# Longitudinal Analysis to Assess the Impact of Method of Delivery on Postpartum Outcomes: The Ontario Mother and Infant Study (TOMIS) III

# Longitudinal Analysis to Assess the Impact of Method of Delivery on Postpartum Outcomes: The Ontario Mother and Infant Study (TOMIS) III

**By**

**YU QING BAI**

A Thesis

Submitted to the School of Graduate Studies

In Partial Fulfillment of the Requirements

For the Degree

Master of Science

McMaster University

MASTER OF SCIENCE (2011) McMaster University

(Statistics) Hamilton, Ontario

TITLE: Longitudinal Analysis to Assess the Impact of Method of Delivery on Postpartum Outcomes: The Ontario Mother and Infant Study (TOMIS) III

AUTHOR: Yu Qing Bai

M. A. Sc. (McMaster University)

B. Eng. (Qingdao University of Science & Technology)

SUPERVISOR: Dr. Lehana Thabane

NUMBER OF PAGES: xi; 156

# Abstract

Postpartum depression has become a major public health concern for women within a specific time period after delivery. Depression is possibly associated with some risk factors such as socioeconomic status, social support, maternal mental and physical health, and history of anxiety. TOMIS III, funded by the Canadian Institutes of Health Research, is a prospective cohort to study the associations between delivery method and health and health resource utilization.

Clinically, we investigated the associations between mode of delivery and outcome of postnatal depression, maternal and infant health, and we implied the risk predictors for outcomes by statistical methodology of marginal model with generalized estimating equations (GEE). Statistically, a variety of regression models, namely, generalized linear mixed effect model (GLMM), hierarchical generalized linear model (HGLM) and Bayesian hierarchical model were applied for this analysis and results were compared with GEEs. Some imputation strategies, namely, mean imputation, last observation carrying forward (LOCF), hot-deck imputation and multiple imputation were employed for handling missing values in this study.

Analysis results demonstrated that there was no statistically significant association between mode of delivery and postpartum depression [OR 0.99, 95% CI (0.73, 1.34)]. However, the development of postpartum depression was found to be associated with low

income, low mental and physical health functioning, lack of social support, the low number of unmet learning needs in hospital, and English or French spoken at home. Results were consistent for all regression models but GEE provided the best fit and an excellent discriminative ability. GEE models were constructed on different datasets imputed by mean, LOCF, hot-deck and multiple imputation, and LOCF was recommended to handle the missing data in this longitudinal study.

Analyses on the outcome of maternal health and infant health stated that method of delivery had a statistically significant influence on maternal health but no significant impact on infant health. Risks of maternal health problems were associated with cesarean delivery, good/fair/poor infant health, low maternal mental and physical health functioning, lack of care for maternal mental health, and good/fair/poor health before pregnancy. Risks of infant health problems were associated with good/fair/poor maternal health before pregnancy and after discharge, inadequate care or help for infant health, fair/poor community services after discharge, low maternal mental health functioning, non-English or non-French spoken at home, and mothers born outside of Canada.

# Acknowledgements

I would like to express my most sincere thanks to my supervisor, Dr. Lehana Thabane, for his expert guidance and encouragement provided towards completion of this project. His working style and enthusiasm were extremely helpful for me to start my career in health research area.

I would also like to put my thanks to my committee members, Dr. Roman Viveros-Aguilera and Dr. Gary Foster, for their valuable advices, and special thanks to Dr. Gary Foster for the discussions on the TOMIS III project. Many thanks to Ick Huh for his proofreading.

I finally would like to thank my wife, Li Hua Wang, and my sons, Luke and Danny Bai, for their support and encouragement for my study.

# Table of Contents

# List of Tables

# List of Figures

# List of Code

# Chapter 1 Introduction

## 1.1 Background

Usually, most vaginal deliveries are non-invasive and are associated with minimal potential harm or side effects for both a pregnant woman and her infant. In many instances, caesarean deliveries, which involve surgery, are considered to be life-saving procedures when a vaginal delivery may unduly risk the health of a woman and her infant[1]. However, rates of caesarean section (C-section) in many countries have shown an increasing trend over the last two decades[2]. For example, in Canada, the C-section rate has been steadily climbing from 17% in 1993 to 26% in 2006[3-5].

The worldwide increase in C-section rate has become an international public health concern. This is because there are maternal risks associated with this procedure including: infection[6], blood loss and hemorrhage[7], rehospitalization due to surgical complications[8]; reduced rate of establishment and ongoing breastfeeding[9]; compromised psychological well-being and increased rate of emotional trauma[10]. Other studies of medical risks of C-section on baby health included: fetal respiratory distress syndrome (RDS)[11-13], persistent pulmonary hypertension (PPH), and surgery-related fetal injuries such as lacerations[14]. After searching the PubMed database using specified terms of "caesarean section, vaginal delivery, breastfeeding, functional health", there is very little published research on

association between delivery methods and the outcomes of breastfeeding duration and functional health status.

Hospital costs associated with C-section are another important issue. Some studies have shown that C-section delivery has higher costs than vaginal delivery[3]. But these studies have focused mainly on increased hospital resources such as increased anesthesia, longer stays, medical supplies, nursing[15,16]. But the post-discharge costs of services used by women and infants have been given only very little attention[17].

## 1.2 The Ontario Mother and Infant Study (TOMIS) III

Funded by the Canadian Institutes of Health Research, the TOMIS III study was designed to address the association between delivery methods and maternal and infant health outcomes, service utilization, and cost of care in the first postpartum year. Over 2500 women were recruited from 11 hospitals across Ontario. The data were collected by self-report questionnaire (baseline measurements) in hospital and scheduled telephone interviews at 6 weeks, 6 months, and 12 months after discharge. The data have a longitudinal structure, that is, the measurements of same individuals are taken repeatedly through time. Therefore, the methods of longitudinal analyses are utilized in this thesis. Based on the results of TOMIS and TOMIS II, there is a potential attrition rate of 30% leading to a substantial amount of missing data in the dataset [17].

## 1.3 Objectives

The overall goal of this project is to determine the optimal modeling strategy to imply the relationship between delivery method and postpartum depression, maternal and infant

health by comparing results from different modeling approaches and missing data handling strategies.

The clinical objective of this thesis is to examine the relationship between delivery method and postpartum depression over time and to examine the relationship between delivery method and maternal health and infant health over time. To achieve this goal, various models such as generalized estimating equations (GEE)[18,19], generalized linear mixed effect models (GLMM)[20], hierarchical generalized linear models (HGLM)[20], and Bayesian hierarchical models[21,22] were applied.

The statistical objective of this thesis consists of two parts. Firstly, to compare different modeling methods of analyzing longitudinal data, including: GEE, GLMM, HGLM, and Bayesian hierarchical models. Secondly, to compare different missing data handling methods, namely, single imputations including mean imputation, last observation carried forward (LOCF), hot-deck imputation and multiple imputation[23-25].

## 1.4 Outline of Thesis

The framework of this thesis is briefly outlined as follows:

Chapter 1 presents a brief introduction and background of TOMIS III, clinical and statistical objectives, and outlines of the thesis.

An introductory overview of statistical methods for longitudinal and clustered data analysis is described in Chapter 2. Four statistical models used for analyzing the outcome of postpartum depression are introduced, and bootstrap variable selection methods and

GEE model for maternal and infant health analysis are discussed. The assessment of discriminative ability for GEE and comparisons of different missing data handling methods for repeated measurements are also discussed.

Comparisons of the results of all analytical methods for the outcomes of postpartum depression and maternal and infant health are presented in Chapter 3. A sensitivity analysis result from different prior distributions in Bayesian hierarchical modeling is presented. Results of assessment of discriminative ability for GEE and missing imputations on postpartum depression are discussed in Chapter 3.

In Chapter 4, analytical results investigating the relationship between postpartum depression and mode of delivery, the effect of different delivery method on maternal and infant health, sensitivity analysis for priors, model validations, and missing data imputations are presented and discussed.

Finally, we present the clinical conclusions and statistical inferences of this work in Chapter 5.

# Chapter 2 Statistical Methods

## 2.1 Reviews on Longitudinal and Clustered Data Analysis

As per the defining feature of longitudinal studies, the measurements of the same patients are taken repeatedly through time, thereby allowing the direct study of change over time. The primary goal of a longitudinal study is to characterize the change in response over time and factors that influence changes[26].

A distinctive feature of longitudinal data is that they are clustered. The clusters are composed of the repeated measurements obtained from a single patient at different occasions. Observations within a cluster typically exhibit positive correlation that must be counted when conducting analyses[26]. Alternatively, the measurements are conducted on patients nested within hospitals, nested within regions such that a multilevel data structure also can be considered. On the other hand, longitudinal data also have a temporal order, where the first measurement for a patient necessarily comes before the second measurement and so on. Indeed the longitudinal study is the only way of capturing the within-individual change over time by studying repeated measures on each individual[26].

To analyze longitudinal data, both marginal models using GEE[27, 28] and mixed effect models[29, 30] are most often used to determine the associations between predictors and longitudinal responses. For longitudinal data with multilevel structures, an HGLM [20] and

Bayesian hierarchical model[21, 22] can be implemented to capture the within-individual change over time.

The intra-cluster correlation coefficient (ICC) was introduced to measure the similarity of individuals within the same cluster. The ICC, denoted by $\rho$, can be interpreted as the proportion of the variability in the outcome due to variation between clusters of individuals.

The TOMIS III dataset has a typical longitudinal structure with multilevel features. The measurements were conducted at four time points, i.e., baseline measures in hospitals and follow-up measures at 6 weeks, 6 months and 12 months after discharge. Four models introduced above can be applied to capture the associations between outcomes and predictors.

## 2.2 Statistical Analyses for TOMIS III Study

The demographic characteristics of patients were summarized using descriptive statistics and expressed as mean (standard deviation) [SD] or median (minimum, maximum) for continuous variables and count (percent) for categorical variables.

Prior to the primary analysis, multicollinearity diagnostics should be conducted for each primary outcome and its corresponding independent variables. A regression model was used to detect the highly correlated factors and the tolerance and variance inflation factor (VIF) were calculated to detect possible multicollinearity.

The marginal models with GEEs were used for the primary analysis on the outcome of postpartum depression, maternal health and infant health. Various approaches such as GLMM, HGLM, and Bayesian hierarchical models were applied for implying the association between postpartum depression and mode of delivery. The fit statistics were compared to GEE.

For the analysis of maternal health and infant health, a bootstrap method was chosen to perform variable selections as it can lead to less bias and gives more stable variables[31, 32].

The ICC for each outcome variable was calculated and the design effect also was computed based on average cluster size using the following formula:

$$Design\ Effect = 1 + (m - 1)\rho \qquad \text{(Eq. 2.1)}$$

Where, $m$ is average cluster size and $\rho$ is ICC.

TOMIS III is a prospective cohort panel study with one year of follow-up. Some participants may drop out from this study or be lost from follow-up. Therefore, three single-imputation methods and multiple imputation method were employed for postpartum depression study to handle missing values. GEE was fitted to the imputed dataset and results were compared.

All classical models were conducted using SAS 9.1 (SAS Institute, Inc.) and Bayesian hierarchical models were fitted using WinBUGS 1.4 (Medical Research Council, UK). For all models the results were reported as estimates of coefficients (or odds ratios [OR] for binary outcomes), corresponding two-sided 95% confidence intervals and associated p-values.

The analysis procedures are summarized and described in Figure 2.1. All SAS programs and WinBUGS codes are presented in Appendix C.

### 2.2.1 Analysis on Postpartum Depression

#### 2.2.1.1 Generalized Estimating Equations

We assume that there are $N$ patients measured repeatedly through time and let $Y_{ij}$ denote the response of postpartum depression for the $i^{th}$ patient in the $j^{th}$ hospital. $Y_{ij}$ is a binary response variable with the values of 0 (denoting "no") and 1 (denoting "yes").

Each $Y_{ij}$ follows Bernoulli distribution and the mean is related to $\boldsymbol{X}$ by logit link function:

$$g(\mu_{ij}) = log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \boldsymbol{X}_{ij}^{'}\boldsymbol{\beta} \qquad \text{(Eq 2.2)}$$

Where,

$\mu_{ij}$: the mean of $Y_{ij}$, which is related to the covariates of $\boldsymbol{X}_{ij}^{'}$ by link function

$\boldsymbol{X}_{ij}$: a $p \times 1$ vector of covariates,

$\boldsymbol{\beta}$: a $p \times 1$ vector of unknown regression coefficients of $\boldsymbol{X}$, and

$g(\,.\,)$: logit link function as $Y_{ij}$ is binary.

To estimate $\boldsymbol{\beta}$, we need to solve the generalized estimating equations:

$$\sum_{i=1}^{N} \boldsymbol{D}_i^{'} \boldsymbol{V}_i^{-1} (\boldsymbol{Y}_i - \boldsymbol{\mu}_i) = 0 \qquad \text{(Eq 2.3)}$$

Where,

$\boldsymbol{D}_i$ : an $n_i \times p$ matrix with the elements of $\partial \boldsymbol{\mu_i} / \partial \boldsymbol{\beta}$,

$\boldsymbol{Y}_i$: a vector of responses measured at 6 weeks, 6 month and 12 month for the $i^{th}$ patient

$\boldsymbol{\mu}_i$: the mean of $\boldsymbol{Y}_i$, $\boldsymbol{\mu}_i = (\mu_{i1}, \mu_{i2}, \ldots, \mu_{in_i})^{'}$, which is function of $\boldsymbol{\beta}$ through logit link, and

$\boldsymbol{V}_i$: working correlation matrix, we used an exchangeable structure that assumes the same correlation between any two participants within a hospital.

We define $\boldsymbol{A}_i$ as an $n \times n$ diagonal matrix with $j^{th}$ element of $V(\mu_{ij})$, and $\boldsymbol{R}_i(\alpha)$ as an $n \times n$ working correlation matrix with unknown parameter $\alpha$, which is assumed to be same for all subjects. Hence, $\boldsymbol{V}_i$ can be decomposed as

$$\boldsymbol{V}_i(\alpha) = \emptyset \boldsymbol{A}_i^{\frac{1}{2}} \boldsymbol{R}_i(\alpha) \boldsymbol{A}_i^{\frac{1}{2}} \qquad \text{(Eq 2.4)}$$

Where, $\emptyset = 1$ for binary outcome of postpartum depression with logit link function.

Thus, Eq 2.3 is solvable and GEE estimator $\widehat{\boldsymbol{\beta}}$ is the solution of Eq 2.3 [26].

$\widehat{\boldsymbol{\beta}}$ can be solved out by SAS procedure of GENMOD with REPEATTED option, then *OR*s for covariates of $\boldsymbol{X}$ were computed using $\widehat{\boldsymbol{OR}} = exp(\widehat{\boldsymbol{\beta}})$. The results of estimates were reported using *OR*s with corresponding 95% CIs.

#### 2.2.1.2 Generalized Linear Mixed Effect Model

The GLMM can be simply considered as a straightforward extension of the generalized linear model, adding random effects to linear predictor and expressing the expected value of the response conditional on the random effects[33].

Similarly, we let $Y_{ij}$ denote binary response of postpartum depression for the $i^{th}$ patient in the $j^{th}$ hospital, taking value of 0 or 1. The link function can be:

$$g(\mu_{ij}) = log\left(\frac{\mu_{ij}}{1-\mu_{ij}}\right) = \boldsymbol{X}_{ij}^{'}\boldsymbol{\beta} + b_i \qquad \text{(Eq 2.5)}$$

Where, $\boldsymbol{X}_{ij}$: covariates of the $i^{th}$ patient in the $j^{th}$ hospital,

$\boldsymbol{\beta}$: regression coefficients of $\boldsymbol{X}_{ij}$,

$\mu_{ij}$: the mean of $Y_{ij}$, which is related to the covariates of $\boldsymbol{X}_{ij}^{'}$ by link function, and

$b_i$: random effect with assumption of multinormal distribution having mean zero and variance $\psi$, i.e., $b_i \sim N(0, \psi)$

The link function of $g(\mu_{ij})$ is as in generalized linear models. The conditional distribution of $y_{ij}|b_i$ belongs to exponential family. The variance of $y_{ij}|b_i$ is function of

$\mu_{ij}$ with a dispersion parameter $\emptyset = 1$ for Bernoulli distribution in our case. So we have conditional variance of $Y_{ij}$:

$$Var(Y_{ij}|b_i) = E(Y_{ij}|b_i)[1 - E(Y_{ij}|b_i)] \qquad \text{(Eq 2.6)}$$

To obtain maximum likelihood estimates, we need to maximize the marginal likelihood by:

$$L(\beta, \theta, y) = \int f(y|b)p(b)db \qquad \text{(Eq 2.7)}$$

Where, $\beta$ and $\theta$ are parameters of likelihood function, $y$ is response variable, $f(y|b)$ is conditional distribution of data, and $p(b)$ is distribution of random effects. However, estimation of generalized linear model by maximum likelihood (ML) method is not so straightforward when random effects are nonlinear form.

In our analysis, a technique of restricted pseudo-likelihood (RPL) estimation[34] was applied as RPL is based on residual likelihood, which can reduce the bias in covariance parameter estimates. This estimation method involving Taylor series first created pseudo data for each optimization. These data were then transferred to have zero mean in residual methods. The estimates of parameters of covariance were ML but the estimates of fixed effects were generalized least squares estimates[35].

This estimation procedure can be performed using GLIMMIX procedure with RANDOM option in SAS. The results were presented by ORs, corresponding 95% CIs and p values.

### 2.2.1.3 Hierarchical Generalized Linear Model

Multilevel data structure is presented when there are clusters existing in the data, especially for the data from health sciences since individuals can be grouped in so many different ways[26]. For TOMIS III, study outcomes were obtained from patients who were nested in hospitals. Also for a longitudinal study, the repeated measurements between time points are correlated. Thus, we considered a three-level structure for the analysis (shown in Figure 2.2 ).

In HGLM illustrations, the notations have some differences from previous discussions. Here, $Y_{ijk}$ denotes postpartum depression for the $i^{th}$patient measurement at the $j^{th}$ time point in the $k^{th}$ hospital with $Cov(y_{ijk}, y_{i'jk}) \neq 0$ and $Cov(y_{ijk}, y_{i'j'k}) \neq 0$. So the models to each level can be demonstrated as follows:

Level 1 (patient level)

$$Y_{ijk} = \pi_{0jk} + \sum_{1}^{p} x_{pijk}\pi_{pjk} + \varepsilon_{ijk} \qquad \text{(Eq 2.8)}$$

Where, $\pi_{pjk}$ (p = 0,1,... , p) : level-1 coefficients, and

$\varepsilon_{ijk}$ : independently and identically distributed random variables, $N(0,\sigma^2)$ .

