal smoking status coefficient) corresponds to the log odds ratio of asthma (wheezing) comparing children with an smoker mother versus those with a non-smoker mother, among 9-year-old children.
- β_3 (interaction coefficient): corresponds to the difference, between children with an smoker mother versus those with a non-smoker mother, of the average increase in the log odds of asthma (wheezing) for a one year increase in age, that is, $\beta_3 = (\beta_1 + \beta_3) - \beta_1$.

Part (3)

We fitted an alternating logistic regression model, using the same model for the mean as in part (a), and with an exchangeable model for the association, that is,

$$\text{logit}(Y_{ij} = 1|Y_{ik} = y_{ik}) = \alpha y_{ik} + \log \left(\frac{\mu_{ij} - \mu_{ijk}}{1 - \mu_{ij} - \mu_{ik} + \mu_{ijk}} \right),$$

where $\mu_{ijk} = P[Y_{ij} = Y_{ik} = 1]$ and μ_{ij} and Y_{ij} as defined in part (a).

The parameter estimates are presented in Table 3.

Table 3: Parameter estimates using ALR

Parameter	Estimate	Std.Error
Intercept	-1.9002	0.1191
Age	-0.14	0.0583
Maternal smoking	0.3141	0.1878
Interaction	0.0711	0.0883
alpha	2.0404	0.174

The point estimates and SE estimates for the coefficients of the mean model are very similar to those estimated under the GEE1 approaches in part (b).

The common OR for any pair of responses Y_{ij} and Y_{ik} is 7.69, 95% CI 5.47 - 10.82, which implies a strong within-subject correlation.

Part (4)

If the participants had started at different ages, the following model allows us to separate the baseline age effect from the longitudinal time effect:

$$\text{logit}(\mu_{ij}) = \gamma_0 + \gamma_1 \text{Age}_{i0} + \gamma_2 \text{Time}_{ij} + \gamma_3 \text{MatSmoking}_{ij} + \gamma_4 \text{Time}_{ij} \times \text{MatSmoking}_{ij},$$

where Age_{i0} is the baseline age, and $\text{Time}_{ij} = \text{Age}_{ij} - \text{Age}_{i0}$ is the time being in the study.

In this model, γ_1 will give us the baseline age effect while γ_3 and γ_4 (if we decided to include the interaction) will give us the longitudinal time effect.

Question 2 (25 points)

There are a few R packages that can simulate correlated binary data, e.g. `bindata`, `MultiRNG` and `SimCorMultRes`. Note that the NORTA method used in `SinCorMultRes` is based on Gaussian copula, and the user-specified correlation matrix corresponds to the correlation of the latent Gaussian random variables, rather than the binary outcome Y . We will use `bindata` for this problem.

Part (1)

Code is attached in the Appendix.

Part (2)

Table 4 summarizes the average estimates of β and ρ across the 200 simulations. We observe that the empirical biases of these estimates are all close to 0.

	β_0	β_1	β_2	ρ
Estimated (avg)	-1.4926	0.4947	0.4989	0.2495
True	-1.5000	0.5000	0.5000	0.2500
Bias	0.0074	-0.0053	-0.0011	-0.0005

Table 4: Empirical biases of GEE parameter estimates

Part (3)

Table 5 compares the average estimated SEs to the empirical SEs, which are reasonably close to each other.

	$SE(\beta_0)$	$SE(\beta_1)$	$SE(\beta_2)$
Estimated (avg)	0.1621	0.1784	0.0784
Empirical	0.1659	0.1571	0.0803

Table 5: Average of estimated SEs compared with empirical SEs

Part (4)

No, trying to simulate binary data with the specified marginal mean and correlation leads to an error (“Error in Element (1,3): Admissible values are in $[0, 0.1824]$ ”).

To understand this error message, note that for two binary random variables $Y_i \sim \text{Bernoulli}(p_i)$, $i = 1, 2$, the marginal mean p_1, p_2 , together with the correlation $\text{corr}(Y_1, Y_2)$, fully determine the joint probability $\Pr(Y_1 = 1, Y_2 = 1)$ via

$$\frac{\Pr(Y_1 = 1, Y_2 = 1) - p_1 p_2}{\sqrt{p_1 p_2 (1 - p_1)(1 - p_2)}} = \text{corr}(Y_1, Y_2). \quad (1)$$

Taking such Y_1 and Y_2 to be the first and third observation ($X_2 = 0$ and 2 respectively) for a male ($X_1 = 0$) in the setting of this problem, we have $p_1 = 0.1824$, $p_2 = 0.3775$. Then, from (1) we obtain $\Pr(Y_1 = 1, Y_2 = 1) = 0.2093$. However, such joint probability should never exceed the marginal probabilities, i.e. we must have $\Pr(Y_1 = 1, Y_2 = 2) \leq p_1 = 0.1824$, which leads to a contradiction.

Question 3 (25 points)

Part (1)

In HW1 we used the following linear mixed model with random intercept and random slopes, to assess whether the change in cholesterol level over time is related to age, gender and BMI:

$$\begin{aligned} \text{Model 1 : } Y_{ij} = & \beta_0 + \beta_1 \text{Time}_{ij} + \beta_2 \text{Age}_i + \beta_3 \text{Female}_i + \beta_4 \text{BMI}_i + \beta_5 \text{Age}_i \times \text{Time}_{ij} \\ & + \beta_6 \text{Female}_i \times \text{Time}_{ij} + \beta_7 \text{BMI}_i \times \text{Time}_{ij} + b_{0i} + b_{1i} \text{Time}_{ij} + \epsilon_{ij}, \end{aligned}$$

where Age_i and BMI_i are age and BMI at enrollment.

Here, we will also fit the following GEE1 model:

$$\begin{aligned} \text{Model 2 : } E[Y_{ij}] = & \beta_0 + \beta_1 \text{Time}_{ij} + \beta_2 \text{Age}_i + \beta_3 \text{Female}_i + \beta_4 \text{BMI}_i + \beta_5 \text{Age}_i \times \text{Time}_{ij} \\ & + \beta_6 \text{Female}_i \times \text{Time}_{ij} + \beta_7 \text{BMI}_i \times \text{Time}_{ij}, \end{aligned}$$

with an exchangeable working correlation matrix.