In this model, we have included all predictors at patient level, $x_{pij}$, and also we included both fixed effect and random effect.

Level 2 (time level)

$$\begin{cases} \pi_{0jk} = \beta_{00k} + \beta_{01k} Time_j + r_{0jk} \\ \pi_{1jk} = \beta_{10k} + \beta_{11k} Time_j + r_{1jk} \\ \pi_{2jk} = \beta_{20k} + \beta_{21k} Time_j + r_{2jk} \\ \ldots\ldots \\ \pi_{pjk} = \beta_{p0k} + \beta_{p1k} Time_j + r_{pjk} \end{cases} \qquad \text{(Eq 2.9 )}$$

Where, $r_{pjk}$ : random effects at time level, $r_{pjk} \sim N(0, \sigma_p^2)$,

$Time_j$ :time of measurements, 1, 2 and 3 represent 6w, 6m and 12m respectively,

$\beta_{pqk}$ : Level-2 coefficients.

In level 2 models (Eq 2.9), the time effect was considered and the random effects follow normal distributions with mean zero.

Level 3 (hospital level)

$$\begin{cases} \beta_{00k} = \gamma_{000} + u_{00k} \\ \beta_{01k} = \gamma_{010} + u_{01k} \\ \beta_{02k} = \gamma_{020} + u_{02k} \\ \beta_{10k} = \gamma_{100} + u_{10k} \\ \beta_{11k} = \gamma_{110} + u_{11k} \\ \beta_{12k} = \gamma_{120} + u_{12k} \\ \ldots\ldots \\ \beta_{p0k} = \gamma_{p00} + u_{p0k} \\ \beta_{p1k} = \gamma_{p10} + u_{p1k} \\ \beta_{p2k} = \gamma_{p20} + u_{p2k} \end{cases} \qquad \text{(Eq 2.10)}$$

Where, $u_{pqk}$ :random effects of hospital and follow normal distribution with mean 0, and

$\gamma_{pjk}$ : intercept at hospital level.

Plugging Eq 2.10 and Eq 2.9 in Eq 2.8 and ignoring random effects on slopes, we obtained a combined form as:

$$Y_{ijk} = \left\{\gamma_{000} + \sum_{p=1}^{p} x_{pijk}\left(\gamma_{p00} + \gamma_{p10}Time_j\right) + \mu_{010}Time_j\right\} +$$

$$\left\{\left(\gamma_{01k}Time_j + \mu_{00k} + r_{0jk} + \varepsilon_{ijk}\right)\right\} \qquad \text{(Eq 2.11)}$$

The final model is expressed as the sum of two parts: a fixed part, which contains three fixed effects (for the intercept, for effects of patient-level factors and for the effect of time) and random part, which contains four random effects (for the intercept, time slope, hospital-level residual $\mu_{00k}$, time-level residual $r_{0jk}$, and within patient residual $\varepsilon_{ijk}$).

The SAS procedure of GLIMMIX with RANDOM option was used to fit HGLM shown in Eq 2.11. The results were presented by ORs, corresponding 95% CIs and p values.

#### 2.2.1.4 Bayesian Hierarchical Model

Compared to classical models, Bayesian approach treats unknown parameters, say $\boldsymbol{\theta}$, as random variables (that is different from classical models) while the observed data $\boldsymbol{y}$ are treated as fixed and known quantities (here is same as classical approach)[36, 37]. Our interest is the distribution of parameter $\boldsymbol{\theta}$ after having observed data $\boldsymbol{y}$.

Let $Y_{ij}$ denote the binary outcome of depression with values of 0 or 1 on the $i^{th}$ patient during $j^{th}$ observation. Here, $j = 1, 2, 3$, which are time points of 6 weeks, 6 months and 12 months, respectively. So we have the following hierarchical model considered:

$$Y_{ij}|\pi_{ij} \sim Bernoulli(\pi_{ij}) \qquad \text{(Eq 2.12)}$$

with the link function:

$$\eta_{ij} = logit(\pi_{ij}) = \beta_{0j} + \sum_{p=1}^{p} x_{pij}\beta_{pj} + b_i \qquad \text{(Eq 2.13)}$$

Where, $\beta_{pj}$: coefficients of parameters,

$b_i$: random effect, follows an independent and identical normal distribution with zero mean and $\sigma^2$, and

$p$: number of covariates involved in model.

The random effect of Bayesian hierarchical model was assumed to follow a normal distribution with a zero mean and an unknown variance $\sigma^2$, that is, $b_i \sim N(0, \sigma^2)$. The observed data $\boldsymbol{Y}_{ij}$ were treated as fixed and known quantities as discussed.

To perform Bayesian inference for our data, we need to include prior information about parameter $\sigma^2$. The inverse of this conditional error variance was taken into account by a Gamma distribution, $\sigma^{-2} \sim Gamma\ (a, b)$, We chose values of parameter $\sigma$ or hyperparameters $a$ and $b$ to characterize the prior distribution. Non-informative parameters were used for our analysis such that the amount of prior information like

researchers' pre belief or ex ante information included in this analysis was minimized or eliminated.

To minimize the influence of these pre-information on our observed data, a distribution of $uniform\,(0, 10)$ was chosen as a non-informative prior based on some researchers' experience[38-40]. One of Markov Chain Monte Carlo (MCMC) methods, namely, Gibbs sampling algorithm, was introduced to summarize posterior distributions. The convergence of MCMC was assessed by comparing MC errors and standard deviations and also both dynamic trace plots and quantile plots were checked to detect the convergence[41, 42].

To ensure the prior beliefs are not unduly affecting the results, sensitivity analyses were carried out. Some uniform priors like U(0, 5), U(0,15), U(0, 20), U(0, 25) and U(0, 50) and conjugate priors like Inverse Gamma(0.001, 0.001), (0.01, 0.01) and (0.1, 0.1) were considered and the results were compared by ORs along with 95% credible intervals.

All Bayesian analyses in this thesis were conducted on WinBUGS 1.4. The number of iterations for each run was set at 20,000 with burn-in number of 5,000. The seed was 0500485. All outputs are similar and the output for U(0,10) as an example is presented in Appendix B.

### 2.2.2 Analysis on Maternal and Infant Health

#### 2.2.2.1 Bootstrap Variable Selections

A different method of variable selections was used for maternal and infant health analysis. Compared to univariate methods, bootstrap selections can give us more stable variables with less bias, especially for variables with correlations. The following discussions are focused on maternal health analysis because infant health analysis has the same procedure.

The original dataset was randomly split into two subsets named derivation set and validation set with proportion of 1:2. All subsequent variable selections and model developments were carried out on derivation set and final model validation was performed on validation set. Prior to variable selections, we checked multicollinearity using previously discussed methods. For variables with collinearity, we kept one and eliminated the others.

**Variable selections**

First, we created 1000 subsets with the same size based on derivation dataset using method of repeated simple random sampling (SRS) with replacement. GEE model was then fitted to each subset. Those variables that were significantly associated with outcome of maternal health with a significant level of $p < 0.25$ were selected[43, 44]. The frequency of each variable presented in models was counted. Finally, we created a series of candidate models for predicting maternal health. Each fitted model contained variables with rate of present at least 80%, 50%, 40%, 36%, 30%, and 20%, respectively. The variable of Mode of Delivery was forced in each model. Final model was screened out using three goodness-of-fit indices, namely, marginal $R^2$, quasi-likelihood under the independence model criterion (QIC), and its sample version called QICu.

**Model Validation**

The discriminative ability of final model was assessed in the validation dataset using c-index[45], which is defined as the probability of concordance between predicted probability and outcome. C-index is identical to the area under a receiver operating characteristic (ROC) curve[43, 46, 47]. A value of 0.5 indicates no predictive discrimination and a value of 1.0 indicates a perfect validity of the model. The variables in the final model were selected for subsequent analyses.

#### 2.2.2.2 Generalized Estimating Equations

A marginal model with GEEs was employed to analyze the association of maternal health and delivery method. Similarly to the discussions on analyses of postpartum depression in Section 2.2.1, maternal health is also a binary outcome, taking value 0 (denoting "having no health problem") and 1 (denoting "having health problem"). A logit link function with exchangeable variance structure was used to fit GEE models. The results of estimates were reported using *OR*s with corresponding 95% CIs.

The analysis on infant health was conducted using the same procedures.

## 2.3 Model Validations

In this section, bootstrap sampling method was introduced to validate final GEE model for postpartum depression outcome. Bootstrap samples were generated using repeated SRS with replacement and each pseudo-data set had the same size to original. Each patient had an equal probability to be sampled with each repetition.

**Bootstrap validation of regression model**

Two hundred bootstrap samples were created and the final GEE model was then constructed repeatedly to these 200 samples. The distribution of parameters estimates and their corresponding SDs were examined using Shapiro-Wilk W tests. The mean values of parameters estimates from 200 samples were calculated and compared with the estimates derived from original set.

**Discriminative ability of final GEE model**

To assess the discriminative ability, we examined the ROC curve and computed area under curve (AUC) for the final model. The distribution of AUCs for GEE models to 200 bootstrap samples in last step was studied and the mean value of AUCs was calculated.

## 2.4 Missing Data

### 2.4.1 Missing Data in TOMIS III

In a longitudinal study in health science, researchers always suffer from the problem of trying to get study subjects to return for each time of follow-up. However, some participants still drop out of study for various reasons. Thus, missing data arise due to the lack of responses. In TOMIS III study, data were collected by face-to-face interview for baseline measurements and scheduled phone interviews for each longitudinal time point. The three most common reasons for missing data in TOMIS III are (1) patients who refused to participate or cannot be reached after discharge, (2) patients who cannot or refuse to answer certain questions, and (3) patients who gave “Don’t Know” answers. The

existence of missing data can lead to serious problems such as reduction of efficiency and biased or unreliable results[48].

How to handle missing data is a challenge for the analysis of longitudinal data. The goal of imputing missing data is not to predict missing values but to draw inference about population quantities[49]. The most appropriate method to use will depend on the nature of the missing data. Three types of missingness, namely, missing completely at random (MCAR), missing at random (MAR) and not missing at random (NMAR), were defined by Rubin[50]. We used the common assumption of MAR for missing values in our analysis so that the missing values can be predicted by other variables[51, 52]. Four approaches to handle missing values were applied in this study, they include: mean imputation, last observation carried forward (LOCF), hot-deck imputation and multiple imputation[23-25].

### 2.4.2 Missing Data Imputations

#### 2.4.2.1 Single Imputations

Three single imputation strategies including mean imputation, last observation carried forward (LOCF) and hot-deck imputation, were employed under the assumption of MAR in this study.

**Mean imputation**

There were four measurements (baseline measurement in hospital included) per patient in TOMIS III study. We replaced the missing values using the mean of non-missing values from data on the given patient. If a variable of one patient missed all of these four

measurements, the imputed result was still missing. For binary outcomes, we rounded to 1 if the mean value was equal to or greater than 0.5 and to 0 otherwise[53].

**Last observation carried forward**

LOCF is a commonly used way of imputing the missing data due to patient dropouts in a longitudinal study in health science[54]. The last observed non-missing value on a given patient is used to fill in the missing values in the subsequent time point. An additional assumption for LOCF imputation method is that the value remains constant along the response time for the imputed variables. If the value of the initial time point was missing, subsequent missing data points were not imputed.

**Hot-deck imputation**

In hot-deck imputation, missing values are replaced using values taken from matched respondents. This method uses data from matched respondents to provide imputed values for the records with missing values[55]. We first sorted and stratified the data by key covariate to determine a matched donor, then filled in the missing value using the value provided by donor.

Single imputation is easy to operationalize by filling in a missing value from an observed value. However, the drawback is that single imputation does not provide an estimate of uncertainty in the imputed value, such that the analysis based on single imputation is underestimated[50].

**2.4.2.2 Multiple Imputation**

Instead of filling in one single value for each missing value, multiple imputation is proposed to deal with each missing value using a set of plausible values that represent the uncertainty about the right value to impute[56]. Three steps are used to address the multiple imputation procedures: to generate imputed datasets, to analyze complete sets, and to combine results for inference. Similarly, multiple imputation is also under the assumption of MAR.

For a longitudinal study, multiple imputation cannot be applied directly due to the repeated measurements structure. We first need to rearrange all repeated measures for each patient to one row to do imputations. Then we turned the imputed dataset back to longitudinal format. Prior to imputation, we are required to detect the missing patterns to determine which imputation method fits for data[57].

GEE model was constructed on each complete dataset using standard procedures and all analytical results were combined for inference which can reflect within-imputation and between-imputation variability.

# Chapter 3 Results

## 3.1 Demographic Characteristics

A total of 2560 women were recruited from 11 hospitals across the province of Ontario to participate in this study. The baseline characteristics of participants are displayed in Table 3.0. Among these participants, 32.31% of them had a C-section delivery. About 70.81% were born in Canada and 81.94% participants spoke English or French at home. 85.08% of patients had an education at level of college or university. 89.73% of participants had a total income more than $20K. The proportion of first pregnancy was 41.90% (1071 in total).

## 3.2 Results on Postpartum Depression Analysis

### 3.2.1 Multicollinearity Diagnostics

Logistic regression detection for multicollinearity showed that variables of SSQBTOT (total social score), AFFECT_S (effective social support), and INSTR_S (instrumental social support) exhibit collinearity. Variables of HIST_DEPRESSION (history of depression) and ANYPRE_DEPRESSION (any previous depression) had VIF values of 32.43 and 34.40 respectively, which were greater than critical value of 10. Also checking tolerance, which measures the proportion of variance that is not explained by all other variables, we found out that the values were 0.03 for HIST_DEPRESSION and 0.03 for

ANYPRE_DEPRESSION (see Table 3.2). These results indicated a linear relationship between both variables. So we kept SSQBTOT and HIST_DEPRESSION variables and eliminated the others from our analyses.

### 3.2.2 Intraclass Correlation Coefficients

For outcome of postpartum depression, the ICC and 95% CI within hospitals were 0.01 (0.00, 0.04). Design effect along with 95% CI was then calculated as 3.59 (1.76, 9.51). The covariates exhibited slight correlations within cluster.

### 3.2.3 Results of GEE

From GEE modeling, mode of delivery (vaginal vs. C-section) was not significantly associated with postpartum depression, and odds ratio along with 95% CI and p value were 0.99 (0.73, 1.34) and $p = 0.9375$.

The results of main effect analysis using GEE indicated that seven predictors had significant effects on depression at level of $\alpha = 0.05$: Language spoken at home; Total income; Unmet learning needs in hospital; SF-12 physical component score; SF-12 mental component score; Total social support and bladder problems.

Patients speaking English or French at home were more likely to have depression than patients speaking other languages at home. The OR and 95% CI (p value) were 0.64 (0.44, 0.93) ($p = 0.0187$) for Language spoken at home (Foreign vs. Canada official languages). Patients with low total income (less than $20K) were more likely to have postpartum depression than those with total income more than $20K [1.99 (1.30, 3.04), ($p =$

$0.0015$) ]. The odds ratios of SF-12 physical component score [0.96 (0.94, 0.98) ($p < 0.0001$)], SF-12 mental component score [0.84 (0.82, 0.85) ($p < 0.0001$)], and Total social support [0.95 (0.93, 0.97) ($p < 0.0001$)] were associated with an increase in one point on score, that is, patients with lower SF-12 physical component score, SF-12 mental component score, or Total social support were more likely to have experiences of depression. Patient with bladder problems had a much high odds of having depression (OR: 1.57, $p = 0.006$) than those without.

### 3.2.4 Results of GLMM

From analytical results using GLMM, mode of delivery showed no statistically significant effect on postpartum depression: 1.00 (0.71, 1.40) ($p = 0.9814$). The ORs and 95% CIs (p value) were 0.64 (0.42, 0.97) ($p = 0.0355$), 1.92 (1.20, 3.09) ($p = 0.0066$), 0.91 (0.87, 0.96) ($p = 0.0008$), 0.74 (0.65, 0.84) ($p < 0.0001$), 0.24 (0.20, 0.28) ($p < 0.0001$), 0.68 (0.59, 0.78) ($p < 0.0001$), and 1.61(1.14, 2.27) ($p = 0.0073$) for Language spoken at home, Total income, the number of Unmet learning needs in hospital, SF-12 physical component score, SF-12 mental component score, the score of Total social support and Bladder problems, respectively.

### 3.2.5 Results of HGLM

In this three-level hierarchical model analysis, mode of delivery also demonstrated no statistically significant influence to depression with OR corresponding 95% CI and p value of 1.03 (0.75, 1.41) and $p = 0.8522$. Compared to GEE, the predictor of Language spoken at home on postpartum depression did not show a significant effect at level

of $\alpha = 0.05$: 0.71 (0.48, 1.05) ($p = 0.0824$). The other covariates demonstrated similar results as GEE: 1.68 (1.08, 2.62) ($p = 0.0203$) for Total income, 0.90 (0.86, 0.95) ($p < 0.0001$) for the number of Unmet learning needs in hospital, 0.95 (0.94, 0.97) ($p < 0.0001$) for SF-12 physical component score, 0.84 (0.82, 0.85) ($p < 0.0001$) for SF-12 mental component score, 0.95 (0.93, 0.97) ($p < 0.0001$) for the score of Total social support and 1.59 (1.15, 2.19) ($p = 0.0052$) for Bladder problems.

### 3.2.6 Results of Bayesian Hierarchical Model

Compared to classical modeling, Bayesian analysis gave similar results. Respectively, the ORs and 95% Credible Intervals were 0.63 (0.43, 0.92), 2.00 (1.32, 3.04), 0.76 (0.66, 0.89), 0.72 (0.64, 0.82), 0.22 (0.19, 0.26), 0.72 (0.64, 0.82), and 1.57(1.14, 2.14) for Language spoken at home, Total income, the number of Unmet learning needs in hospital, SF-12 physical component score, SF-12 mental component score, the score of Total social support and Bladder problems. The effect of mode of delivery was still not statistically significant.

### 3.2.7 Impacts of Priors for Bayesian Analysis

To perform Bayesian analysis on postpartum depression, we chose a non-informative prior of U(0, 10) based on previous research works. A non-informative prior should have no or minimum impact on estimates of uncertainty. To evaluate the influence of prior for the analysis results, we introduced different non-informative distributions for this sensitivity analysis, which included U(0, 5), U(0,15), U(0, 20), U(0, 25) U(0, 50), Inverse Gamma(0.001, 0.001), Inverse Gamma (0.01, 0.01) and Inverse Gamma(0.1, 0.1). The

ORs and 95% CIs corresponding to above priors were 0.98 (0.72, 1.33), 0.98 (0.72, 1.35), 0.98 (0.72, 1.34), 0.98 (0.71, 1.35), 0.98 (0.72, 1.34), 0.98 (0.72, 1.33), 0.98 (0.72, 1.34) and 0.98 (0.72, 1.35). The results are so consistent that we cannot even find difference for estimates under two decimals. The results of comparisons are shown in Figure 2.7 and Table 3.7 in Appendix.

## 3.3 Results on Maternal Health Analysis and Infant Health Analysis

### 3.3.1 Multicollinearity Diagnostics

For the outcome of maternal health, regression multicollinearity detection showed that predictors of Total social score, Effective social support, Confidant social support, and Instrumental social support had a linear combination. Meanwhile, predictor Breastfeeding initiated had collinearity with intercept. The tolerance and VIF were 0.01 and 77.11 for History of depression, and 0.01 and 79.46 for Any previous depression, indicating they were highly correlated. In subsequent analyses, we kept Total social score and History of depression in the model.

For infant health, the same linear combination of covariates of Total social score, Effective social support, Confidant social support, and Instrumental social support was detected. The predictors of History of depression and Any previous depression were highly correlated with tolerances of 0.01 and 0.01, and VIFs of 76.85 and 79.22. Breastfeeding initiated had collinearity with intercept for the outcome of infant health. We used Total social score and History of depression in the future analyses.

### 3.3.2 Bootstrapping for Variable Selections

One thousand bootstrap samples were generated from derivation dataset using MCMC methods and GEE models were then fitted on these samples. The frequency of each covariate that appeared in these 1000 models at significance level of $p < 0.25$ had been counted and shown in Table 3.10 for maternal health and Table 3.11 for infant health.

**Maternal Health**

For the outcome of maternal health, we categorized variables in terms of rates of frequency that one variable presented in 1000 models at $p < 0.25$. GEEs were fitted on each category, that is, variables in at least 80%, 40%, 36% and 20% of bootstrap models, and the predictor of Mode of delivery was forced in each model. The fit statistics of marginal $R^2$, QIC and QICu were calculated for each model and compared (Table 3.12) using criteria: for QIC and QICu, the smaller is the better, and for marginal $R^2$, the greater is the better. The results showed that variables with appearance rate > 30% had the greatest marginal $R^2$ (0.90) and smaller QIC (519.13) and QICu (519.14). Thus, we used the variables with appearance rate at least 30% as the final variables.

**Infant Health**

For the outcome of infant health, the categories were at least 80%, 50%, 40%, 30% and 20%. The predictor of Mode of delivery still was involved in each model. We applied the same methods and the results indicated that variables with appearance rate > 30% had the greatest marginal $R^2$ (0.85) and smallest QIC (324.55) and QICu (324.56) (see Table

3.13). Therefore, we used the variables with appearance rate greater than 30% as the final variables.

The validation of the model with final variables constructed on validation datasets showed that c-index (equivalent to AUC) and 95% confidence limit were 0.91 (0.88, 0.94) for the outcome of maternal health and 0.86 (0.78, 0.94) for the outcome of infant health. Both final models had an excellent internal validity[58-60].

### 3.3.3 Results of GEEs for Maternal Health and Infant Health Analyses

We constructed GEEs on original dataset (TOMIS III dataset) using variables selected from bootstrap approaches. GEE models were finalized by using backward method.

**Maternal Health**

Six variables exhibited significant influences on outcome of maternal health. Mode of delivery had a significant effect for maternal health and OR and 95% CI were 0.82 (0.68, 0.99) ($p = 0.0404$). ORs and 95% CIs along with p values for other predictors were 0.90 (0.89, 0.91) ($p < 0.0001$), 0.80 (0.79, 0.81) ($p < 0.0001$), 6.95 (5.70, 8.47) ($p < 0.0001$), 2.66 (1.61, 4.40), and 3.05 (2.25, 4.13) for SF-12 mental component score, SF-12 physical component score, Health before pregnancy (good/fair/poor vs. excellent/very good), Unable to get care for a maternal mental health problem (yes vs. no), and Baby's health (good/fair/poor vs. excellent/very good). The Predictors of Health before pregnancy and Baby's health gave the highest ORs.

**Infant Health**

The predictor of Mode of delivery [0.97 (0.75, 1.24), $p = 0.8004$] was not significantly associated with infant health. An excellent status of infant health was significantly associated with 7 predictor variables: Able to get care or help for baby's health [2.19 (1.55, 3.08), $p < 0.0001$]; Maternal healthy before pregnancy [1.58 (1.21, 2.06), $p = 0.0008$]; Low mental functioning score [0.98 (0.96, 0.99), $p = 0.0003$]; English or French spoken at home [2.02 (1.42, 2.87), $p < 0.0001$]; Excellent/good rating of services in the community after discharge [1.48 (1.12, 1.97), $p = 0.0066$]; Excellent/ good maternal health since delivery [3.00 (2.29, 3.93), $p < 0.0001$]; and Born in Canada [1.42 (1.02, 1.98), $p = 0.0367$]. ORs were highest for Excellent/ good maternal health since delivery and Able to get care or help for baby's health.

## 3.4 Model Validations

The distribution of parameters estimates using GEE models on 200 bootstrap samples was examined. We used the mean to describe the central tendency of the distribution of bootstrap estimates. The results from Shapiro-Wilk W tests for each parameter indicated that there was no evidence to reject the null hypothesis that parameter was normally distributed (Table 3.18)[61, 62] (all Shapiro-Wilk W test p values were greater than 0.05). The mean values of parameter estimates from bootstrap models were extremely similar to those derived from the original dataset for the outcome of depression (Table 3.19 and Figure 3.15).

ROC curve for testing discriminative ability of final GEE model for postpartum depression is shown in Figure 3.16. The area under ROC curve was computed as 0.92,

which demonstrated an excellent discriminative ability. We also tested each bootstrap GEE model and computed AUCs, and found out that the mean of AUCs was 0.92 with 95% CI of (0.9197, 0.9221). The distribution of AUCs is shown in Figure 3.17.

## 3.5 Missing Data Imputations

In TOMIS III study, the missing data were primarily due to dropouts. The missing proportions were less that 27% at 6-week measurement, but, it increased up to 29.5% at 6-month visit. The rate of attrition was closed to 50% at final measurement (see Table 3.20). The missing exhibited an arbitrary pattern (Figure 3.18).

The missing proportion was reduced to 12% after mean imputation. The estimates from imputed data compared to original data were quite similar. Mode of delivery showed no statistically significant association with depression [1.06 (0.82, 1.38), $p = 0.7100$] (Table 3.21 and 3.22). The imputation rate was up to 18% for LOCF approach and the estimates had a very minor change, but association between mode of delivery and depression was not statistically significant still [1.03 (0.79, 1.36), $p = 0.8134$] (Table 3.23 and 3.24). Results from hot-deck method showed a different imputation rate, that is, all missing values were filled in completely, due to imputation mechanisms (Table 3.24). The OR and 95% CI (p value) were 0.97 (0.76, 1.24) ($p = 0.8072$), which indicated that there were not any obvious changes compared to original estimates.

In the multiple imputation, five datasets were randomly imputed using Monte Carlo Markov Chain (MCMC) method. The combined analytical results showed that the

covariate of Language spoken at home was not statistically significant associated with depression [0.76 (0.55, 1.05), $p = 0.0956$]. The other estimates were consistent with ones estimated from original data. The influence of mode of delivery on depression was not statistically significant [1.02 (0.75, 1.38), $p = 0.8975$] (Table 3.27).

# Chapter 4 Discussion

To imply the associations between mode of delivery and primary outcome of postpartum depression, maternal health and infant health, a series of marginal models with GEEs were introduced to this analysis. A comparison of a variety of modeling approaches including GLMM, HGLM and Bayesian hierarchical model to the outcome of postpartum depression was conducted. Four missing data imputation strategies were carried out and compared on the primary outcome of postpartum depression. We have addressed all the results from different modeling and different imputation approaches in Chapter 3. The similarities and differences of these approaches are discussed in this section.

## 4.1 Modeling Comparisons in Postpartum Depression Analysis

The results from estimates of GEE indicated that covariate of mode of delivery had no statistically significant association at level of $\alpha = 0.05$ with depression [0.99 (0.73, 1.34)], which was very similar to the estimates from GLMM [1.00 (0.71, 1.40)], HGLM [1.03 (0.75, 1.41)], and Bayesian analysis [0.98 (0.71, 1.34)] (see Table 4.2 and Figure 3.19 for details). The other predictors demonstrated the same influences as shown in GEE on depression, except the covariate of Language spoken at home estimated from HGLM [0.71 (0.48, 1.05)], which indicated no statistically significant impact to depression with p value of 0.0824. Interestingly, for continuous covariates of MCS12, PCS12 and Total

social score, ORs estimated from GEE were much closer to the ones from HGLM, but were slightly different from estimates of GLMM and Bayesian.

A modified Akaike information criterion (AIC) for GEE, namely, quasi-likelihood information criterion (QIC), was used for regression model comparisons[63,64]. Compared to GLMM and HGLM, GEE provided the smallest AIC and BIC and the largest log-likelihood value (Table 4.2), and also demonstrated an excellent discriminative ability as discussed in Section 3.4. Therefore, GEE was considered to be a better approach than the others for TOMIS III data.

## 4.2 Comparisons of Imputation Approaches

ORs estimated by GEE model based on mean, LOCF and hot-deck imputed data were consistent. The predictor of Language spoken at home was not significantly associated with depression at level of $\alpha = 0.05$ based on multiple imputed data. Mode of delivery was not a statistically significant predictor of depression based on all imputed data analyses. Consistent results can be found in covariates of Total income, Unmet learning needs in hospital, PCS12, MCS12, Total social support and Bladder problems (see Table 4.4 for more details).

Compared to single imputation, benefits of multiple imputation have been discussed by many researchers[49,51,52,56, 65], however, we cannot see differences from Q-Q plots of residuals in this analysis (see details in Figure 4.1-4.4 ). Checking the values of log-likelihood and Deviance/DF, we found out that the hot-deck approach provided the smallest Deviance/DF and LOCF had a maximum likelihood value with relatively small

ratio of deviance/DF. Multiple imputation resulted a Deviance/DF value of 0.34 and a minimum log-likelihood value of -1309.32, which did not lead GEE to having a best performance (Table 4.3). Thus, LOCF had a better performance than the others.

## 4.3 Findings on Analyses of Maternal and Infant Health

ORs from GEE estimates for outcome of maternal health implied that mode of delivery was significantly associated with maternal health, and baby's health and maternal health before pregnancy had high influences on postpartum maternal health. Moreover, the care for maternal mental health problem stated a significant effect on mother's health. Generally speaking, women having low mental or physical functioning, excellent health status before pregnancy, easy to get care for mental health problem, or whose baby's health was in excellent status, were most likely to present an excellent postpartum health status. Compared to vaginal delivery, women experiencing C-section reported a higher risk of postpartum health problems.

Analyses on infant health suggested that method of delivery had no statistically significant association with infant health. However, baby's health was highly correlated to predictors of Maternal health after delivery and Unable to get care or help for baby's health. That a mother had an excellent health status after delivery or a baby can be easy to get care or help for the health problems were likely to have great benefit for infant health. The other factors with high impact on maternal health also included country of birth, health before pregnancy, community services after discharge and mental functioning.

## 4.4 Comparisons of Findings from Other studies

### 4.4.1 Postpartum Depression

Previous analyses[66] of TOMIS III stated that delivery mode had no significant impact on the development of postpartum depression at 6 weeks. High risk of postpartum depression was associated with low mental health functioning, low subjective social status, high number of unmet learning needs in hospital, young maternal age, maternal hospital readmission, non-initiation of breastfeeding, and maternal postpartum health.

A longitudinal study by Patel et al. (2005) revealed that there was no reason for women at risk of postnatal depression to be managed differently with regard to mode of delivery[67].

Seguin et al. (1999)[68] specifically studied women with low income within Canada and found out that the financial strain was an important factor to develop depression. A study conducted by the University of Iowa reported that low-income women in Iowa are much more likely to suffer from clinically significant postpartum depression[69].

O'Hara and Swain[70] conducted a mate-analysis and found out that there was a strong negative relationship between social support and postpartum depression. A study[71] on lack of social support demonstrated that social support was a strong risk factor for postpartum depressive symptoms.

Hullfish et al (2007)[72] performed a cross-sectional study on 100 patients at the University of Michigan Hospital and 46 patients at the University of Virginia Hospital and revealed that there was an association between urinary incontinence and postpartum depression.

This finding suggested that depressed patients had more symptoms and a greater impact on their lives from urge urinary incontinence.

In our analysis, we have already shown that risks of postpartum depression were associated with low income, low mental and physical functioning, lack of social support, and the low number of unmet learning needs in hospital. Our results provided evidence that mode of delivery had no significant influence to postpartum depression. For the predictor of Language spoken at home, it was a complicated issue as it was potentially related to patient culture, family structure, country of birth, and education and geographic[73,74]. Future work will be needed to reveal the relationship between depression and language spoken at home.

### 4.4.2 Maternal Health

A prospective cohort study by Wang et al (2010) [75]on 602 patients in Shanghai, China showed that women with caesarean section had a high risk of chronic abdominal pain compared to those having vaginal delivery, and rehospitalization of patients with planned caesarean was more likely than for these with planned vaginal delivery in the first one to two months after giving birth. The WHO global survey on maternal and perinatal health 2007-08 resulted in that risk of maternal mortality and morbidity increased for all types of C-sections[76].

Gjerdingen et al (1990) studied the relationship of women's postpartum health to social support, length of leave, and complications of childbirth and reported that maternal mental disorders and physical health had a reciprocal relationship and infant’s health was

associated with maternal health[77]. Sufficient and appropriate care for mental health had positive effects on maternal health and physical function[78,79].

These results were consistent with our findings where analyses implied that maternal health was associated with method of delivery, infant health, maternal mental and physical functions, care for maternal mental health, and health before pregnancy.

### 4.4.3 Infant Health

Glantz (2011)[80], a researcher at University of Rochester School of Medicine, reviewed data from 10 hospitals in a New York area and found out that there was no link between C-section and infant health. Results from a prospective cohort study in China by Wang et al. (2007)[81] indicated that the incidence of neonatal complications and infant morbidities at all measurement occasions did not differ significantly between vaginal and C-section delivery.

Social supports were proved to be significant benefit effects for infant health[82,83]. Maternal mental health in pregnancy and after delivery also had associations with infant health[84]. Researchers from University of Texas (USA) found out that significantly higher proportions of children in non-English-primary-language households were not in excellent/very good health compared to English-primary-language households[85].

Our analysis to infant health demonstrated a compatible result with previous research works discussed above but one more significant factor of country of birth. The results

showed that for mothers born outside Canada, their infants experienced a higher health risk than those whose mothers had been born in Canada.

## 4.5 Limitations and Future Work

**Sample size and missing values**

The designed sample size for TOMIS III was 3774 based on attrition rate of 30% and ICC of 0.018[66]. Thirty independent variables and one response variable were involved in primary analysis for outcome of postpartum depression. Our analysis using GEE model included only 37.5% [4250/(3774×3)] of the target sample size due to missing values. Similarly, the rates were 38.8% and 38.5% for outcome of maternal health and infant health, respectively. The missing rate for predictors of physical score and mental score were up to 50% at follow-up time of 12 months. Only half (50.7%) of participants completed 12-month interview with Edinburgh Postnatal Depression Scale (EPDS) data. These may decline the power of analysis.

**Variable selection method**

Univariate variable selection method was applied for postpartum depression analysis. However, univariate approaches are designed to test one feature at a time for their ability to discriminate a dependent variable such that it is not able to capture the correlations of variables. Bootstrap selection methods are recommended.

**Main effects modeling**

In this analysis, GEEs or other models just included the main effects for each outcome. In practice, some combined effects (i.e., interactions) should be considered in regression

models. For example, an interaction term of delivery of model and country of birth was reported to have a statistically significant influence on depression at 6 weeks[66]. Future research work is recommended for the analysis with interaction terms.

**Analyses regarding mode of delivery**

In this analysis, we only detected the association between method of delivery (C-section vs. vaginal delivery) and outcomes of depression, maternal and infant health. For the different types of cesareans like planned or emergent C-section and different types of vaginal deliveries like assisted vaginal birth or spontaneous vaginal birth, they were not involved in this analysis. Some researchers[67, 86, 87] have studied the associations between postpartum depression and different cesarean or different vaginal delivery methods.

# Chapter 5 Conclusions

Clinically, we have evaluated the association between mode of delivery and outcomes of postpartum depression, maternal health and infant health using longitudinal analysis methodologies. The predictors for clinical outcomes have been investigated using modeling strategies. Statistically, a variety of modeling approaches including classical regressions, i.e., GEE, GLMM and HGLM, and Bayesian models have been compared for analysis of postpartum depression. Four missing imputation strategies were applied for depression analysis and analytical results from complete data were compared. Both clinical and statistical conclusions have been described as follows:

**Clinical conclusions**

There was no statistically significant association detected between mode of delivery and postpartum depression. Total family income, the number of unmet learning needs in hospital, physical health functioning scores and mental health functioning scores were identified as high risk factor of postpartum depression. Lack of social supports increased the chance of development of postnatal depression. Patients with bladder issues were more likely to have risk of depression than those without. Mothers speaking English or French at their home had a higher possibility to get depression than others.

For outcome of maternal postnatal health, caesarean delivery showed a significant influence within one year after delivery. Low mental and physical health functioning and low support for maternal mental health increased the risk of developing maternal health problems. An excellent health status before pregnancy or a great shape of infant health was an outstanding benefit for maternal health.

Inferences from analysis of infant health indicated that mothers having great health before pregnancy or after delivery had a positive impact on infant health; for mothers born in Canada or mothers with English or French spoken at home, their babies had a better health status than those whose mothers were born outside Canada or non-English and non-French spoken at home. A good community service or a good care support for infant health after discharge had a positive effect on infant health. Low maternal mental health functioning resulted in a high risk of health problem for infants.