We will also compare the Model 2 with the following reduced model:

$$\text{Model 0 : } E[Y_{ij}] = \beta_0 + \beta_1 \text{Time}_{ij} + \beta_2 \text{Age}_i + \beta_3 \text{Female}_i + \beta_4 \text{BMI}_i,$$

with no interactions of time with age, gender or BMI at baseline. By performing a Wald test (`anova` function in R), we obtained a p-value of $p < 0.001$, then, we can reject the null hypothesis that $\beta_5 =$

$\beta_6 = \beta_7 = 0$ and we keep the full model (Model 2). In HW1 we also rejected the corresponding null hypothesis for comparing Model 1 with a reduced model without interactions.

The parameter estimates for the marginal parameters obtained for the fixed effects model 1 and the GEE model 2 are presented in Table 6. The parameter estimates are very similar with both approaches.

Table 6: Population parameter estimates

	LMM		GEE1	
	Estimate	Std. Error	Estimate	Std.err
(Intercept)	129.5611	5.4861	129.9114	4.5300
time	5.7345	0.4929	5.6709	0.7958
age	1.4788	0.0936	1.4692	0.0781
female	-2.6364	1.5260	-3.0547	1.2322
bmi0	1.0823	0.1808	1.0950	0.1568
time:age	-0.0532	0.0083	-0.0597	0.0134
time:female	1.1249	0.1355	1.1331	0.2135
time:bmi0	-0.0448	0.0164	-0.0295	0.0276

Finally, the expected change of cholesterol level by year for a male participant of x years old and z kg/m² of BMI at baseline is $5.73 - 0.05x - 0.05z$ using the LMM and $5.67 - 0.06x - 0.03z$ using the GEE1 approach, and the expected change of cholesterol level by year for a female participant of x years old and z kg/m² of BMI at baseline is $6.86 - 0.05x - 0.05z$ using the LMM and $6.80 - 0.06x - 0.03z$ using the GEE1 approach. In both cases, males and females, the estimated expected change of cholesterol level by year for a participant of x years old and z kg/m² of BMI at baseline is practically the same with both approaches.

Part (2)

(1) Linear mixed model

Merits:

- Provide consistent estimators for the regression coefficients (of the fixed effects) even if the covariance model is misspecified, because they are calculated using weighted least square estimation.
- Useful for data in which the measurements on different units are not made at a common set of times.
- Useful for prediction when incorporate random effects. For example, if we want to predict subject-specific response profiles.
- Allows us to model two sources of variability: between subject and within subject variability.
- The between subject variability allows to capture heterogeneity due to unmeasured factors.

- Interpretation about the regression parameter is independent of the correlation structure.

Limitations:

- The selected correlation model is used to compute SEs for the regression coefficients of the fixed effects. Therefore, when the the covariance model is misspecified, the inference about these coefficients will be compromised.

(2) Generalized Estimating Equations

Merits:

- Regression coefficient estimates are consistent, in large samples, even if the covariance model is misspecified.
- The empirical variance gives valid standard errors for the estimated regression coefficients (in large sample size) even if the correlation model is wrong. Therefore, inference using the empirical (robust) standard errors is correct even with a poor choice for the correlation model.
- Accommodates wide variety of correlation structures: exchangeable, autoregressive, unstructured.

Limitations:

- The empirical variance is valid only in “large samples”.
- It’s a marginal model, its focus is on estimating the average response over the population.

To assess these approaches, we simulated data using the following model

$$Y_{ij} = \beta_0 + \beta_1 \text{Time}_{ij} + \gamma \text{Trt}_i \times \text{Time}_{ij} + b_{0i} + \tau b_{1i} + \epsilon_{ij},$$

where $i = 1, \dots, m$, $j = 1, \dots, n$, Y_{ij} is the biomarker level for patient i at time Time_{ij} , Trt_i is an indicator variable for receiving treatment A , $b_i \sim N_2\left(0, \begin{bmatrix} 1 & 1 \\ 1 & 3 \end{bmatrix}\right)$, and $\epsilon_i = (\epsilon_{i1}, \dots, \epsilon_{in_i}) \sim N_n(0, 4I_n)$.

Note that we did not include a term containing only Trt_i which is the case when we perform a randomized clinical trial, we may assume that the baseline biomarker level in the same in both treatment groups, and we model within subject variability with the random effects.

Under the marginal version of the chosen model, the change in biomarker level per unit increase in time is given by $\beta_1 + \gamma$ in patients receiving treatment A, and by β_1 in patients receiving treatment B. The parameter of interest for this simulation study will be the difference of the change in biomarker level per unit increase in time between both treatment groups, that is, we are interested in γ .

We will calculate empirical bias, power and coverage probability for estimating γ for $\beta_0 = 10$, $\beta_1 = 5$, and combinations of $\gamma \in \{0, 1, 5, 10\}$, $m \in \{20, 50, 100, 200\}$, $n_i = n \in \{2, 6, 10, 20\}$ and $\tau \in \{0, 1\}$.

Note that for $\tau = 0$, the model only includes a random intercept and, for $\tau = 1$, the model also includes both, a random intercept and a random slope.

We will compare a LMM with only random intercept and a GEE model with the identity as the link function and an exchangeable correlation covariance matrix. Hence, when $\tau = 0$ the correlation model is correct with both approaches, and when $\tau = 1$, the working covariance-matrix is misspecified with both approaches.

The R code for these simulations is provided as appendix at the end of the document.

Bias

The absolute empirical bias for different values of γ , when the covariance model is correct ($\tau = 0$), is displayed in the following plots.

The empirical bias is low in all cases and improves as sample size increases. This make sense since both approaches provide consistent estimates for the regression coefficients.

The absolute empirical bias for different values of γ , when the covariance model is incorrect ($\tau = 1$), is displayed in the following plots.

The empirical bias is low in all cases and improves as sample size increases, however is higher than the empirical bias when the model is correctly specified.

Power

The power for different values of γ , when the covariance model is correct ($\tau = 0$), is displayed in the following plots.

We appreciate that, when the covariance model is correct, both approaches provided good type I error and high power for all sample sizes and all the studied values of n (number of longitudinal measurements per subject). As expected, the type I error improves as sample size increases.

The power for different values of γ , when the covariance model is incorrect ($\tau = 1$), is displayed in the following plots.