**Statistical inferences**

Fit statistics have demonstrated that GEE exhibited the best fit for depression analysis. TOMIS III is a longitudinal study with clustered and correlated data and GEE was proved to be suitable for this analysis. GEE model also provided an excellent discriminative ability for analysis of depression.

Multiple imputation approach did not show any advantages for the data. LOCF was the best choice for handling missing values from TOMIS III study. In practice, LOCF has been proved to have a good performance to handle missing data for longitudinal study due to patient dropouts[54].

# References

1 . Koroukian SM. Relative Risk of Postpartum Complications in the Ohio Medicaid Population: Vaginal versus Cesarean Delivery. Med Care Res Rev. 2004; 61(2):203-24.

2 . Mukherjee SN. Rising Cesarean Section Rate. J Obstet Gynecol India. 2006; 56(4):298-300.

3 . Giving Birth in Canada Providers of Maternity and Infant Care [internet]. 2004 [Cited 2011 Feb 17]. Available from: http://dsp-psd.pwgsc.gc.ca/Collection/H118-25-2004E.pdf.

4. Shamshad B. Factors Leading to Increased Cesarean Section Rate. Gomal J Med Sci. 2008; 6(1):1-5.

5. C-section Rate in Canada Continues Upward Trend [internet]. [Cited 2011 Feb 17]. Available from: http://www.canada.com/topics/bodyandhealth/story.html.

6. Henderson EJ, Love EJ. Incidence of Hospital-acquired Infections Associated with Cesarean Section. J Hosp Infect. 1995; 29:245-55.

7. van Ham MA, van Dongen PW, Mulder J. Maternal Consequences of Caesarean Section: A Retrospective Study of Intra-operative and Postoperative Maternal Complications of Caesarean Section During A 10-year Period. Eur J Obstet Gynecol Reprod Biol. 1997; 74:1-6.

8. Lydon-Rochelle M, Holt VL, Martin DP, et al. Association Between Method of Delivery and Maternal Rehospitalization. J Amer Med Assoc. 2000; 283(18):2411-16.

9. Lydon-Rochelle MT, Holt VL, Martin DP. Delivery Method and Self-reported Postpartum General Health Status Among Primiparous Women. Paediatr Perinat Epidemiol. 2001 Jul; 15(3):241-2.

10. Ryding EL, Wijma K, Wijma B. Experiences of Emergency Cesarean Section: A Phenomenological Study of 53 Women. Birth. 1998; 25(4):246-51.

11. Morrison JJ, Rennie JM, Milton PJ. Neonatal Respiratory Morbidity and Mode of Delivery at Term: Influence of Timing of Elective Caesarean Section. Br J Obstet Gynaecol. 1995; 102:101-6.

12. Hales KA, Morgan MA, Thurnau GR. Influence of Labor and Route of Delivery on the Frequency of Respiratory Morbidity in Term Neonates. Int J Gynaecol Obstet. 1993; 43(1):35-40.

13. Levine EM, Ghai V, Barton JJ, Strom CM. Mode of Delivery and Risk of Respiratory Diseases in Newborns. Obstet Gynecol. 2001; 97:439–42.

14. Smith J, Hernandez C, Wax J. Fetal Laceration Injury at Cesarean Delivery. Obstet Gynecol. 1997; 90:344-6.

15. Khan A, Zaman S. Costs of Vaginal Delivery and Caesarean Section at a Tertiary Level Public Hospital in Islamabad, Pakistan. BMC Pregnancy Childbirth. 2010 Jan 20; 10:2.

16. Allen V, O'Connell C, Farrell S, et al. Economic Implications of Method of Delivery. Obstet Gynecol. 2005; 193:192–7.

17. Sword W, Watt S, Krueger P, et al. The Ontario Mother and Infant Study (TOMIS) III: A multi-site cohort study of the impact of delivery method on health, service use, and costs of care in the first postpartum year. BMC Pregnancy and Childbirth. 2009; 9:16.

18. Liang KY, Zeger L. Longitudinal Data Analysis for Discrete and Continuous Outcomes. Biomatrics. 1986; 44:121-30.

19. Liang KY, Zeger L. Models for Longitudinal Data: A Generalized Estimating Equation Approach. Biomatrics. 1988; 44:1049-60.

20. Hardin JW, Hilbe JM. Generalized Linear Models and Extensions, 2nd Edition. Stata Press; 2001.

21. Senthamaraikannan K, Senthilkumar B, Ponnuraja C et al. A Bayesian Hierarchical Model for Longitudinal Data. Int J Curr Res. 2010, 10:12-20.

22. Daniels MJ, Hogan JW. Missing Data in Longitudinal Studies: Strategies for Bayesian Modeling and Sensitivity Analysis. Chapman &Hall/CRC. 2008.

23. Nakai M, Ke W. Review of the Methods for Handling Missing Data in Longitudinal Data Analysis. Int Journal of Math Analysis. 2011; 1(5):1-13.

24. Myers W. Handling Missing Data in Clinical Trials: an Overview. Drug Inf J. 2000; 34:525–33.

25. Kenward M, Carpenter J. Multiple Imputation: Current Perspectives. Stat Methods Med Res. 2007; 16:199–218.

26 . Fitzmaurice GM, Laird NM, Ware JH. Applied Longitudinal Analysis. John Wiley & Son, Inc. 2002.

27. Liang KY, Zeger L. Models for Longitudinal Data: A Generalized Estimating Equation Approach. Biomatrics. 1988; 44:1049-60.

28. Liang KY, Zeger L. Longitudinal Data Analysis for Discrete and Continuous Outcomes. Biomatrics. 1986; 44:121-30.

29. Breslow NE and Clayton DG. Approximate inference in generalized linear mixed models. Journal of the American Statistical Association, 1993;88: 9-25.

30. Laird NM and Ware JH. Random-Effects Models for Longitudinal Data. Biometrics. 1982; 38: 963-974.

31. Chen CH, and George SL. The bootstrap and identification of prognostic factors via Cox's proportional hazards regression model. Statistics in Medicine. 1985;4:39-46.

32. Sauerbrei W and Schumacher M. A Bootstrap Resampling Procedure for Model Building: Application to the Cox Regression Model. Statistics in Medicine, 1992;11:2093-109.

33. Molenberghs G and Verbeke G. Models for Discrete Longitudinal Data. Springer Series in Statistics (Part IV). 2005; 265-280.

34. Wolfinger R. and O'Connell M. Generalized Linear Mixed Models: A Pseudo-Likelihood Approach. J Stat Comput Simul. 1993;4, 233–43.

35. The GLIMMIX Procedure. SAS Support Documentation. SAS Institute, Inc.

36. Gill J. Bayesian Methods: A Social and Behavioral Sciences Approach. Boca Raton: Chapman & Hall/CRC. 2002.

37. Gelman A, Carlin JB, Stem HS et al. Bayesian Data Analysis. Chapman & Hall/CRC, second edition. 2004.

38. Spiegelhalter, DJ, Abrams KR, and Myles JP. Bayesian Approaches to Clinical Trials and Health-Care Evaluation, section 5.7.3. Chichester: Wiley. 2004.

39. Gelman A. Prior distributions for variance parameters in hierarchical models. Bayesian Analysis. 2006;3(1), 515-33.

40. Spiegelhalter DJ, Thomas A, Best NG et al. BUGS: Bayesian inference using Gibbs sampling. MRC Biostatistics Unit, Cambridge, England. 1994, 2003.

41. Cowles, M. K., and B. P. Carlin. Markov Chain Monte Carlo Diagnostics: A Comparative Review. Journal of the American Statistical Association. 1995;91:883–904.

42. Brooks, S. P., and G. O. Roberts. Assessing Convergence of Markov Chain Monte Carlo Algorithms. Statistics and Computing. 1998;8:319–335.

43. Austin PC and Tu JV. Bootstrap Methods for Developing Predictive Models. Am Stat. 2004;58(2), 131-7.

44. Hosmer DW and Lemeshow S. Applied Logistic Regression. New York : Wiley. 1989.

45. Harrell FE Jr. Regression Modelling Strategies: With Applications to Linear Models, Logistic Regression, and Survival analyses. New York: Springer-Verlag. 2001.

46. Harrell FE Jr, Lee KL, Mark DB. Tutorial in Biostatistics: Multivariable prognosis model: Issues in developing models, evaluating assumptions and adequacy, and measuring and reducing errors. Statistics in Medicine 1996, 15:361-87.

47. Pencina MJ and D'Agostino RB. Overall C as a measure of discrimination in survival analysis: model specific population value and confidence interval estimation. Statistics in Medicine 2004, 23:2109-23.

48. Kromrey JD and Hines CV. Nonrandomly missing data in multiple regression: An empirical comparison of common missing data treatments. Educ Psychol Meas. 1994; 54 (3), 573-93.

49. Schafer J. Missing Data in Longitudinal Studies: A Review. 2005 [Cited 2011 Aug 17]. Available from: http://www.stat.psu.edu/~jls/aaps_schafer.pdf.

50. Rubin DB, Inference and Missing data. Biometrika. 1976; 63:581–92.

51. Little RJA and Rubin DB. Statistical analysis with missing data. John Wiley & Sons, Inc. 1987.

52. Sterne JAC, White IR, Carlin JB et al. Multiple imputation for missing data in epidemiological and clinical research: potential and pitfalls BMJ 2009; 338:b2393.

53. Ma J, Akhtar-Danesh N, Dolovich L, Thabane L, and the CHAT investigators. Imputation strategies for missing binary outcomes in cluster randomized trials. BMC Medical Research Methodology 2011;11:18.

54. Hamer RM, and Simpson PM. Last Observation Carried Forward Versus Mixed Models in the Analysis of Psychiatric Clinical Trials. Am J Psychiatry. 2009;166:639-41.

55. Madow WG, Olkin I, Rubin DB. An overview of hot-deck procedures, in: Incomplete data in sample surveys. Academic Press, New York. 1983.

56. Rubin DB. Multiple Imputation for Nonresponse in Surveys. New York: John Wiley & Sons, Inc. 1987.

57. SAS Support Documentation. Multiple Imputation for Missing Data. SAS/STAT Software Enhancements. SAS Institute Inc., Cary, NC, USA.

58. O'Connor CM, Hasselblad V, Mehta RH, et al. Triage After Hospitalization With Advanced Heart Failure. J Am Coll Cardiol. 2010; 55:872-878.

59. Mazouni C, Delaloge S, Rimareix F, et al: Nomogram for risk of relapse after breast-conserving surgery in ductal carcinoma in situ. J Clin Oncol. 2011; 29:e44.

60. Mehta RH, Honeycutt E, Shaw LK, et al. Clinical correlates of long-term mortality after percutaneous interventions of saphenous vein grafts. American Heart Journal. 2006; 152(4), 801-6.

61. Park, HM. Univariate Analysis and Normality Test Using SAS, Stata, and SPSS. Working Paper. The University Information Technology Services (UITS) Center for Statistical and Mathematical Computing, Indiana University. 2008.

62. Shapiro SS, Wilk MB. An Analysis of Variance Test for Normality (Complete Samples). Biometrika. 1965; 52(3/4), 591-611.

63. Pan W. Akaike's Information Criterion in Generalized Estimating Equations. Biometrics. 2001;57, 120-5.

64. SAS/STAT® 9.2 User's Guide The GENMOD Procedure. SAS Institute Inc., Cary, NC, USA. 2008; 1965-70.

65. Schafer JL and Graham JW. Missing data: our view of the state of the art. Psychological Methods. 2002;7, 147-77.

66. Sword W, Landy CK, Thabane L et al. Is mode of delivery associated with postpartum depression at 6 weeks: a prospective cohort study. BJOG. 2011;118(8):966-77.

67. Patel R, Murphy D, Peters T. Operative delivery and postnatal depression: a cohort study. BMJ. 2005; 330(7496): 879.

68. Séguin L, Potvin L, St-Denis M et al. Depressive symptoms in the late postpartum among low socioeconomic status women. Birth. 1999;26(3):157-63.

69. Segre LS, O'Hara MW, Arndt S et al. The prevalence of postpartum depression: the relative significance of three social status indices. Soc Psychiatry Psychiatr Epidemiol. 2007;42(4):316-21.

70. O'Hara MW and Swain AM. Rates and risk of postpartum depression-a meta-analysis. Int Rve Psychiatr. 1996; 8, 37-54.

71. Forman DN, Videbech P, Hedegaard M, et al. Postpartum depression:identification of women at risk. Brit J Obstet Gynaec. 2000; 107, 1210-7.

72. Hullfish KL, Fenner DE, Sorser SA, et al. Postpartum depression, urge urinary incontinence, and overactive bladder syndrome: is there an association? Int Urogynecol J Pelvic Floor Dysfunct. 2007;18(10):1121-6.

73. Languages of Canada [internet]. [Cited 2011 Aug 17]. Available from: http://en. wikipedia.org /wiki/Languages_of_Canada.

74. Statistics Canada. Community belonging and self-reported health [internet]. The Daily. Statistics Canada. [Cited in August 20, 2011]. Availabe from: http://www.statcan.gc.ca/daily-quotidien/080416/dq080416c-eng.htm.

75. Wang BS, Zhou LF, Coulter D, et al. Effects of caesarean section on maternal health in low risk nulliparous women: a prospective matched cohort study in Shanghai, China. BMC Pregnancy and Childbirth, 2010, 10:78.

76. Lumbiganon P, Laopaiboon M, Gülmezoglu AM, et al; World Health Organization Global Survey on Maternal and Perinatal Health Research Group. Method of delivery and pregnancy outcomes in Asia: the WHO global survey on maternal and perinatal health 2007-08. Lancet. 2010; 375(9713):490-9.

77. Gjerdingen DK, Froberg DG, Fontaine P. A causal model describing the relationship of women's postpartum health to social support, length of leave, and complications of childbirth. Women Health. 1990;16(2):71-87.

78. Hanna BA, Edgecombe G, Jackson CA, Newman S. The importance of first-time parent groups for new parents. Nurs Health Sci 2002; 4(4): 209-14.

79. Bahadoran P, Azimi A, Valiyani M, et al. The relation between social support and postpartum physical health in mothers. IJNMR 2009; 14(1): 19-23

80 . Glantz JC. Rates of Labor Induction and Primary Cesarean Delivery Do Not Correlate with Rates of Adverse Neonatal Outcome in Level 1 Hospitals. Journal of Maternal-Fetal & Neonatal Medicine. 2011; 24(4): 636-42.

81 . Zhou LF, Liang H, Wang BS, et al. Effects of Cesarean Section on Infant Health in China: Matched Prospective Cohort Study. J Reproduction Contraception. 2007; 18(3): 221-30.

82. Shonkoff JP. Social support and the development of vulnerable children. Am J Public Health. 1984;74:310-312.

83. Pascoe JM, Earp JA. The effect of mothers' social support and life changes on the stimulation of their children in the home. Am J Public Health. 1984;74:358-360.

84. Punamäki RL, Repokari L, Vilska S, et al. Maternal mental health and medical predictors of infant developmental and health problems from pregnancy to one year: Does former infertility matter? Infant Behavior and Development. 2006; 29(2), 230-42.

85 Flores G, Tomany-Korman SC. The language spoken at home and disparities in medical and dental health, access to care, and use of services in US children. Pediatrics. 2008;121(6): 1703-14.

86. Van Son M, Gerda Verkerk G, van der Hart O, et al. Prenatal depression, mode of delivery and perinatal dissociation as predictors of postpartum posttraumatic stress: an empirical study. Clin Psychol Psychother. 2005; 12(4): 297-312.

87. Baskett TF, Allen VM, O‘Connell CM , et al. Predictors of respiratory depression at birth in the term infant. BJOG. 2006; 113(7): 769-74.

# Appendices

# Appendix A Tables

**Table 3.0 Characteristics of TOMIS III Participants**

| Characteristics | Vaginal delivery n (%) | C-section delivery n (%) | Total n (%) |
|---|---|---|---|
| Mothers' age | n=1733 | n=827 | n=2560 |
| Less than 25 | 292 (16.85) | 76 (9.19) | 368 (16.38) |
| Equal or greater than 25 | 1441 (83.15) | 751 (90.87) | 2192 (85.63) |
| Living Status | n=1721 | n=821 | n=2542 |
| With partner | 1600 (92.97) | 781 (95.13) | 2381 (93.67) |
| Alone | 121 (7.03) | 40 (4.87) | 161 (6.33) |
| Country birth | n=1722 | n=823 | n=2545 |
| Canada | 1221 (70.91) | 581 (70.60) | 1802 (70.81) |
| Others | 501 (29.09) | 242 (29.40) | 743 (29.19) |
| Language spoken at home | n=1728 | n=824 | n=2552 |
| English or French | 1417 (82.00) | 647 (81.80) | 2091 (81.94) |
| Others | 311 (18.00) | 150 (18.20) | 461 (18.06) |
| Highest level of education | n=1724 | n=823 | n=2547 |
| High school or less | 292 (16.94) | 88 (10.69) | 380 (14.92) |
| College, university | 1432 (83.06) | 735 (89.31) | 2167 (85.08) |
| Total income | n=1669 | n=804 | n=2473 |
| Less than 20K | 187 (11.20) | 67 (8.33) | 254 (10.27) |
| 20K or more | 1482 (88.80) | 737 (91.67) | 2219 (89.73) |
| First pregnancy | n=1729 | n=827 | n=2556 |
| Yes | 715 (41.35) | 356 (43.05) | 1071 (41.90) |
| No | 1014 (58.65) | 471 (56.95) | 1485 (58.10) |

**Table 3.1 Variable Descriptions and Code**

**Variables involved in primary analysis of postpartum depression**

| Variable | Description | Code |
|---|---|---|
| mom_age | Mothers age | 1: <25, 0: >=25 |
| pp6m | Was this your first pregnancy? | 0 = yes, 1 = no |
| pp25m | Ready to be discharged? | 0 = definitely or probably yes |
| | | 1 = don't know, definitely or probably not |
| pp26m | Healt related concerns | numeric, 0 - 12 |
| pp28m | Language spoken at home | 0 = English or French |
| | | 1 = others |
| pp30m | Country of birth | 0 = yes, 1 = no |
| pp31m | Marital status | 0 = married, common-law living with a partner |
| | | 1 = single, widowed separated/divorced |
| pp33m | Total income | 1 = <$20k, 0 = $20k or more |
| pp34m | Highest level of education | 1 = elementary school or less, some HS, completed HS |
| | | 0 = all other categories |
| pp35 | Social status: MacArthur SES Ladder | numeric, 0 - 12 |
| bh5m | Would you like to learn more about…? | numeric |
| bh6m | Baby's health | 1 = good, fair, poor 0 = excellent, very good |
| bh7m | Cannot tell when baby is sick | 0 = yes, 1 = no, don't know |

| Variable | Description | Code |
|---|---|---|
| wb10m | Maternal hospital readmission | 1 = yes, 0 = no |
| gh1m | Health before pregnancy | 0 = excellent, very good 1 = good, fair, poor |
| gh1_2m | Health since delivery | 0 = excellent, very good 1 = good, fair, poor |
| pcs12 | SF-12 physical component score | numeric, 13 - 72 |
| mcs12 | SF-12 mental component score | numeric, 14 - 68 |
| wb25m | Physical health problems | numeric, 0 - 20 |
| AFFECT_S | Affective Social support | numeric |
| CONFIDANT_S | Confidant social support | numeric |
| INSTR_S | Instrumental social support | numeric |
| SSQBTOT | Total social support | numeric |
| W6_BLADDER | Bladder problems | 1 = yes, 0 = no |
| SE92m | Unable to get help for maternal physical health problem | 1 = yes, 0 = no |
| SE94m | Unable to get care for a maternal mental health problem | 1 = yes, 0 = no |
| PP14m | Rating of community health services | 0 = excellent, good 1 = fair, poor, didn't use |
| HS3m | Rating of services in hospital during labour | 0 = excellent, good 1 = fair, poor |
| SE89m | Rating of services in the community after discharge | 0 = excellent, good 1 = fair, poor, didn't use |
| hist_depression | History of depression | 1 = yes, 0 = no |
| preg_depression | Depression in pregnancy | 1 = yes, 0 = no |
| anypre_depression | Any previous depression | 1 = yes, 0 = no |
| typevc | Delivery method | 1 = c-section 0 = vaginal |
| ppd_num | postpartum depression | 1 = yes 0 = no |

**Table 3.2 Tolerances and VIFs of Predictors for Postpartum Depression**

| Variable | Tolerance | VIF* |
|---|---|---|
| mom_age | 0.77 | 1.29 |
| pp24m | 0.92 | 1.08 |
| pp25m | 0.93 | 1.07 |
| pp26m | 0.87 | 1.15 |
| pp28m | 0.53 | 1.87 |
| pp30m | 0.54 | 1.84 |
| pp31m | 0.86 | 1.17 |
| pp33m | 0.76 | 1.32 |
| pp34m | 0.79 | 1.26 |
| pp35 | 0.97 | 1.03 |
| bh5m | 0.84 | 1.19 |
| bh6m | 0.91 | 1.10 |
| bh7m | 0.95 | 1.05 |
| bh9am | 0.91 | 1.10 |
| wb10m | 0.96 | 1.04 |
| gh1m | 0.79 | 1.27 |
| gh1_2m | 0.56 | 1.79 |
| PCS12 | 0.53 | 1.89 |
| MCS12 | 0.57 | 1.76 |
| wb25m | 0.67 | 1.49 |
| affect_s | 0.45 | 2.22 |
| confidant_s | 0.53 | 1.89 |
| instr_s | 0.58 | 1.74 |
| w6_bladder | 0.92 | 1.09 |
| se92m | 0.92 | 1.09 |
| se94m | 0.90 | 1.11 |
| pp14m | 0.93 | 1.07 |
| hs3m | 0.95 | 1.05 |
| se89m | 0.89 | 1.13 |
| **hist_depression** | **0.03** | **32.43** |
| Preg_depression | 0.64 | 1.56 |
| **anypre_depression** | **0.03** | **34.40** |
| TYPEVC | 0.89 | 1.12 |

*: variance inflation factor

**Table 3.3 Estimates and Odds Ratios from GEE Model**

| Factor | Description | Estimate | Standard Error | 95% CI | | Odds Ratio | 95% CI | | Pr > ChiSq |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Lower | Upper | | Lower | Upper | |
| pp28m | Language spoken at home | -0.4536 | 0.1929 | -0.8317 | -0.0755 | 0.64 | 0.44 | 0.93 | 0.0187 |
| pp33m | Total income | 0.6883 | 0.2168 | 0.2635 | 1.1132 | 1.99 | 1.30 | 3.04 | 0.0015 |
| bh5m | Unmet learning needs in hospital | -0.0893 | 0.0244 | -0.1371 | -0.0416 | 0.91 | 0.87 | 0.96 | 0.0002 |
| PCS12[a] | SF-12 physical component score | -0.0405 | 0.0087 | -0.0576 | -0.0234 | 0.96 | 0.94 | 0.98 | <0.0001 |
| MCS12 [a] | SF-12 mental component score | -0.1762 | 0.0089 | -0.1937 | -0.1587 | 0.84 | 0.82 | 0.85 | <0.0001 |
| ssqbtot [a] | Total social support | -0.0513 | 0.0095 | -0.0699 | -0.0327 | 0.95 | 0.93 | 0.97 | <0.0001 |
| W6_bladder | Bladder problems | 0.4494 | 0.1636 | 0.1288 | 0.77 | 1.57 | 1.14 | 2.16 | 0.0060 |
| TYPEVC | Delivery method | -0.0123 | 0.1573 | -0.3207 | 0.296 | 0.99 | 0.73 | 1.34 | 0.9375 |

[a]: ORs associated with an increase one point on score

**Table 3.4 Estimates and Odds Ratios from GLMM Model**

| Factor | Description | Estimate | Standard Error | 95% CI | | Odds Ratio | 95% CI | | P value |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Lower | Upper | | Lower | Upper | |
| pp28m | Language spoken at home | -0.4487 | 0.2133 | -0.8669 | -0.0304 | 0.64 | 0.42 | 0.97 | 0.0355 |
| pp33m | Total income | 0.6549 | 0.2408 | 0.1826 | 1.1271 | 1.92 | 1.20 | 3.09 | 0.0066 |
| bh5m | Willing to learn more | -0.0900 | 0.0268 | -0.1425 | -0.0376 | 0.91 | 0.87 | 0.96 | 0.0008 |
| PCS12 [a] | SF-12 physical component score | -0.3024 | 0.0093 | -0.4300 | -0.1748 | 0.74 | 0.65 | 0.84 | <0.0001 |
| MCS12 [a] | SF-12 mental component score | -1.4418 | 0.0095 | -1.5952 | -1.2883 | 0.24 | 0.20 | 0.28 | <0.0001 |
| ssqbtot [a] | Total social support: SSQBTOT | -0.3918 | 0.0104 | -0.5307 | -0.2529 | 0.68 | 0.59 | 0.78 | <0.0001 |
| w6_bladder | Bladder problems | 0.4743 | 0.1767 | 0.1278 | 0.8209 | 1.61 | 1.14 | 2.27 | 0.0073 |
| TYPEVC | Delivery method | -0.0041 | 0.1738 | -0.3449 | 0.3368 | 1.00 | 0.71 | 1.40 | 0.9814 |

[a]: ORs associated with an increase one point on score

**Table 3.5 Estimates and Odds Ratios from HGLM Model**

| Factor | Description | Estimate | Standard Error | 95% CI | | Odds ratio | 95% CI | | P value |
|---|---|---|---|---|---|---|---|---|---|
| | | | | Lower | Upper | | Lower | Upper | |
| pp28m | Language spoken at home | -0.3493 | 0.2010 | -0.7434 | 0.0448 | 0.71 | 0.48 | 1.05 | 0.0824 |
| pp33m | Total income | 0.5215 | 0.2247 | 0.0810 | 0.9620 | 1.68 | 1.08 | 2.62 | 0.0203 |
| bh5m | Willing to learn more | -0.1008 | 0.0253 | -0.1505 | -0.0512 | 0.90 | 0.86 | 0.95 | <0.0001 |
| PCS12 [a] | SF-12 physical component score | -0.0463 | 0.0090 | -0.0641 | -0.0286 | 0.95 | 0.94 | 0.97 | <0.0001 |
| MCS12 [a] | SF-12 mental component score | -0.1799 | 0.0092 | -0.1978 | -0.1619 | 0.84 | 0.82 | 0.85 | <0.0001 |
| ssqbtot [a] | Total social support | -0.0493 | 0.0097 | -0.0684 | -0.0301 | 0.95 | 0.93 | 0.97 | <0.0001 |
| w6_bladder | Bladder problems | 0.4609 | 0.1649 | 0.1375 | 0.7842 | 1.59 | 1.15 | 2.19 | 0.0052 |
| TYPEVC | Delivery method | 0.0300 | 0.1608 | -0.2853 | 0.3452 | 1.03 | 0.75 | 1.41 | 0.8522 |

[a]: ORs associated with an increase one point on score

**Table 3.6 Estimates and Odds Ratios from Bayesian Analysis**

| Factor | Description | Mean | 95% CI | | Odds ratio | 95% CI | |
|---|---|---|---|---|---|---|---|
| | | | Lower | Upper | | Lower | Upper |
| pp28m | Language spoken at home | -0.4590 | -0.8425 | -0.0870 | 0.63 | 0.43 | 0.92 |
| pp33m | Total income | 0.6917 | 0.2756 | 1.1120 | 2.00 | 1.32 | 3.04 |
| pcs12 [a] | SF-12 physical component score | -0.3262 | -0.4523 | -0.2012 | 0.72 | 0.64 | 0.82 |
| mcs12 [a] | SF-12 mental component score | -1.4940 | -1.6430 | -1.3500 | 0.22 | 0.19 | 0.26 |
| ssqbtot [a] | Total social support: SSQBTOT | -0.3221 | -0.4464 | -0.1988 | 0.72 | 0.64 | 0.82 |
| bh5m | Willing to learn more | -0.2691 | -0.4182 | -0.1193 | 0.76 | 0.66 | 0.89 |
| w6_bladder | Bladder problems | 0.4494 | 0.1289 | 0.7624 | 1.57 | 1.14 | 2.14 |
| Typevc | Delivery method | -0.0208 | -0.3415 | 0.2925 | 0.98 | 0.71 | 1.34 |

[a]: ORs associated with an increase one point on score

**Table 3.7 Sensitivity Analysis for Various Prior Distributions**

| Prior | | Outcome: postpartum depression<br>Factor: delivery methods | |
|---|---|---|---|
| **Type of prior** | **Prior distribution** | **Odds ratio** | **95% C.I.** |
| Non-informative | Uniform (0, 5) | 0.98 | (0.72, 1.33) |
| | Uniform (0, 10) | 0.98 | (0.71, 1.34) |
| | Uniform (0, 15) | 0.98 | (0.72, 1.35) |
| | Uniform (0, 20) | 0.98 | (0.72, 1.34) |
| | Uniform (0, 25) | 0.98 | (0.71, 1.35) |
| | Uniform (0, 50) | 0.98 | (0.72, 1.34) |
| Conjugate | Igamma (0.01, 0.01) | 0.98 | (0.72, 1.33) |
| | Igamma (0.1, 0.1) | 0.98 | (0.72, 1.34) |
| | Igamma (1, 1) | 0.98 | (0.72, 1.35) |

**Table 3.8 Tolerances and VIFs of Predictors for Maternal Health**

| Variable | Tolerance | VIF |
|---|---|---|
| mom_age | 0.65 | 1.54 |
| pp24m | 0.90 | 1.11 |
| pp25m | 0.85 | 1.18 |
| pp26m | 0.80 | 1.26 |
| pp28m | 0.56 | 1.78 |
| pp30m | 0.58 | 1.73 |
| pp31m | 0.85 | 1.17 |
| pp33m | 0.67 | 1.49 |
| pp34m | 0.67 | 1.50 |
| pp35 | 0.93 | 1.08 |
| bh5m | 0.75 | 1.33 |
| bh6m | 0.85 | 1.18 |
| bh7m | 0.93 | 1.07 |
| bh9bm | 0.75 | 1.33 |
| wb10m | 0.87 | 1.15 |
| gh1m | 0.81 | 1.23 |
| ppd_num | 0.69 | 1.45 |
| PCS12 | 0.59 | 1.71 |
| MCS12 | 0.49 | 2.05 |
| wb25m | 0.60 | 1.68 |
| affect_s | 0.41 | 2.42 |
| confidant_s | 0.54 | 1.84 |
| instr_s | 0.51 | 1.95 |
| w6_bladder | 0.83 | 1.20 |
| se92m | 0.79 | 1.26 |
| se94m | 0.83 | 1.21 |
| pp14m | 0.83 | 1.20 |
| hs3m | 0.86 | 1.16 |
| se89m | 0.79 | 1.26 |
| hs1m | 0.59 | 1.69 |
| hs2m | 0.85 | 1.18 |
| hs4m | 0.81 | 1.24 |
| wb11m | 0.83 | 1.21 |
| se90m | 0.81 | 1.23 |
| sx81m | 0.80 | 1.24 |
| **hist_depression** | **0.01** | **77.11** |
| Preg_depression | 0.75 | 1.33 |
| **anypre_depression** | **0.01** | **79.46** |
| TYPEVC | 0.61 | 1.63 |

**Table 3.9 Tolerances and VIFs of Predictors for Maternal Health**

| Variable | Tolerance | Inflation |
|---|---|---|
| mom_age | 0.65 | 1.53 |
| pp24m | 0.91 | 1.10 |
| pp25m | 0.85 | 1.17 |
| pp26m | 0.79 | 1.26 |
| pp28m | 0.57 | 1.77 |
| pp30m | 0.58 | 1.73 |
| pp31m | 0.85 | 1.17 |
| pp33m | 0.67 | 1.48 |
| pp34m | 0.67 | 1.50 |
| pp35 | 0.93 | 1.07 |
| bh5m | 0.75 | 1.33 |
| bh7m | 0.93 | 1.07 |
| bh9bm | 0.75 | 1.33 |
| wb10m | 0.87 | 1.15 |
| gh1m | 0.80 | 1.26 |
| gh1_2m | 0.55 | 1.83 |
| ppd_num | 0.69 | 1.45 |
| PCS12 | 0.44 | 2.28 |
| MCS12 | 0.46 | 2.18 |
| wb25m | 0.60 | 1.68 |
| affect_s | 0.41 | 2.43 |
| confidant_s | 0.54 | 1.84 |
| instr_s | 0.51 | 1.95 |
| w6_bladder | 0.83 | 1.20 |
| se92m | 0.79 | 1.27 |
| se94m | 0.81 | 1.23 |
| pp14m | 0.84 | 1.20 |
| hs3m | 0.87 | 1.15 |
| se89m | 0.79 | 1.26 |
| hs1m | 0.59 | 1.69 |
| hs2m | 0.85 | 1.17 |
| hs4m | 0.81 | 1.24 |
| wb11m | 0.83 | 1.20 |
| se90m | 0.83 | 1.21 |
| sx81m | 0.80 | 1.24 |
| **hist_depression** | **0.01** | **76.85** |
| Preg_depression | 0.75 | 1.33 |
| **anypre_depression** | **0.01** | **79.22** |
| TYPEVC | 0.61 | 1.63 |

**Table 3.10 Frequency of Candidate Variables for Outcome of Maternal Health**

| Variables | Count 1 ($p \leq 0.05$) | Count 2 ($0.05 < p \leq 0.15$) | Count 3 ($0.15 < p \leq 0.25$) | Total Present | Present Rate % | Selection |
|---|---|---|---|---|---|---|
| MCS12 | 998 | 2 | 0 | 1000 | 100.0 | >80% |
| PCS12 | 1000 | 0 | 0 | 1000 | 100.0 | |
| gh1m | 928 | 48 | 10 | 986 | 98.6 | |
| se94m | 790 | 129 | 31 | 950 | 95.0 | |
| pp26m | 502 | 220 | 80 | 802 | 80.2 | |
| bh5m | 499 | 194 | 91 | 784 | 78.4 | >50% |
| pp34m | 355 | 206 | 111 | 672 | 67.2 | |
| se92m | 358 | 182 | 119 | 659 | 65.9 | |
| se89m | 298 | 177 | 118 | 593 | 59.3 | |
| hs3m | 211 | 204 | 132 | 547 | 54.7 | |
| bh7m | 200 | 173 | 120 | 493 | 49.3 | >40% |
| sx81m | 168 | 184 | 128 | 480 | 48.0 | |
| wb10m | 193 | 142 | 103 | 438 | 43.8 | |
| Preg_depression | 154 | 162 | 115 | 431 | 43.1 | |
| bh6m | 185 | 147 | 98 | 430 | 43.0 | |
| hist_depression | 161 | 166 | 103 | 430 | 43.0 | |
| wb11m | 131 | 138 | 125 | 394 | 39.4 | >36% |
| pp30m | 146 | 131 | 110 | 387 | 38.7 | |
| pp31m | 132 | 150 | 105 | 387 | 38.7 | |
| pp28m | 145 | 130 | 111 | 386 | 38.6 | |
| ppd_num | 130 | 131 | 103 | 364 | 36.4 | |
| mom_age | 120 | 147 | 94 | 361 | 36.1 | |
| w6_bladder | 121 | 121 | 117 | 359 | 35.9 | >30% |
| hs1m | 106 | 134 | 110 | 350 | 35.0 | |
| bh9bm | 108 | 117 | 115 | 340 | 34.0 | |
| pp14m | 75 | 117 | 117 | 309 | 30.9 | |
| pp24m | 76 | 129 | 104 | 309 | 30.9 | |
| ssqbtot | 77 | 99 | 126 | 302 | 30.2 | |
| wb25m | 76 | 123 | 102 | 301 | 30.1 | |
| **TYPEVC** | **83** | **102** | **114** | **299** | **29.9** | >25% |
| pp35 | 39 | 125 | 122 | 286 | 28.6 | |
| pp33m | 67 | 107 | 107 | 281 | 28.1 | |
| hs2m | 72 | 112 | 96 | 280 | 28.0 | |
| pp25m | 75 | 107 | 92 | 274 | 27.4 | |
| se90m | 64 | 104 | 104 | 272 | 27.2 | |
| hs4m | 52 | 98 | 100 | 250 | 25.0 | |