We appreciate that, when the covariance model is misspecified, the GEE approach still provides low type I error and this error improves as sample size increases. In contrast, the LMM approach gives high type I error for all sample sizes and for all different values of n . Also, the power is good with both approaches when the effect is relatively large ($\gamma \in \{5, 10\}$). Moreover, for a small effect ($\gamma = 1$), the LMM approach has better power than the GEE approach for small sample sizes, and both approaches have good power for large sample size. As previously discussed, for large sample size, the inference using the empirical (robust) standard errors with the GEE approach is correct even with a poor choice for the correlation model.

Coverage probability

The coverage probability for different values of γ , when the covariance model is correct ($\tau = 0$), is displayed in the following plots.

The coverage probability is high with both approaches, always above 0.90.

The coverage probability for different values of γ , when the covariance model is incorrect ($\tau = 1$), is displayed in the following plots.

As expected, when the covariance model is incorrect, the coverage probability is low with the LMM approach but very good with the GEE approach. Also, as observed with the power, for the GEE approach the coverage probability improves as sample size increases.

Finally, if giving more time, I would simulate scenarios with other different covariance matrices, scenarios where the mean outcome also depends on other covariates (for example age and sex), scenarios where the number of observations and the time at which the outcomes were measured differ among the subjects. And maybe scenarios with missing data with the three types of missing mechanisms: MCAR, MAR and non-ignorable missing.

Question 4 (25 points)

Let $\mathbf{Y}_i = (Y_{i1}, Y_{i2})$ where Y_{i1} is continuous and Y_{i2} is binary. We specify the mean model as $\mu_{i1} = \mathbb{E}[Y_{i1}] = X_i^\top \beta_1$ and $\mu_{i2} = \mathbb{E}[Y_{i2}] = g^{-1}(X_i^\top \beta_2)$, where $g(\cdot)$ is a link function such as the logit, probit or complementary log-log link. We illustrate with the logit link in our derivation for concreteness.

For the variance model, we have $v(\mu_{i1}) = 1$ and $v(\mu_{i2}) = \mu_{i2}(1 - \mu_{i2})$. We assume a dispersion parameter σ^2 for Y_{i1} so that $Var(Y_{i1}) = \sigma^2 v(\mu_{i1}) = \sigma^2$. Note that we do not have a dispersion parameter for Y_{i2} , since the variance of a Bernoulli random variable is fully determined by the mean. Also, we adopt an independent working correlation structure.

Putting these together and letting $\beta := (\beta_1, \beta_2)^\top$, we define the following quantities:

$$\begin{aligned}\boldsymbol{\mu}_i &:= \begin{bmatrix} \mu_{i1} \\ \mu_{i2} \end{bmatrix} = \begin{bmatrix} X_i^\top \beta_1 \\ \text{expit}(X_i^\top \beta_2) \end{bmatrix} \in \mathbb{R}^{2 \times 1} \\ \mathbf{D}_i &:= \frac{\partial \boldsymbol{\mu}_i}{\partial \beta^\top} = \begin{bmatrix} \frac{\partial \mu_{i1}}{\partial \beta_1^\top} & \frac{\partial \mu_{i1}}{\partial \beta_2^\top} \\ \frac{\partial \mu_{i2}}{\partial \beta_1^\top} & \frac{\partial \mu_{i2}}{\partial \beta_2^\top} \end{bmatrix} = \begin{bmatrix} X_i^\top & 0 \\ 0 & X_i^\top \frac{\exp(X_i^\top \beta_2)}{[1 + \exp(X_i^\top \beta_2)]^2} \end{bmatrix} = X_i^\top \begin{bmatrix} 1 & 0 \\ 0 & \mu_{i2}(1 - \mu_{i2}) \end{bmatrix} \in \mathbb{R}^{2 \times 2p} \\ \mathbf{V}_i &:= \begin{bmatrix} \sigma^2 & 0 \\ 0 & \mu_{i2}(1 - \mu_{i2}) \end{bmatrix} \in \mathbb{R}^{2 \times 2}.\end{aligned}$$

Then our model is

$$\begin{aligned}\mathbb{E}[\mathbf{Y}_i] &= \boldsymbol{\mu}_i = \begin{bmatrix} X_i^\top \beta_1 \\ \text{expit}(X_i^\top \beta_2) \end{bmatrix}, \\ Cov(\mathbf{Y}_i) &= \mathbf{V}_i = \begin{bmatrix} \sigma^2 & 0 \\ 0 & \mu_{i2}(1 - \mu_{i2}) \end{bmatrix}\end{aligned}$$

where the off-diagonal entries of \mathbf{V}_i do not need to be correctly specified. With these notations defined, we solve for β from the GEE

$$\sum_{i=1}^m \mathbf{D}_i^\top \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) = 0.$$

In addition, we can define the sandwich estimator for $Cov(\hat{\beta})$ in our usual way, i.e. we can estimate $\widehat{Cov}(\hat{\beta}) = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$ where

$$\begin{aligned} \mathbf{A} &= \frac{1}{m} \sum_{i=1}^m \mathbf{D}_i^\top \mathbf{V}_i^{-1} \mathbf{D}_i \\ \mathbf{B} &= \frac{1}{m} \sum_{i=1}^m \mathbf{D}_i^\top \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) (\mathbf{Y}_i - \boldsymbol{\mu}_i)^\top \mathbf{V}_i^{-1} \mathbf{D}_i. \end{aligned}$$

Question 5 (Bonus)

Consider longitudinal/clustered data $(Y_{ij}, \mathbf{X}_{ij})$, where i indicates subject i ($i = 1, \dots, m$) and j indicates the j th observation of the i th subject ($j = 1, \dots, n_i$). For simplicity, we assume $n_i = n$. Denote by $\mu_{ij} = E(Y_{ij} | \mathbf{X}_i)$ the marginal mean of Y_{ij} conditional on the covariates \mathbf{X}_i and $var(Y_{ij} | \mathbf{X}_i) = v(\mu_{ij})$ the marginal variance of Y_{ij} . Suppose the marginal mean μ_{ij} satisfies the generalized linear model

$$g(\mu_{ij}) = \mathbf{X}_{ij}^T \boldsymbol{\beta}, \quad (2)$$

where $g(\cdot)$ is a monotonic differential link function, $\boldsymbol{\beta}$ is a $p \times 1$ vector of regression coefficients. Let $\mathbf{Y}_i = (Y_{i1}, \dots, Y_{in})^T$, $\boldsymbol{\mu}_i = (\mu_{i1}, \dots, \mu_{in})^T$ and $\mathbf{X}_i = (\mathbf{X}_{i1}, \dots, \mathbf{X}_{in})^T$. Consider the Generalized Estimating Equations (GEEs)

$$\mathbf{U}_\beta(\boldsymbol{\beta}, \alpha) = \sum_{i=1}^m \mathbf{D}_i^T \mathbf{V}_i^{-1}(\alpha) \{\mathbf{Y}_i - \boldsymbol{\mu}_i(\boldsymbol{\beta})\} \quad (3)$$

where $\mathbf{D}_i = \partial \boldsymbol{\mu}_i / \partial \boldsymbol{\beta}^T = \boldsymbol{\Delta}_i^{-1} \mathbf{X}_i$ and $\boldsymbol{\Delta}_i = \text{diag}\{g'(\mu_{i1}), \dots, g'(\mu_{in})\}$, $\mathbf{V}_i(\alpha) = \mathbf{V}_{Mi}^{1/2} \mathbf{R}_i(\alpha) \mathbf{V}_{Mi}^{1/2}$ is a working covariance matrix, $\mathbf{R}_i(\alpha)$ is a working correlation matrix assuming an exchangeable correlation, α is the corresponding working exchangeable correlation parameter, $\mathbf{V}_{Mi} = \text{diag}\{v(\mu_{ij})\}$ contains the marginal variances of the Y_{ij} .

Suppose the mean model is correctly specified by the GLM (2). The working correlation $\mathbf{R}_i(\alpha)$ might be misspecified. Show that under appropriate regularity conditions, the GEE estimator $\hat{\boldsymbol{\beta}}$ is asymptotically normal

$$\sqrt{m}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \xrightarrow{d} N(0, \boldsymbol{\Sigma}),$$

where $\boldsymbol{\Sigma} = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$, and

$$\mathbf{A} = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \mathbf{D}_i^T \mathbf{V}_i^{-1} \mathbf{D}_i \quad (4)$$

$$\mathbf{B} = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \mathbf{D}_i^T \mathbf{V}_i^{-1} (\mathbf{Y}_i - \boldsymbol{\mu}_i) (\mathbf{Y}_i - \boldsymbol{\mu}_i)^T \mathbf{V}_i^{-1} \mathbf{D}_i, \quad (5)$$

\mathbf{A}^{-1} is called the model-based covariance of $\hat{\beta}$ and \mathbf{B}^{-1} is called the empirical covariance of $\hat{\beta}$. When the working covariance \mathbf{V}_i is correctly specified, $\mathbf{A} = \mathbf{B}$ and $\Sigma = \mathbf{A}^{-1}$.

Answer. First, let's assume there exist an estimator $\hat{\alpha}$ of α , such that $\hat{\alpha} = \hat{\alpha}(\beta)$ only depends on β and $\sqrt{m}(\hat{\alpha} - \alpha) = O_p(1)$. For the exchangeable case, we know that the moment estimator for α satisfies $\sqrt{m}(\hat{\alpha} - \alpha) = O(1)$ and only depends on β .

Then, under some regularity conditions we may apply the implicit function theorem to expressed the GEE as

$$\mathbf{U}_\beta(\beta, \hat{\alpha}(\beta)) = \sum_{i=1}^m \mathbf{D}_i^T \mathbf{V}_i(\hat{\alpha}(\beta))^{-1} \{\mathbf{Y}_i - \boldsymbol{\mu}_i(\beta)\} = \mathbf{0},$$

and since $\hat{\beta}$ solves the previous equations, using the Taylor expansion around β (the true parameter), that is, the Delta method, we get

$$\mathbf{0} = \mathbf{U}_\beta(\hat{\beta}, \hat{\alpha}(\hat{\beta})) = \mathbf{U}_\beta(\beta, \hat{\alpha}(\beta)) + \frac{\partial \mathbf{U}_\beta(\beta, \hat{\alpha}(\beta))}{\partial \beta} [\hat{\beta} - \beta] + \tilde{\mathbf{R}}_m,$$

where $\tilde{\mathbf{R}}_m = o_P(\|\hat{\beta} - \beta\|)$, then

$$\sqrt{m} [\hat{\beta} - \beta] = \left[-\frac{\partial \mathbf{U}_\beta(\beta, \hat{\alpha}(\beta))}{\partial \beta} \right]^{-1} \sqrt{m} [\mathbf{U}_\beta(\beta, \hat{\alpha}(\beta))] + \underbrace{\sqrt{m} \left[-\frac{\partial \mathbf{U}_\beta(\beta, \hat{\alpha}(\beta))}{\partial \beta} \right]^{-1} \tilde{\mathbf{R}}_m}_{\tilde{\mathbf{R}}_m^*}, \quad (**)$$

with $\tilde{\mathbf{R}}_m^* = o_P(\sqrt{m}\|\hat{\beta} - \beta\|) = o_P(1)$, and

$$\frac{\partial \mathbf{U}_\beta(\beta, \hat{\alpha}(\beta))}{\partial \beta} = \frac{\partial \mathbf{U}_\beta(\beta, \hat{\alpha}(\beta))}{\partial \beta} + \frac{\partial \mathbf{U}_\beta(\beta, \hat{\alpha}(\beta))}{\partial \alpha} \frac{\partial \hat{\alpha}}{\partial \beta}.$$

Now, by Delta method again, using that $\sqrt{m}(\hat{\alpha}(\beta) - \alpha) = O_p(1)$, we obtain

$$\mathbf{U}_\beta(\beta, \hat{\alpha}(\beta)) = \mathbf{U}_\beta(\beta, \alpha) + \frac{\partial \mathbf{U}_\beta(\beta, \alpha)}{\partial \alpha} (\hat{\alpha} - \alpha) + o_P(|\hat{\alpha} - \alpha|),$$

which implies

$$\frac{\mathbf{U}_\beta(\beta, \hat{\alpha}(\beta))}{\sqrt{m}} = \frac{\mathbf{U}_\beta(\beta, \alpha)}{\sqrt{m}} + \underbrace{\frac{1}{m} \frac{\partial \mathbf{U}_\beta(\beta, \alpha)}{\partial \alpha}}_{o_P(1)} \underbrace{\sqrt{m}(\hat{\alpha} - \alpha)}_{O_P(1)} + o_P(1) = \frac{\mathbf{U}_\beta(\beta, \alpha)}{\sqrt{m}} + o_P(1),$$

since $o_P(1)O_P(1) + o_P(1) = o_P(1) + o_P(1) = o_P(1)$.