**Table 3.11 Frequency of Candidate Variables for Outcome of Infant Health**

| Source | Count 1 ($p \leq 0.05$) | Count 2 ($0.05 < p \leq 0.15$) | Count 3 ($0.15 < p \leq 0.25$) | Total Present | Present Rate % | Selection |
|---|---|---|---|---|---|---|
| pp25m | 696 | 163 | 61 | 920 | 92.0 | >80% |
| gh1m | 738 | 120 | 44 | 902 | 90.2 | |
| se90m | 705 | 142 | 46 | 893 | 89.3 | |
| pp33m | 522 | 219 | 78 | 819 | 81.9 | |
| pp24m | 564 | 171 | 74 | 809 | 80.9 | |
| pp35 | 527 | 214 | 62 | 803 | 80.3 | |
| hs3m | 427 | 183 | 97 | 707 | 70.7 | >50% |
| MCS12 | 361 | 189 | 105 | 655 | 65.5 | |
| pp14m | 314 | 196 | 101 | 611 | 61.1 | |
| bh7m | 291 | 178 | 129 | 598 | 59.8 | |
| hs1m | 303 | 184 | 107 | 594 | 59.4 | |
| pp28m | 305 | 169 | 102 | 576 | 57.6 | |
| hs2m | 253 | 162 | 111 | 526 | 52.6 | |
| pp34m | 185 | 202 | 105 | 492 | 49.2 | >40% |
| hist_depression | 220 | 162 | 109 | 491 | 49.1 | |
| pp30m | 189 | 161 | 116 | 466 | 46.6 | |
| se89m | 198 | 141 | 109 | 448 | 44.8 | |
| gh1_2m | 185 | 168 | 92 | 445 | 44.5 | |
| bh5m | 189 | 150 | 97 | 436 | 43.6 | |
| pp26m | 181 | 142 | 110 | 433 | 43.3 | |
| wb11m | 177 | 151 | 102 | 430 | 43.0 | |
| mom_age | 174 | 143 | 100 | 417 | 41.7 | |
| pp31m | 149 | 152 | 114 | 415 | 41.5 | |
| ppd_num | 165 | 148 | 96 | 409 | 40.9 | |
| PCS12 | 112 | 155 | 134 | 401 | 40.1 | |
| **TYPEVC** | **125** | **138** | **113** | **376** | **37.6** | >30% |
| w6_bladder | 109 | 138 | 112 | 359 | 35.9 | |
| ssqbtot | 103 | 133 | 108 | 344 | 34.4 | |
| sx81m | 105 | 124 | 109 | 338 | 33.8 | |
| Preg_depression | 79 | 127 | 123 | 329 | 32.9 | |
| bh9bm | 72 | 124 | 121 | 317 | 31.7 | |
| wb10m | 59 | 127 | 116 | 302 | 30.2 | |
| wb25m | 75 | 124 | 94 | 293 | 29.3 | >20% |
| se94m | 91 | 95 | 97 | 283 | 28.3 | |
| hs4m | 61 | 96 | 92 | 249 | 24.9 | |
| se92m | 45 | 103 | 99 | 247 | 24.7 | |

**Table 3.12 Fit statistics of Bootstrap Model for Outcome of Maternal Health**

| Model | MarginalR2 | QIC | QICU |
|---|---|---|---|
| >80% | 0.52 | 2195.56 | 2195.30 |
| >50% | 0.53 | 2154.57 | 2154.00 |
| >40% | 0.70 | 1387.75 | 1386.86 |
| >36% | 0.71 | 1362.41 | 1361.61 |
| **>30%** | **0.90** | **519.13** | **519.14** |
| >0% | 0.90 | 515.19 | 515.20 |

**Table 3.13 Fit statistics of Bootstrap Model for Outcome of Infant Health**

| Model | MarginalR2 | QIC | QICU |
|---|---|---|---|
| 80% | 0.21 | 1458.81 | 1458.77 |
| 50% | 0.27 | 1339.68 | 1339.41 |
| 40% | 0.31 | 1278.70 | 1278.57 |
| **30%** | **0.85** | **324.55** | **324.56** |
| 20% | 0.85 | 332.13 | 332.16 |

**Table 3.14 AUC and 95% CI for Validation for Maternal Health**

| | ROC Area | 95% Confidence Limits | |
|---|---|---|---|
| **validation** | 0.9124 | 0.883 | 0.9417 |
| **intercept** | 0.5 | 0.5 | 0.5 |

**Table 3.15 AUC and 95% CI for Validation for Infant Health**

| | ROC Area | 95% Confidence Limits | |
|---|---|---|---|
| **validation** | 0.8592 | 0.7779 | 0.9405 |
| **intercept** | 0.5 | 0.5 | 0.5 |

**Table 3.16 Results of GEE for Maternal Health Analysis**

| Factor | Description | Estimate | 95% CI | | Odds Ratio | 95% CI | | Pr > ChiSq |
|---|---|---|---|---|---|---|---|---|
| | | | Lower | Upper | | Lower | Upper | |
| MCS12 | SF-12 mental component score | -0.1066 | -0.1181 | -0.0951 | 0.90 | 0.89 | 0.91 | <0.0001 |
| PCS12 | SF-12 physical component score | -0.2243 | -0.2412 | -0.2075 | 0.80 | 0.79 | 0.81 | <0.0001 |
| gh1m | Health before pregnancy | 1.9386 | 1.7403 | 2.1368 | 6.95 | 5.70 | 8.47 | <0.0001 |
| se94m | Unable to get care for a maternal mental health problem | 0.9793 | 0.4770 | 1.4816 | 2.66 | 1.61 | 4.40 | 0.0001 |
| bh6m | Baby's health | 1.1158 | 0.8122 | 1.4193 | 3.05 | 2.25 | 4.13 | <0.0001 |
| TYPEVC | Delivery methods | -0.1966 | -0.3846 | -0.0086 | 0.82 | 0.68 | 0.99 | 0.0404 |

**Table 3.17 Results of GEE for Infant Health Analysis**

| Factor | Description | Estimate | 95% CI | | Odds Ratio | 95% CI | | Pr > ChiSq |
|---|---|---|---|---|---|---|---|---|
| | | | Lower | Upper | | Lower | Upper | |
| se90m | Unable to get care or help for baby's health | 0.7818 | 0.4402 | 1.1234 | 2.19 | 1.55 | 3.08 | <0.0001 |
| gh1m | Health before pregnancy | 0.4568 | 0.1911 | 0.7225 | 1.58 | 1.21 | 2.06 | 0.0008 |
| MCS12 | SF-12 mental component score | -0.0238 | -0.0366 | -0.011 | 0.98 | 0.96 | 0.99 | 0.0003 |
| pp28m | Language spoken at home | 0.7014 | 0.3487 | 1.0541 | 2.02 | 1.42 | 2.87 | <0.0001 |
| se89m | Rating of services in the community after discharge | 0.3936 | 0.1094 | 0.6778 | 1.48 | 1.12 | 1.97 | 0.0066 |
| gh1_2m | Maternal health since delivery | 1.0997 | 0.83 | 1.3695 | 3.00 | 2.29 | 3.93 | <0.0001 |
| pp30m | Were you born in Canada? | 0.3521 | 0.0217 | 0.6824 | 1.42 | 1.02 | 1.98 | 0.0367 |
| TYPEVC | Delivery methods | -0.0323 | -0.2827 | 0.2181 | 0.97 | 0.75 | 1.24 | 0.8004 |

**Table 3.18 Results of Normality Test for Bootstrap Estimates**

| Parameter | Shapiro-Wilk W | P value |
|---|---|---|
| pp28m | 0.9946 | 0.6903 |
| pp33m | 0.9878 | 0.0844 |
| bh5m | 0.9932 | 0.4880 |
| MCS12 | 0.9886 | 0.1115 |
| PCS12 | 0.9907 | 0.2234 |
| ssqbtot | 0.9911 | 0.2582 |
| w6_bladder | 0.9907 | 0.2236 |
| TYPEVC | 0.9889 | 0.1221 |

**Table 3.19 Comparison of GEE Estimates on Original and Bootstrap Data**

| Parameter | Original Data | | | | Bootstrap Data | | | |
|---|---|---|---|---|---|---|---|---|
| | Estimate | Standard Deviation | 95% C.L. | | Estimate | Standard Deviation | 95% C.L. | |
| | | | Lower | Upper | | | Lower | Upper |
| pp28m | -0.45 | 0.19 | -0.83 | -0.08 | -0.46 | -0.46 | -0.84 | -0.08 |
| pp33m | 0.69 | 0.22 | 0.26 | 1.11 | 0.67 | 0.67 | 0.23 | 1.09 |
| bh5m | -0.09 | 0.02 | -0.14 | -0.04 | -0.09 | -0.09 | -0.14 | -0.04 |
| MCS12 | -0.18 | 0.01 | -0.19 | -0.16 | -0.18 | -0.18 | -0.20 | -0.16 |
| PCS12 | -0.04 | 0.01 | -0.06 | -0.02 | -0.04 | -0.04 | -0.06 | -0.02 |
| ssqbtot | -0.05 | 0.01 | -0.07 | -0.03 | -0.05 | -0.05 | -0.07 | -0.03 |
| w6_bladder | 0.45 | 0.16 | 0.13 | 0.77 | 0.47 | 0.47 | 0.15 | 0.79 |
| TYPEVC | -0.01 | 0.16 | -0.32 | 0.30 | -0.01 | -0.01 | -0.32 | 0.30 |

**Table 3.20 Summary of Missing Data in TOMISIII Study**

| Timepoint | Total Observation | Label | No. of Missing | No. of Observed | Missing Percentage |
|---|---|---|---|---|---|
| Baseline | 2560 | Language spoken at home | 8 | 2552 | 0.31 |
| | | Total income | 87 | 2473 | 3.40 |
| | | Unmet learning needs in | 663 | 1897 | 25.90 |
| | | Delivery method | 0 | 2560 | 0 |
| 1 | 2560 | Postpartum depression | 672 | 1888 | 26.25 |
| | | SF-12 physical component score | 684 | 1876 | 26.72 |
| | | SF-12 mental component score | 684 | 1876 | 26.72 |
| | | Total social support | 670 | 1890 | 26.17 |
| | | Bladder problems | 677 | 1883 | 26.45 |
| 2 | 2560 | Postpartum depression | 751 | 1809 | 29.34 |
| | | SF-12 physical component score | 755 | 1805 | 29.49 |
| | | SF-12 mental component score | 755 | 1805 | 29.49 |
| | | Total social support | 745 | 1815 | 29.10 |
| | | Bladder problems | 752 | 1808 | 29.38 |
| 3 | 2560 | Postpartum depression | 1263 | 1297 | 49.34 |
| | | SF-12 physical component score | 1269 | 1291 | 49.57 |
| | | SF-12 mental component score | 1269 | 1291 | 49.57 |
| | | Total social support | 1263 | 1297 | 49.34 |
| | | Bladder problems | 1264 | 1296 | 49.38 |
| Total | 7680 | Postpartum depression | 2686 | 4994 | 34.97 |
| | | SF-12 physical component score | 2708 | 4972 | 35.26 |
| | | SF-12 mental component score | 2708 | 4972 | 35.26 |
| | | Total social support | 2678 | 5002 | 34.87 |
| | | Bladder problems | 2693 | 4987 | 35.07 |

**Table 3.21 Comparison of Mean Imputed Data and Original Data**

| Factor | Description | Original Data | | | Mean Imputed Data | | |
|---|---|---|---|---|---|---|---|
| | | N. Obs'd | N. Miss | P. Miss | N. Obs'd | N. Miss | P. Miss |
| ppd_num | Postpartum depression | 4994 | 2686 | 34.97 | 6765 | 915 | 11.91 |
| pp28m | Language spoken at home | 7656 | 24 | 0.31 | 7668 | 12 | 0.16 |
| pp33m | Total income | 7419 | 261 | 3.40 | 7419 | 261 | 3.40 |
| bh5m | Unmet learning needs in hospital | 5691 | 1989 | 25.90 | 5691 | 1989 | 25.90 |
| PCS12 | SF-12 physical component score | 4972 | 2708 | 35.26 | 6765 | 915 | 11.91 |
| MCS12 | SF-12 mental component score | 4972 | 2708 | 35.26 | 6765 | 915 | 11.91 |
| ssqbtot | Total social support | 5002 | 2678 | 34.87 | 6768 | 912 | 11.88 |
| w6_bladder | Bladder problems | 4987 | 2693 | 35.07 | 6765 | 915 | 11.91 |
| TYPEVC | Delivery method | 7680 | 0 | 0 | 7680 | 0 | 0 |

**Table 3.22 Comparison of GEE Estimates of Mean Imputed and Original Data**

| Factor | Description | Original Data | | | | Mean Imputed Data | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Odds Ratio | 95% CI | | P Value | Odds Ratio | 95% CI | | P Value |
| | | | Lower | Upper | | | Lower | Upper | |
| pp28m | Language spoken at home | 0.64 | 0.44 | 0.93 | 0.0187 | 0.67 | 0.49 | 0.91 | <0.0001 |
| pp33m | Total income | 1.99 | 1.30 | 3.04 | 0.0015 | 1.67 | 1.18 | 2.37 | 0.0118 |
| bh5m | Unmet learning needs in hospital | 0.91 | 0.87 | 0.96 | 0.0002 | 0.92 | 0.88 | 0.95 | 0.0042 |
| PCS12 | SF-12 physical component score | 0.96 | 0.94 | 0.98 | <0.0001 | 0.95 | 0.94 | 0.97 | <0.0001 |
| MCS12 | SF-12 mental component score | 0.84 | 0.82 | 0.85 | <0.0001 | 0.84 | 0.82 | 0.85 | <0.0001 |
| ssqbtot | Total social support | 0.95 | 0.93 | 0.97 | <0.0001 | 0.95 | 0.94 | 0.97 | <0.0001 |
| w6_bladder | Bladder problems | 1.57 | 1.14 | 2.16 | 0.006 | 1.45 | 1.11 | 1.91 | <0.0001 |
| TYPEVC | Delivery method | 0.99 | 0.73 | 1.34 | 0.9375 | 1.06 | 0.82 | 1.38 | 0.71 |

**Table 3.23 Comparison of LOCF Imputed Data and Original Data**

| Factor | Description | Original Data | | | LOCF Imputed Data | | |
|---|---|---|---|---|---|---|---|
| | | N. Obs'd | N. Miss | P. Miss | N. Obs'd | N. Miss | P. Miss |
| ppd_num | Postpartum depression | 4994 | 2686 | 34.97 | 6337 | 1343 | 17.49 |
| pp28m | Language spoken at home | 7656 | 24 | 0.31 | 7656 | 24 | 0.31 |
| pp33m | Total income | 7419 | 261 | 3.40 | 7419 | 261 | 3.40 |
| bh5m | Unmet learning needs in hospital | 5691 | 1989 | 25.90 | 5691 | 1989 | 25.90 |
| PCS12 | SF-12 physical component score | 4972 | 2708 | 35.26 | 6322 | 1358 | 17.68 |
| MCS12 | SF-12 mental component score | 4972 | 2708 | 35.26 | 6322 | 1358 | 17.68 |
| ssqbtot | Total social support | 5002 | 2678 | 34.87 | 6340 | 1340 | 17.45 |
| w6_bladder | Bladder problems | 4987 | 2693 | 35.07 | 6331 | 1349 | 17.57 |
| TYPEVC | Delivery method | 7680 | 0 | 0 | 7680 | 0 | 0 |

**Table 3.24 Comparison of GEE Estimates from LOCF Imputed Data and Original Data**

| Factor | Description | Original Data | | | | LOCF Imputed Data | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Odds Ratio | 95% CI | | Pr > ChiSq | Odds Ratio | 95% CI | | Pr > ChiSq |
| | | | Lower | Upper | | | Lower | Upper | |
| pp28m | Language spoken at home | 0.64 | 0.44 | 0.93 | 0.0187 | 0.60 | 0.43 | 0.83 | 0.0022 |
| pp33m | Total income | 1.99 | 1.30 | 3.04 | 0.0015 | 1.52 | 1.05 | 2.19 | 0.0271 |
| bh5m | Unmet learning needs in hospital | 0.91 | 0.87 | 0.96 | 0.0002 | 0.91 | 0.88 | 0.95 | <0.0001 |
| PCS12 | SF-12 physical component score | 0.96 | 0.94 | 0.98 | <0.0001 | 0.95 | 0.94 | 0.97 | <0.0001 |
| MCS12 | SF-12 mental component score | 0.84 | 0.82 | 0.85 | <0.0001 | 0.84 | 0.82 | 0.85 | <0.0001 |
| ssqbtot | Total social support | 0.95 | 0.93 | 0.97 | <0.0001 | 0.95 | 0.93 | 0.96 | <0.0001 |
| w6_bladder | Bladder problems | 1.57 | 1.14 | 2.16 | 0.006 | 1.39 | 1.05 | 1.85 | 0.0235 |
| TYPEVC | Delivery method | 0.99 | 0.73 | 1.34 | 0.9375 | 1.03 | 0.79 | 1.36 | 0.8134 |

**Table 3.25 Comparison of Hot-Deck Imputed Data and Original Data**

| Factor | Description | Original Data | | | Hot-Deck Imputed Data | | |
|---|---|---|---|---|---|---|---|
| | | N. Obs'd | N. Miss | P. Miss | N. Obs'd | N. Miss | P. Miss |
| ppd_num | Postpartum depression | 4994 | 2686 | 34.97 | 7680 | 0 | 0 |
| pp28m | Language spoken at home | 7656 | 24 | 0.31 | 7680 | 0 | 0 |
| pp33m | Total income | 7419 | 261 | 3.40 | 7680 | 0 | 0 |
| bh5m | Unmet learning needs in hospital | 5691 | 1989 | 25.90 | 7680 | 0 | 0 |
| PCS12 | SF-12 physical component score | 4972 | 2708 | 35.26 | 7680 | 0 | 0 |
| MCS12 | SF-12 mental component score | 4972 | 2708 | 35.26 | 7680 | 0 | 0 |
| ssqbtot | Total social support | 5002 | 2678 | 34.87 | 7680 | 0 | 0 |
| w6_bladder | Bladder problems | 4987 | 2693 | 35.07 | 7680 | 0 | 0 |
| TYPEVC | Delivery method | 7680 | 0 | 0 | 0 | 0 | 0 |