Note that, since $E[\mathbf{Y}_i - \boldsymbol{\mu}_i] = \mathbf{0}$ and the m subjects are independent, by the Lindeberg-Feller CLT (assuming the Linderberg condition holds), we have

$$\frac{\mathbf{U}_\beta(\beta, \hat{\alpha}(\beta))}{\sqrt{m}} = \sqrt{m} \frac{1}{m} \sum_{i=1}^m \mathbf{D}_i^T \mathbf{V}_i^{-1}(\alpha) \{\mathbf{Y}_i - \boldsymbol{\mu}_i(\beta)\} \xrightarrow{d} N(\mathbf{0}, \mathbf{B}), \quad (***)$$

where

$$\mathbf{B} = \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m \mathbf{D}_i^T \mathbf{V}_i^{-1} E[(\mathbf{Y}_i - \boldsymbol{\mu}_i)(\mathbf{Y}_i - \boldsymbol{\mu}_i)^T] \mathbf{V}_i^{-1} \mathbf{D}_i.$$

Now, note that

$$\begin{aligned}
E \left[\frac{\delta (D_i^T V_i^{-1}(\hat{\alpha})[Y_i - \mu_i])}{\delta \beta} \right] &= E \left[\frac{\partial (D_i^T V_i^{-1}(\hat{\alpha})[Y_i - \mu_i])}{\partial \beta} \right] + E \left[\frac{\partial (D_i^T V_i^{-1}(\hat{\alpha})[Y_i - \mu_i])}{\partial \alpha} \frac{\partial \hat{\alpha}}{\partial \beta} \right] \\
&= E \left[\frac{\partial (D_i^T V_i^{-1}(\hat{\alpha})[Y_i - \mu_i])}{\partial \beta} \right] + \frac{\partial (D_i^T V_i^{-1}(\hat{\alpha}))}{\partial \alpha} \underbrace{E[Y_i - \mu_i]}_{=0} \\
&= D_i^T V_i^{-1}(\hat{\alpha}) E \left[\frac{\partial}{\partial \beta} (Y_i - \mu_i) \right] + \frac{\partial}{\partial \beta} (D_i^T V_i^{-1}(\hat{\alpha})) \underbrace{E[Y_i - \mu_i]}_{=0} \\
&= -D_i^T V_i^{-1}(\hat{\alpha}) D_i.
\end{aligned}$$

Then, by the general WLLN and the consistency of $\hat{\alpha}$, we have

$$\frac{1}{m} \frac{\delta U_{\beta}(\beta, \hat{\alpha}(\beta))}{\delta \beta} \xrightarrow{p} \mathbf{A},$$

where

$$\mathbf{A} = - \lim_{m \rightarrow \infty} \frac{1}{m} \sum_{i=1}^m D_i^T V_i^{-1} D_i.$$

Substituting last expression and $(\star\star\star)$ in $(\star\star)$, and applying continuing mapping theorem and Slutsky's lemma, gives us

$$\sqrt{m} [\hat{\beta} - \beta] \xrightarrow{d} N(\mathbf{0}, \Sigma),$$

where $\Sigma = \mathbf{A}^{-1} \mathbf{B} \mathbf{A}^{-1}$, as desired.

When the working covariance V_i is correctly specified, we have $E[(Y_i - \mu_i)(Y_i - \mu_i)^T] = V_i$, and \mathbf{B} becomes \mathbf{A} , then $\Sigma = \mathbf{A}^{-1}$.

Appendix: R code

Question 1

(a)

```
df1 <- read.table("sixcity.dat", header = F, col.names = c("wheezing", "id", "Age",
  mod1 <- glm(wheezing~Age*Maternal_smoking, family = binomial(link = "logit"), data =
df1)

results_q1a <- summary(mod1)$coefficients[,c(1,2)] kbl(round(results_q1a,4), align
= "lcc", booktabs = TRUE, caption = "Parameter estimates using a GLM") %>% kableExtra::kable_sty
= "center", latex.options = "hold.position")
```

(b)

```
mod2 <- geeglm(wheezing~Age*Maternal_smoking, id = id, data = df1, family = binomial, corstr =
"independence")
```

```
mod3 <- geeglm(wheezing~Age*Maternal_smoking, id = id, data = df1, family = binomial, corstr =
"exchangeable")
```

```
results_q1b1 <- summary(mod2)$coefficients[,c(1,2)]
results_q1b2 <- summary(mod3)$coefficients[,c(1,2)]
kbl(cbind(results_q1b1,results_q1b2, results_q1a),
    align = "cccc", booktabs = TRUE,
    caption = "Parameter estimates with GLM and GEE1 approaches", digits = 4) %>%
    add_header_above(c("Working covariance:" = 1, "Independence" = 2, "Exchangeable" = 2,
        "Naive logistic regression"=2))%>%
    add_header_above(c("Approach"=1, "GEE1"=4, "GLM"=2)) %>%
    kableExtra::kable_styling(position = "center", latex_options = "hold_position")
```

(c)

```
df1$int_age_matsmk <- df1$Age*df1$Maternal_smoking y <- as.matrix(1*df1$wheezing) x <-
as.matrix(cbind(1,df1[,3:5])) id <- as.matrix(df1$id)
```

```
mod4 <- alr(y~x-1, id = id, depm="exchangeable", ainit=0.01)
```

```
results_q1c <- c(mod4coefficients,mod4alpha) results_q1c <- matrix(c(c("Intercept", "Age", "Mater-
nal smoking", "Interaction", "alpha"), round(results_q1c,4),round(sqrt(diag(mod4$variance)),4)), ncol
= 3)
```

```
alpha_est <- round(exp(mod4alpha + c(0, -1, 1) * qnorm(0.975) * sqrt(mod4variance[5,5])),2)
```

```
kbl(results_q1c,
    align = "lcc", booktabs = TRUE, caption = "Parameter estimates using ALR", digits = 4) %>%
    add_header_above(c("Parameter"=1, "Estimate"=1, "Std.Error"=1))%>%
    kableExtra::kable_styling(position = "center", latex_options = "hold_position")
```