**Table 3.26 Comparison of GEE Estimates for Hot-Deck Imputed and Original Data**

| Factor | Description | Original Data | | | | Hot-Deck Imputed Data | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Odds Ratio | 95% CI | | P Value | Odds Ratio | 95% CI | | P Value |
| | | | Lower | Upper | | | Lower | Upper | |
| pp28m | Language spoken at home | 0.64 | 0.44 | 0.93 | 0.0187 | 0.66 | 0.49 | 0.87 | 0.0038 |
| pp33m | Total income | 1.99 | 1.30 | 3.04 | 0.0015 | 1.91 | 1.40 | 2.59 | <0.0001 |
| bh5m | Unmet learning needs in hospital | 0.91 | 0.87 | 0.96 | 0.0002 | 0.95 | 0.91 | 0.98 | 0.0027 |
| PCS12 | SF-12 physical component score | 0.96 | 0.94 | 0.98 | <0.0001 | 0.96 | 0.94 | 0.97 | <0.0001 |
| MCS12 | SF-12 mental component score | 0.84 | 0.82 | 0.85 | <0.0001 | 0.85 | 0.83 | 0.86 | <0.0001 |
| ssqbtot | Total social support | 0.95 | 0.93 | 0.97 | <0.0001 | 0.94 | 0.93 | 0.96 | <0.0001 |
| w6_bladder | Bladder problems | 1.57 | 1.14 | 2.16 | 0.006 | 1.47 | 1.15 | 1.89 | 0.0024 |
| TYPEVC | Delivery method | 0.99 | 0.73 | 1.34 | 0.9375 | 0.97 | 0.76 | 1.24 | 0.8072 |

**Table 3.27 Comparison of GEE Estimates from Multiple Imputed Data and Original Data**

| Factor | Description | Original Data | | | | Multiple Imputed Data | | | |
|---|---|---|---|---|---|---|---|---|---|
| | | Odds Ratio | 95% CI | | P Value | Odds Ratio | 95% CI | | P Value |
| | | | Lower | Upper | | | Lower | Upper | |
| pp28m | Language spoken at home | 0.64 | 0.44 | 0.93 | 0.0187 | 0.76 | 0.55 | 1.05 | 0.0956 |
| pp33m | Total income | 1.99 | 1.30 | 3.04 | 0.0015 | 1.66 | 1.04 | 2.64 | 0.0345 |
| bh5m | Unmet learning needs in hospital | 0.91 | 0.87 | 0.96 | 0.0002 | 0.95 | 0.91 | 0.99 | 0.0112 |
| PCS12 | SF-12 physical component score | 0.96 | 0.94 | 0.98 | <0.0001 | 0.96 | 0.94 | 0.98 | <0.0001 |
| MCS12 | SF-12 mental component score | 0.84 | 0.82 | 0.85 | <0.0001 | 0.85 | 0.84 | 0.87 | <0.0001 |
| ssqbtot | Total social support | 0.95 | 0.93 | 0.97 | <0.0001 | 0.96 | 0.94 | 0.98 | 0.0003 |
| w6_bladder | Bladder problems | 1.57 | 1.14 | 2.16 | 0.0060 | 1.43 | 1.07 | 1.91 | 0.0176 |
| TYPEVC | Delivery method | 0.99 | 0.73 | 1.34 | 0.9375 | 1.02 | 0.75 | 1.38 | 0.8975 |

**Table 4.1 Summary of Estimates of Variables from Different Models**

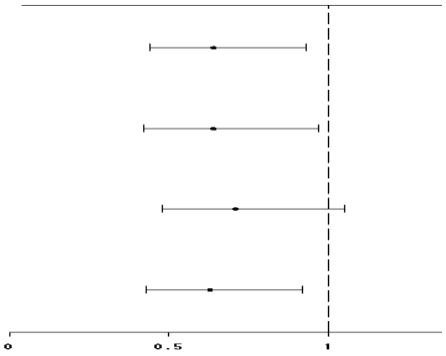
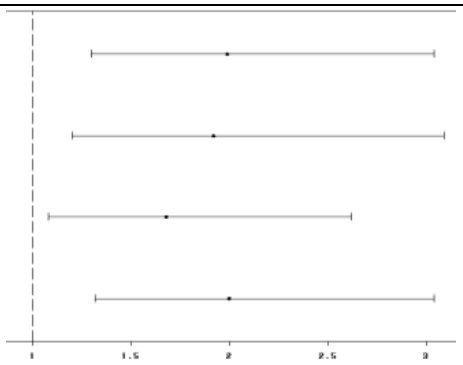
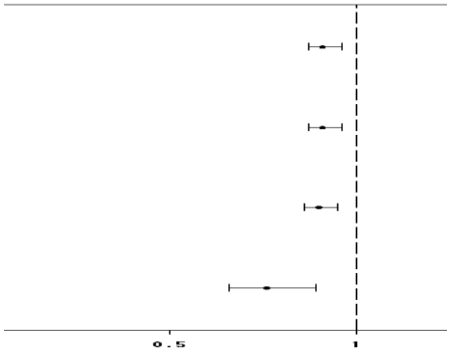
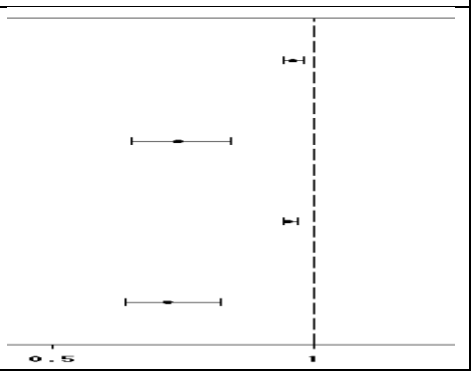
| Variables | Model | Odds ratio (95%CI) | Forest plot |
|---|---|---|---|
| Language spoken at home | GEE | 0.64 (0.44, 0.93) |  |
| | GLMM | 0.64 (0.42, 0.97) | |
| | HGLM | 0.71 (0.48, 1.05) | |
| | Bayesian | 0.63 (0.43, 0.92) | |
| Total income | GEE | 1.99 (1.30, 3.04) |  |
| | GLMM | 1.92 (1.20, 3.09) | |
| | HGLM | 1.68 (1.08, 2.62) | |
| | Bayesian | 2.00 (1.32, 3.04) | |
| Unmet learning needs in hospital | GEE | 0.91 (0.87, 0.96) |  |
| | GLMM | 0.91 (0.87, 0.96) | |
| | HGLM | 0.90 (0.86, 0.95) | |
| | Bayesian | 0.76 (0.66, 0.89) | |
| SF-12 physical component score | GEE | 0.96 (0.94, 0.98) |  |
| | GLMM | 0.74 (0.65, 0.84) | |
| | HGLM | 0.95 (0.94, 0.97) | |
| | Bayesian | 0.72 (0.64, 0.82) | |

**Table 4.1 Summary of Estimates of Variables from Different Models (Continued)**

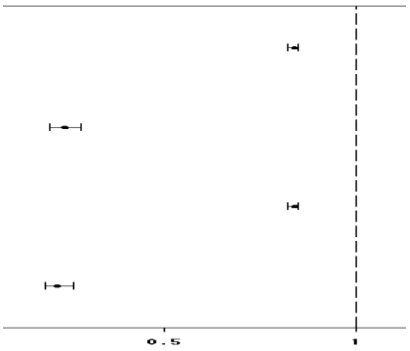
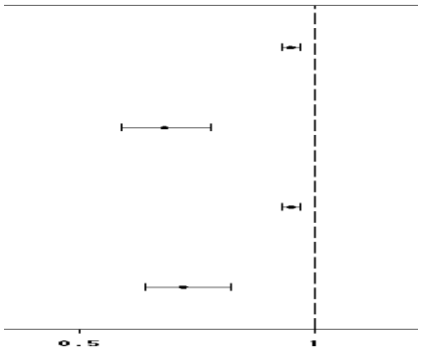
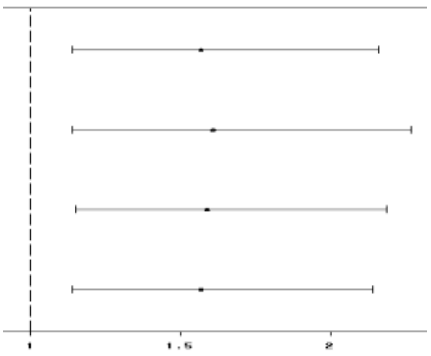
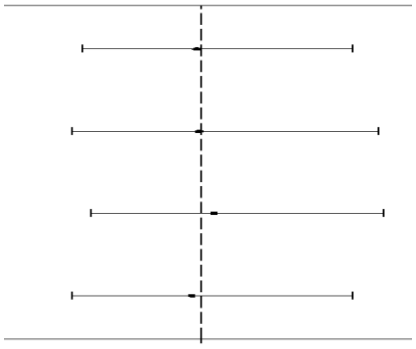
| Variables | Model | Odds ratio (95%CI) | Forest plot |
|---|---|---|---|
| SF-12 mental component score | GEE | 0.84 (0.82, 0.85) |  |
| | GLMM | 0.24 (0.20, 0.28) | |
| | HGLM | 0.84 (0.82, 0.85) | |
| | Bayesian | 0.22 (0.19, 0.26) | |
| Total social support | GEE | 0.95 (0.93, 0.97) |  |
| | GLMM | 0.68 (0.59, 0.78) | |
| | HGLM | 0.95 (0.93, 0.97) | |
| | Bayesian | 0.72 (0.64, 0.82) | |
| Bladder problems | GEE | 1.57 (1.14, 2.16) |  |
| | GLMM | 1.61(1.14, 2.27) | |
| | HGLM | 1.59(1.15, 2.19) | |
| | Bayesian | 1.57(1.14, 2.14) | |
| Delivery method | GEE | 0.99 (0.73, 1.34) |  |
| | GLMM | 1.00 (0.71, 1.40) | |
| | HGLM | 1.03 (0.75, 1.41) | |
| | Bayesian | 0.98 (0.71, 1.34) | |

**Table 4.2 Comparison of Fit Statistics for GEE, GLMM, and HGLM**

| Model | AIC | BIC | Chi-Square/DF | Log-likelihood |
|---|---|---|---|---|
| GEE* | 1403.73 | 1408.62 | 0.74 | -690.20 |
| GLMM | 27151.05 | 27156.56 | 0.44 | -13579.53 |
| HGLM | 28317.57 | 28320.44 | 0.75 | -14156.88 |

*: a modified AIC for GEE

**Table 4.3 Comparison of Fit Statistics for GEE on Different Imputation Methods**

| Imputation | Deviance/DF | Log-likelihood |
|---|---|---|
| Mean | 0.35 | -956.32 |
| LOCF | 0.33 | -889.23 |
| Hot-deck | 0.29 | -1116.10 |
| MI* | 0.34 | -1309.32 |

*: Values are average of fit statistics of 5 imputation datasets

**Table 4.4 Summary of Estimates from Different Imputation Methods**

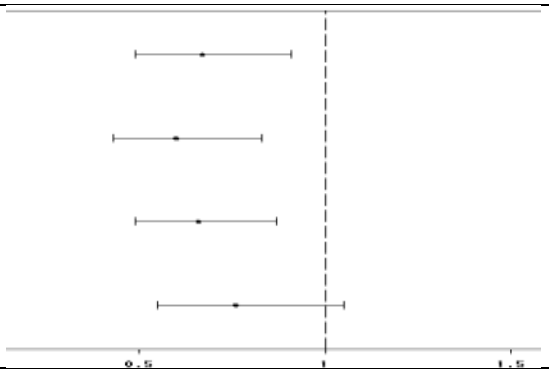
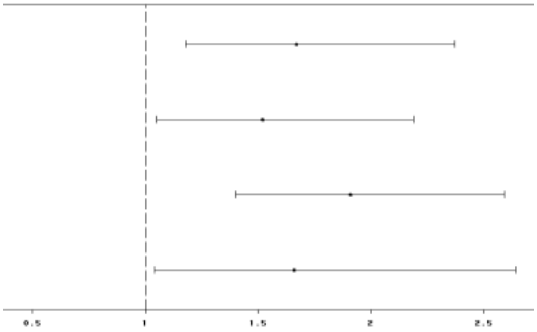
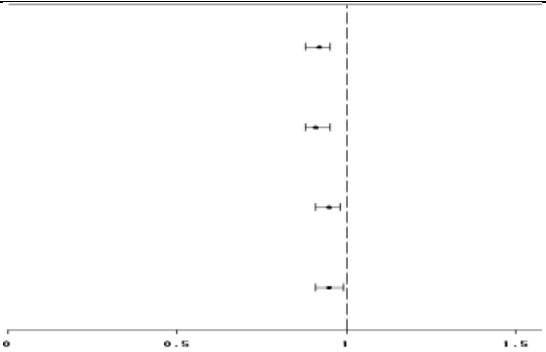
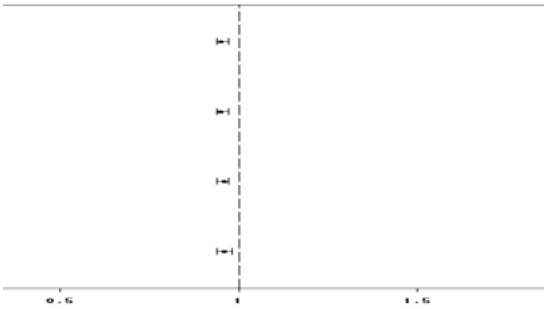
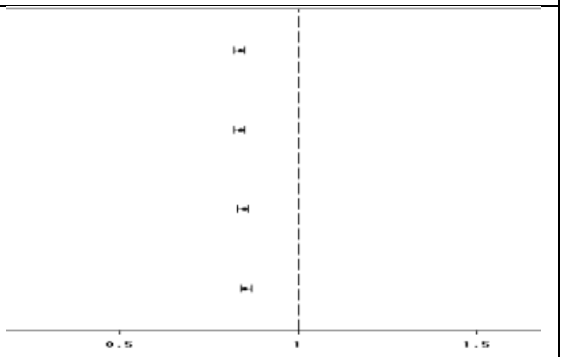
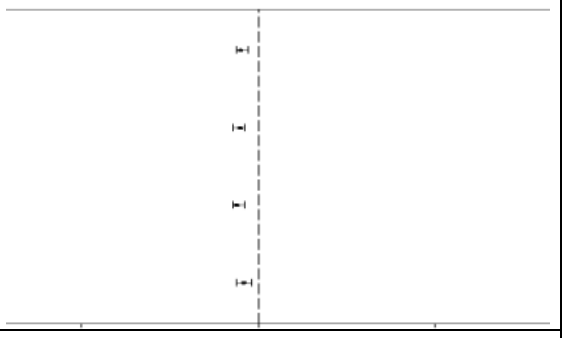
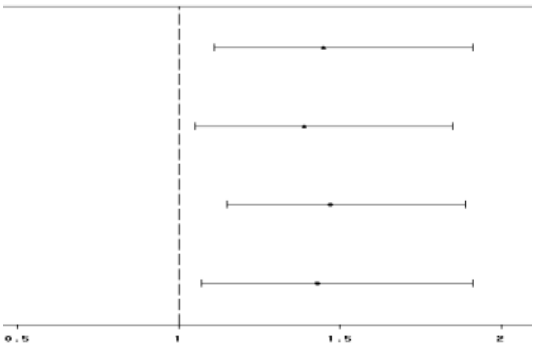
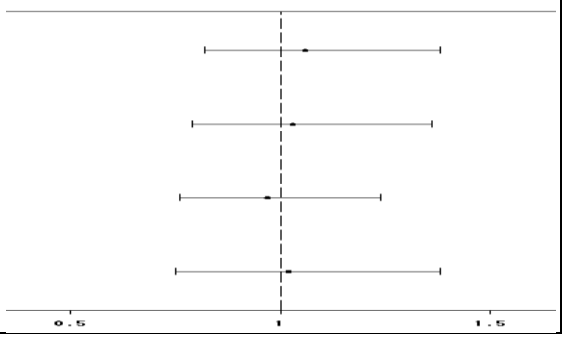
| Variables | Model | Odds ratio (95%CI) | Forest plot |
|---|---|---|---|
| Language spoken at home | Mean | 0.67 (0.49, 0.91) |  |
| | LOCF | 0.60 (0.43, 0.83) | |
| | Hot-Deck | 0.66 (0.49, 0.87) | |
| | MI | 0.76 (0.55, 1.05) | |
| Total income | Mean | 1.67 (1.18, 2.37) |  |
| | LOCF | 1.52 (1.05, 2.19) | |
| | Hot-Deck | 1.91 (1.40, 2.59) | |
| | MI | 1.66 (1.04, 2.64) | |
| Unmet learning needs in hospital | Mean | 0.92 (0.88, 0.95) |  |
| | LOCF | 0.91 (0.88, 0.95) | |
| | Hot-Deck | 0.95 (0.91, 0.98) | |
| | MI | 0.95 (0.91, 0.99) | |
| SF-12 physical component score | Mean | 0.95 (0.94, 0.97) |  |
| | LOCF | 0.95 (0.94, 0.97) | |
| | Hot-Deck | 0.96 (0.94, 0.97) | |
| | MI | 0.96 (0.94, 0.98) | |

**Table 4.4 Summary of Estimates from Different Imputation Methods (continued)**

| Variables | Model | Odds ratio (95%CI) | Forest plot |
|---|---|---|---|
| SF-12 mental component score | Mean | 0.84 (0.82, 0.85) |   |
| | LOCF | 0.84 (0.82, 0.85) | |
| | Hot-Deck | 0.85 (0.83, 0.86) | |
| | MI | 0.85 (0.84, 0.87) | |
| Total social support | Mean | 0.95 (0.94, 0.97) |  |
| | LOCF | 0.95 (0.93, 0.96) | |
| | Hot-Deck | 0.94 (0.93, 0.96) | |
| | MI | 0.96 (0.94, 0.98) | |
| Bladder problems | Mean | 1.45 (1.11, 1.91) |   |
| | LOCF | 1.39 (1.05, 1.85) | |
| | Hot-Deck | 1.47 (1.15, 1.89) | |
| | MI | 1.43 (1.07, 1.91) | |
| Delivery method | Mean | 1.06 (0.82, 1.38) |   |
| | LOCF | 1.03 (0.79, 1.36) | |
| | Hot-Deck | 0.97 (0.76, 1.24) | |
| | MI | 1.02 (0.75, 1.38) | |