Question 2

```
library(bindata)
library(gee)
```

```
m = 300
n = 3
bt = c(-1.5, .5, .5)
rho = .25
```

```
sim.one = function(seed){
  set.seed(seed)
```

```

# design matrices for male and female
X.m = cbind(1, rep(0,n), 0:(n-1))
X.f = cbind(1, rep(1,n), 0:(n-1))
# generate outcome y
y.m = rmvbin(n = m/2, margprob = exp(X.m%*%bt)/(1 + exp(X.m%*%bt)),
             bincorr = matrix(rho, ncol = n, nrow = n) + diag(1-rho, n))
y.f = rmvbin(n = m/2, margprob = exp(X.f%*%bt)/(1 + exp(X.f%*%bt)),
             bincorr = matrix(rho, ncol = n, nrow = n) + diag(1-rho, n))
# combine into a dataset
dat = data.frame(id = rep(1:m, each = n),
                 female = factor(rep(c(0,1), each = m/2*n)),
                 time = rep(0:(n-1), m),
                 y = c(t(rbind(y.m, y.f))))
# fit GEE1
mod = gee(y ~ female + time, family = binomial("logit"),
          id = id, data = dat, corstr = "exchangeable")

return(list(bt.est = coef(mod),
            sandwich.se = summary(mod)$coefficients[, "Robust S.E."],
            corr.est = summary(mod)$working.correlation[1,2]))
}

rslt = lapply(1:200, sim.one)
rslt.summary = matrix(unlist(rslt), nrow = 200, byrow = TRUE)
colnames(rslt.summary) = c("bt0.est", "bt1.est", "bt2.est",
                          "bt0.se", "bt1.se", "bt2.se", "corr.est")
colMeans(rslt.summary) # average estimates
apply(rslt.summary[,1:3], 2, sd) # empirical sd of estimated beta's

out.part2 = rbind(colMeans(rslt.summary[,-(4:6)]), c(bt, rho),
                  colMeans(rslt.summary[,-(4:6)]) - c(bt, rho))
rownames(out.part2) = c("Estimated (avg)", "True", "Bias")
xtable::xtable(out.part2, digits = 4)

out.part3 = rbind(colMeans(rslt.summary[,4:6]), apply(rslt.summary[,1:3], 2, sd))
rownames(out.part3) = c("Estimated (avg)", "Empirical")
xtable::xtable(out.part3, digits = 4)

# trying to generate data with correlation 0.75 leads to an error
try({
  rho = .75
  X.m = cbind(1, rep(0,n), 0:(n-1))
  X.f = cbind(1, rep(1,n), 0:(n-1))
  # generate outcome y
  y.m = rmvbin(n = m/2, margprob = exp(X.m%*%bt)/(1 + exp(X.m%*%bt)),
               bincorr = matrix(rho, ncol = n, nrow = n) + diag(1-rho, n))

```

```

y.f = rmvbin(n = m/2, margprob = exp(X.f%%bt)/(1 + exp(X.f%%bt)),
             bincorr = matrix(rho, ncol = n, nrow = n) + diag(1-rho, n))
})

```

Question 3

(a)

```

dataf <- read.table("framingham.dat", header = F, na.strings = "-9", col.names = c("age",
                                                                                     "gender", "bmi0

```

```

dataf <- dataf[!is.na(dataf$bmi0),] datafid <- -1 : nrow(dataf) dataffemale <- 1(dataf$gender==2)
dataf <- dataf %>% mutate( bmi0_cat = 1(bmi0>18.5)+1(bmi0>25)+1(bmi0>30), BMI_classification
= fct_recode(as.factor(bmi0_cat), "Underweighted"="0", "Normal weight" = "1", "Overweighted"="2",
"Obese" = "3"), age_cat = 1(age>35) + 1(age>45) + 1*(age>55), Age_category = fct_recode(as.factor(age_cat),
"26 - 35 yr" = "0", "36 - 45 yr" = "1", "46 - 55 yr" = "2", "56 - 65 yr" = "3"
), Gender = fct_recode(as.factor(gender), "Female" = "2", "Male" = "1" ) )

```

```

dat.long <- reshape(dataf, idvar = "id",
                    varying = list(c(6,7,8,9,10,11)),
                    v.names = "chol", times=c(0,2,4,6,8,10),
                    direction = "long")

```

```

{r eval=FALSE} mod5 <- lmer (chol~ time + age + female +bmi0+ time*age + time*female+ time*bmi0
+ (1+time|id), REML = T, data = dat.long)

```

```

mod6 <- geeglm(chol~ time + age + female +bmi0+ timeage + timefemale+ time*bmi0, id = id, data
= dat.long, corstr = "exchangeable") summary(mod6)

```

```

mod6red <- geeglm(chol~ time + age + female +bmi0, id = id, data = dat.long, corstr = "exchange-
able")

```

```

anova(mod6, mod6red)

```

```

results_q3a1 <- summary(mod5)$coefficients[,c(1,2)]
results_q3a2 <- summary(mod6)$coefficients[,c(1,2)]
kbl(cbind(round(results_q3a1,4),round(results_q3a2,4)),
     align = "cccc", booktabs = TRUE,
     caption = "Population parameter estimates", digits = 4) %>%
add_header_above(c(" " =1, "LMM"=2, "GEE1"=2)) %>%
kableExtra::kable_styling(position = "center", latex_options = "hold_position")

```

(b)

```
cov_matrix <- matrix(c(1,1,1,3), ncol = 2) simulate_data <- function(m, n, g, tau){ id = 1:m trt =  
rep(0,m) trt[sample(id, m/2)] = 1 # Assigning patients to trt A Y = array(0, dim = c(n,m)) fu_times  
= 3*(1:n-1)
```

```
for(i in 1:m){ b = rmvnorm(1, c(0,0), cov_matrix) e = rnorm(n,2) Y[,i] = 10+5fu_times+gtrt[i]fu_times  
+ b[1]+ taub[2]fu_times + e } y = as.vector(Y) id <- as.vector(t(array(rep(1:m, n), dim = c(m,n))))  
trt <- as.vector(matrix(rep(1,n), nrow = n)%%trt) time <- rep(fu_times, m) data.frame(id = id, trt =  
trt, time = time, y = y) }
```

```
lmm_approach <- function(df){  
  mod <- lmer (y~ time + I(time*trt) + (1|id), REML = F, data = df)  
  d_hat <- summary(mod)$coefficients[3,1]  
  se_dhat <- summary(mod)$coefficients[3,2]  
  conf_int <- d_hat+c(-1,1)*qnorm(0.975)*se_dhat  
  mod0 <- lmer(y~ time+(1|id), REML = F, data = df)  
  
  p.value <- anova(mod0,mod)$'Pr(>Chisq)'[2]  
  c(d_hat, p.value, conf_int)  
}
```

```
gee_approach <- function(df){  
  mod <- geeglm(y~time + I(time*trt), id = id, data = df, corstr = "exchangeable")  
  d_hat <- summary(mod)$coefficients[3,1]  
  se_dhat <- summary(mod)$coefficients[3,2]  
  conf_int <- d_hat+c(-1,1)*qnorm(0.975)*se_dhat  
  p.value <- summary(mod)$coefficients[3,4]  
  c(d_hat, p.value, conf_int)  
}
```

```
nsim <- 200
```

```
delta <- c(0,1,5,10) msamp<-c(20,50,100,200,400) nobs <- c(3,6,10) tau <- c(0,1)
```

```
alpha_lmm <- array(rep(0, 120), dim = c(4,5,3,2)) alpha_gee <- array(rep(0, 120), dim = c(4,5,3,2))
```

```
bias_lmm <- array(rep(0, 120), dim = c(4,5,3,2)) bias_gee <- array(rep(0, 120), dim = c(4,5,3,2))
```

```
coverage_lmm <- array(rep(0, 120), dim = c(4,5,3,2)) coverage_gee <- array(rep(0, 120), dim = c(4,5,3,2))
```

```
for(k in seq_along(delta)){ for(i in seq_along(msamp)){ for(j in seq_along(nobs)){ for(h in seq_along(tau)){  
data <- map(1:nsim, function(x)simulate_data(msamp[i],nobs[j],delta[k],tau[h]))
```

```
## LMM approach
```

```
results <- map(1:nsim, function(x)lmm_approach(data[[x]]))
```

```

alpha_lmm[k,i,j,h] = mean(map_dbl(results,2)<0.05)
bias_lmm[k,i,j,h] = abs(mean(map_dbl(results,1))-delta[k])
coverage_lmm[k,i,j,h] = mean((map_dbl(results,3)<delta[k])*(map_dbl(results,4)>delta[k]))

## GEE approach
results <- map(1:nsim, function(x)gee_approach(data[[x]]))
alpha_gee[k,i,j,h] = mean(map_dbl(results,2)<0.05)
bias_gee[k,i,j,h] = abs(mean(map_dbl(results,1))-delta[k])
coverage_gee[k,i,j,h] = mean((map_dbl(results,3)<delta[k])*(map_dbl(results,4)>delta[k]))
}
}

} }

## Empirical bias under correct covariance model

for(k in 1:4){
  ## a = 0
  label_plot <- paste("Gamma =", delta[k], ", Tau = 0")
  plot(msamp, bias_lmm[k,,1,1], type = "b", col="tomato", ylim = c(0,0.15),
       xlab = c("n_samp"), ylab=c("|Empirical bias|"), main = label_plot, lty = 2)
  lines(msamp, bias_gee[k,,1,1], type = "b", col = "lightblue", lty = 3, lwd = 1)
  lines(msamp, bias_lmm[k,,2,1], type = "b", col = "red", lty = 2, lwd = 1)
  lines(msamp, bias_gee[k,,2,1], type = "b", col = "blue", lty = 3, lwd = 1)
  lines(msamp, bias_lmm[k,,3,1], type = "b", col = "darkred", lty = 2, lwd = 1)
  lines(msamp, bias_gee[k,,3,1], type = "b", col = "darkblue", lty = 3, lwd = 1)
  if(k == 1){
    legend(c("LMM, n = 3", "LMM, n = 6", "LMM, n = 10", "GEE, n = 3","GEE, n = 6", "GEE, n = 10"),
          x = "topright",cex=.8,col=c(c("tomato", "red", "darkred", "lightblue", "blue","darkblue"),
          lty=c(2,2,2,3,3,3))
  }else{
    legend(c("LMM, n = 3", "LMM, n = 6", "LMM, n = 10", "GEE, n = 3","GEE, n = 6", "GEE, n = 10"),
          x = "right",cex=.8,col=c(c("tomato", "red", "darkred", "lightblue", "blue","darkblue")),
          lty=c(2,2,2,3,3,3))
  }
}

## Empirical bias under incorrect covariance model

for(k in 1:4){ ## a = 0 label_plot <- paste("Gamma =", delta[k], ", Tau = 1") plot(msamp, bias_lmm[k,1,2],
type = "b", col="tomato", ylim = c(0,0.15), xlab = c("n_samp"), ylab=c("|Empirical bias|"), main
= label_plot, lty = 2) lines(msamp, bias_gee[k,1,2], type = "b", col = "lightblue", lty = 3, lwd = 1)
lines(msamp, bias_lmm[k,2,2], type = "b", col = "red", lty = 2, lwd = 1) lines(msamp, bias_gee[k,2,2],

```



```

type = "b", col = "blue", lty = 3, lwd = 1) lines(msamp, bias_lmm[k,,3,2], type = "b", col = "dark-
red", lty = 2, lwd = 1) lines(msamp, bias_gee[k,,3,2], type = "b", col = "darkblue", lty = 3, lwd
= 1) if(k == 1){ legend(c("LMM, n = 3", "LMM, n = 6", "LMM, n = 10", "GEE, n = 3","GEE,
n = 6", "GEE, n = 10"), x = "topright",cex=.8,col=c(c("tomato", "red", "darkred", "lightblue",
"blue","darkblue")), lty=c(2,2,2,3,3,3)) }else{ legend(c("LMM, n = 3", "LMM, n = 6", "LMM, n =
10", "GEE, n = 3","GEE, n = 6", "GEE, n = 10"), x = "right",cex=.8,col=c(c("tomato", "red",
"darkred", "lightblue", "blue","darkblue")), lty=c(2,2,2,3,3,3)) } }

```

Power under correct covariance model

```

for(k in 1:4){
  ## a = 0
  label_plot <- paste("Gamma =", delta[k], ", Tau = 0")
  plot(msamp, alpha_lmm[k,,1,1], type = "b", col="tomato", ylim = c(0,1),
       xlab = c("m_samp"), ylab=c("Power"), main = label_plot, lty = 2)
  lines(msamp, alpha_gee[k,,1,1], type = "b", col = "lightblue", lty = 3, lwd = 1)
  lines(msamp, alpha_lmm[k,,2,1], type = "b", col = "red", lty = 2, lwd = 1)
  lines(msamp, alpha_gee[k,,2,1], type = "b", col = "blue", lty = 3, lwd = 1)
  lines(msamp, alpha_lmm[k,,3,1], type = "b", col = "darkred", lty = 2, lwd = 1)
  lines(msamp, alpha_gee[k,,3,1], type = "b", col = "darkblue", lty = 3, lwd = 1)
  if(k == 1){
    legend(c("LMM, n = 3", "LMM, n = 6", "LMM, n = 10", "GEE, n = 3","GEE, n = 6", "GEE, n = 10"),
          x = "topright",cex=.8,col=c(c("tomato", "red", "darkred", "lightblue", "blue","darkblue"),
          lty=c(2,2,2,3,3,3))
  }else{
    legend(c("LMM, n = 3", "LMM, n = 6", "LMM, n = 10", "GEE, n = 3","GEE, n = 6", "GEE, n = 10"),
          x = "right",cex=.8,col=c(c("tomato", "red", "darkred", "lightblue", "blue","darkblue")),
          lty=c(2,2,2,3,3,3))
  }
}
}

```

Power under incorrect covariance model

```

for(k in 1:4){ ## a = 0 label_plot <- paste("Gamma =", delta[k], ", Tau = 1") plot(msamp, al-
pha_lmm[k,,1,2], type = "b", col="tomato", ylim = c(0,1), xlab = c("m_samp"), ylab=c("Power"),
main = label_plot, lty = 2) lines(msamp, alpha_gee[k,,1,2], type = "b", col = "lightblue", lty = 3,
lwd = 1) lines(msamp, alpha_lmm[k,,2,2], type = "b", col = "red", lty = 2, lwd = 1) lines(msamp,
alpha_gee[k,,2,2], type = "b", col = "blue", lty = 3, lwd = 1) lines(msamp, alpha_lmm[k,,3,2], type =
"b", col = "darkred", lty = 2, lwd = 1) lines(msamp, alpha_gee[k,,3,2], type = "b", col = "darkblue",
lty = 3, lwd = 1) if(k == 1){ legend(c("LMM, n = 3", "LMM, n = 6", "LMM, n = 10", "GEE,
n = 3","GEE, n = 6", "GEE, n = 10"), x = "topright",cex=.8,col=c(c("tomato", "red", "darkred",
"lightblue", "blue","darkblue")), lty=c(2,2,2,3,3,3)) }else{ legend(c("LMM, n = 3", "LMM, n = 6",
"LMM, n = 10", "GEE, n = 3","GEE, n = 6", "GEE, n = 10"), x = "right",cex=.8,col=c(c("tomato",
"red", "darkred", "lightblue", "blue","darkblue")), lty=c(2,2,2,3,3,3)) } }

```

```
## Coverage probability under correct covariance model
```

```
for(k in 1:4){
  ## a = 0
  label_plot <- paste("Gamma =", delta[k], ", Tau = 0")
  plot(msamp, coverage_lmm[k,,1,1], type = "b", col="tomato", ylim = c(0,1),
       xlab = c("n_samp"), ylab=c("Coverage probability"), main = label_plot, lty = 2)
  lines(msamp, coverage_gee[k,,1,1], type = "b", col = "lightblue", lty = 3, lwd = 1)
  lines(msamp, coverage_lmm[k,,2,1], type = "b", col = "red", lty = 2, lwd = 1)
  lines(msamp, coverage_gee[k,,2,1], type = "b", col = "blue", lty = 3, lwd = 1)
  lines(msamp, coverage_lmm[k,,3,1], type = "b", col = "darkred", lty = 2, lwd = 1)
  lines(msamp, coverage_gee[k,,3,1], type = "b", col = "darkblue", lty = 3, lwd = 1)
  legend(c("LMM, n = 3", "LMM, n = 6", "LMM, n = 10", "GEE, n = 3", "GEE, n = 6", "GEE, n = 10"),
        x = "bottomright", cex=.8, col=c(c("tomato", "red", "darkred", "lightblue", "blue", "darkblue"),
        lty=c(2,2,2,3,3,3))
}
```

```
## Coverage probability under incorrect covariance model
```

```
for(k in 1:4){ ## a = 0 label_plot <- paste("Gamma =", delta[k], ", Tau = 1") plot(msamp, coverage_lmm[k,,1,2], type = "b", col="tomato", ylim = c(0,1), xlab = c("n_samp"), ylab=c("Coverage probability"), main = label_plot, lty = 2) lines(msamp, coverage_gee[k,,1,2], type = "b", col = "lightblue", lty = 3, lwd = 1) lines(msamp, coverage_lmm[k,,2,2], type = "b", col = "red", lty = 2, lwd = 1) lines(msamp, coverage_gee[k,,2,2], type = "b", col = "blue", lty = 3, lwd = 1) lines(msamp, coverage_lmm[k,,3,2], type = "b", col = "darkred", lty = 2, lwd = 1) lines(msamp, coverage_gee[k,,3,2], type = "b", col = "darkblue", lty = 3, lwd = 1) legend(c("LMM, n = 3", "LMM, n = 6", "LMM, n = 10", "GEE, n = 3", "GEE, n = 6", "GEE, n = 10"), x = "bottomright", cex=.8, col=c(c("tomato", "red", "darkred", "lightblue", "blue", "darkblue")), lty=c(2,2,2,3,3,3)) }
```