# Appendix B Figures

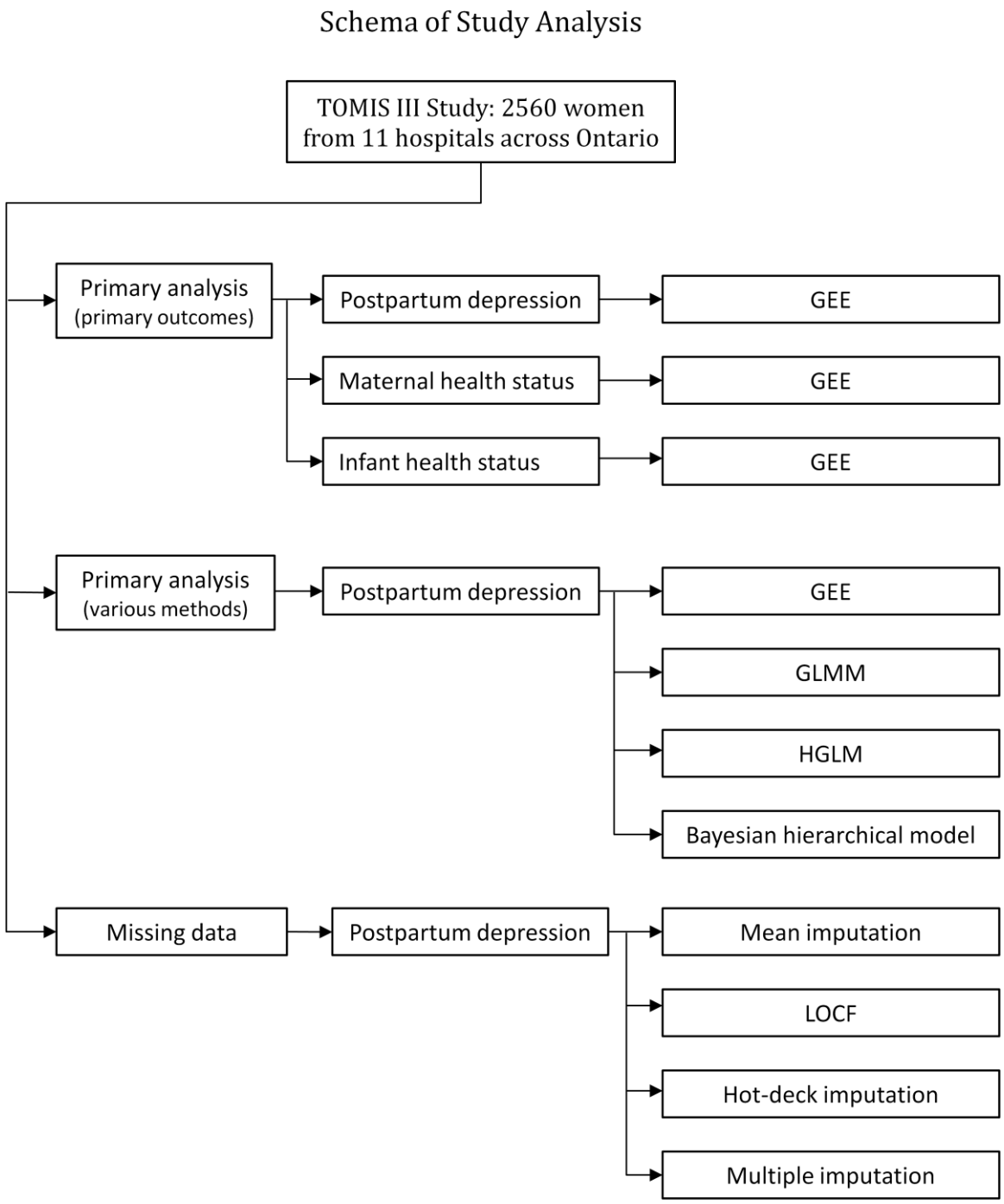



**Figure 2.1 Schema of Study Analysis**

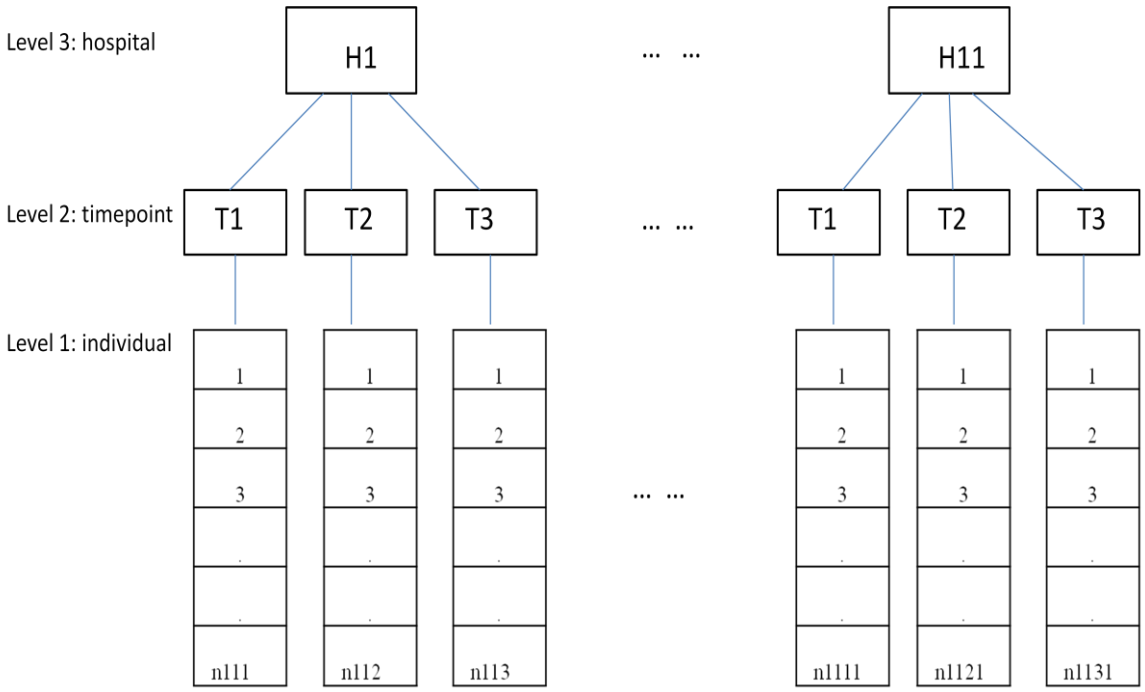



**Figure 2.2 Three-Level Data Structures**

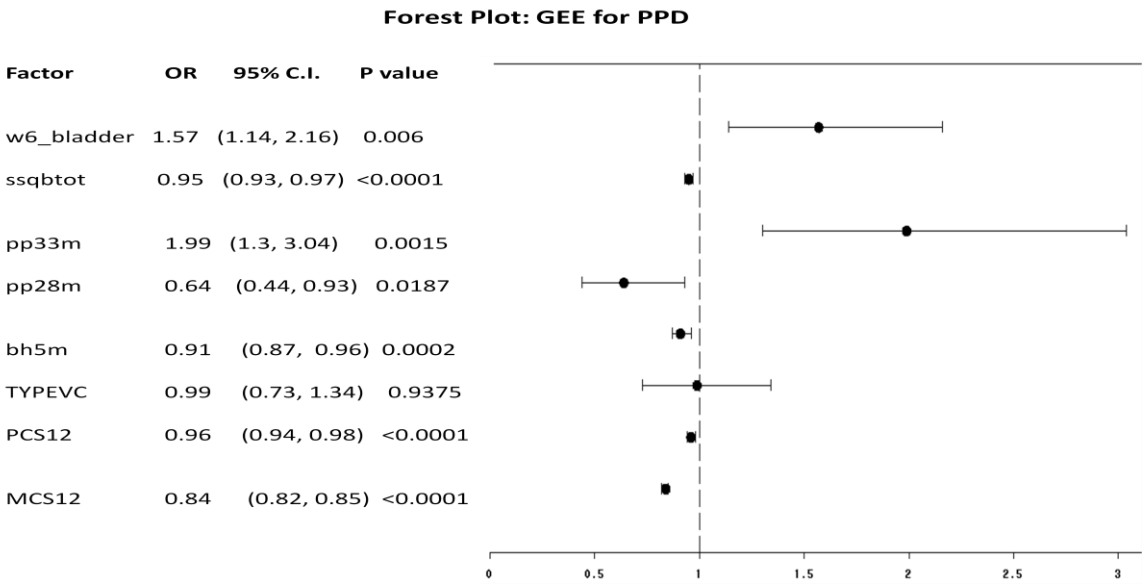



**Figure 3.1 Forest Plot of Postpartum Depression for Covariates (GEE)**

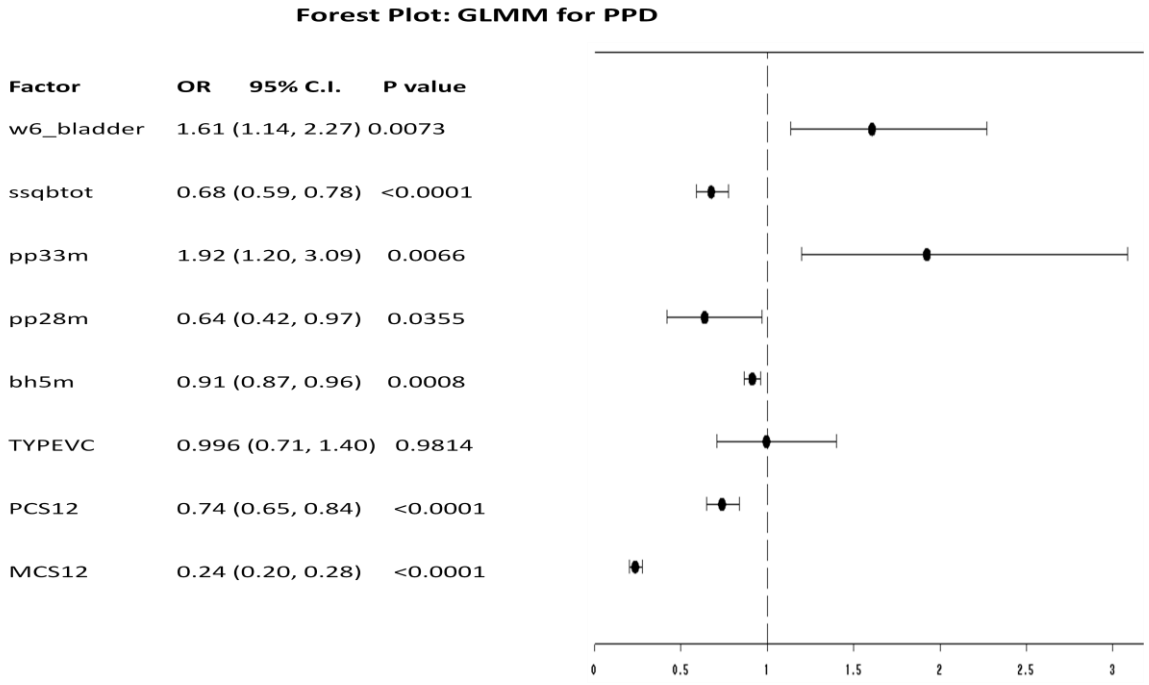



**Figure 3.2 Forest Plot of Postpartum Depression for Covariates (GLMM)**

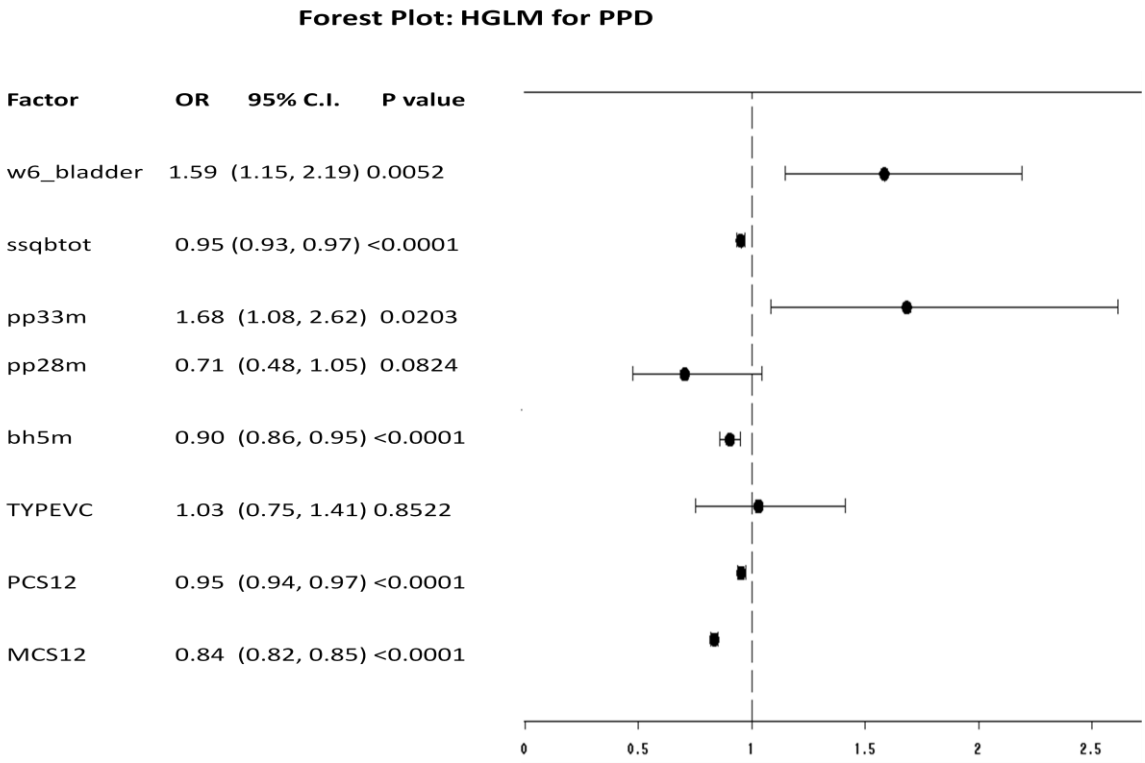



**Figure 3.3 Forest Plot of Postpartum Depression for Covariates (HGLM)**

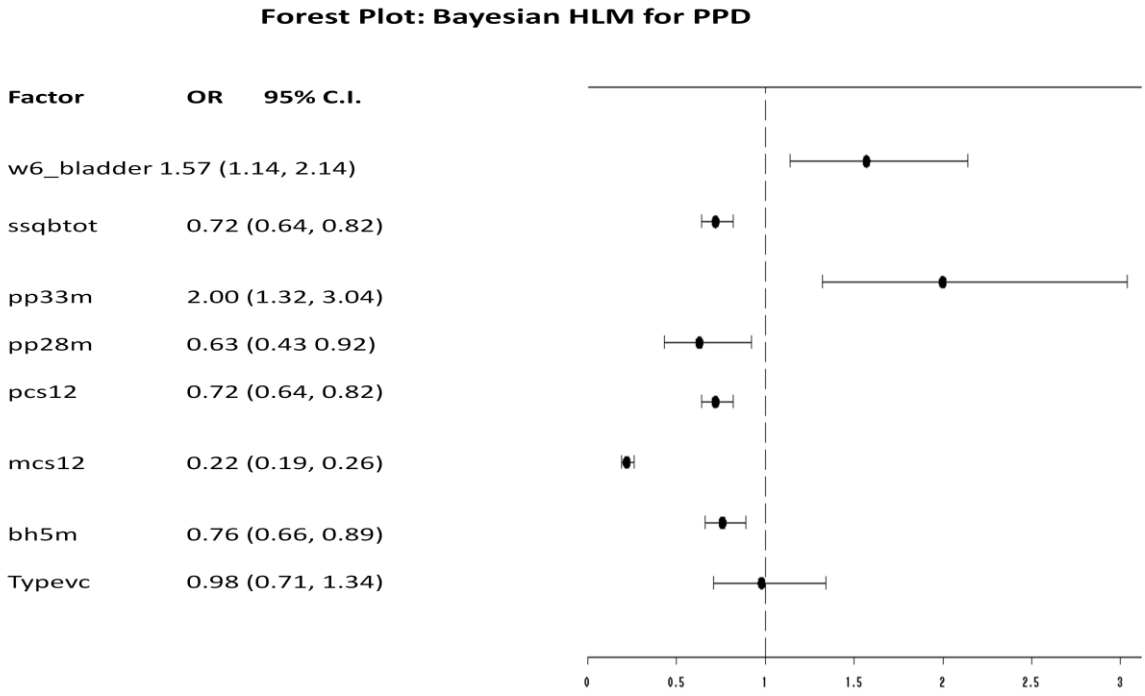



**Figure 3.4 Forest Plot of Postpartum Depression for Covariates (Bayesian)**

**Forest Plot: Sensitivity Analysis**
(Outcome: ppd, covariate: typevc)

| Prior | OR | 95% C.I. |
|---|---|---|
| Uniform (0, 50) | 0.98 | (0.72, 1.34) |
| Uniform (0, 5) | 0.98 | (0.72, 1.33) |
| Uniform (0, 25) | 0.98 | (0.71, 1.35) |
| Uniform (0, 20) | 0.98 | (0.72, 1.34) |
| Uniform (0, 15) | 0.98 | (0.72, 1.35) |
| Uniform (0, 10) | 0.98 | (0.71, 1.34) |
| Igamma (1, 1) | 0.98 | (0.72, 1.35) |
| Igamma (0.1, 0.1) | 0.98 | (0.72, 1.34) |
| Igamma (0.01, 0.01) | 0.98 | (0.72, 1.33) |

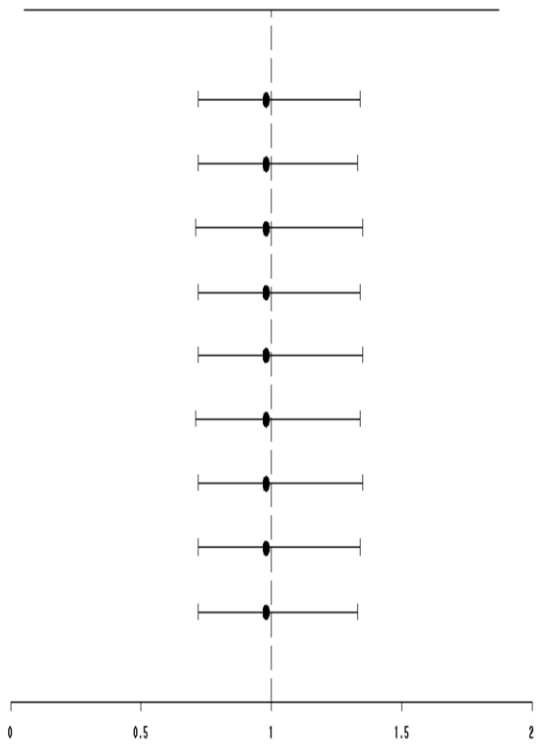



**Figure 3.5 Forest Plot of Sensitivity Analysis for Various Prior Distributions**

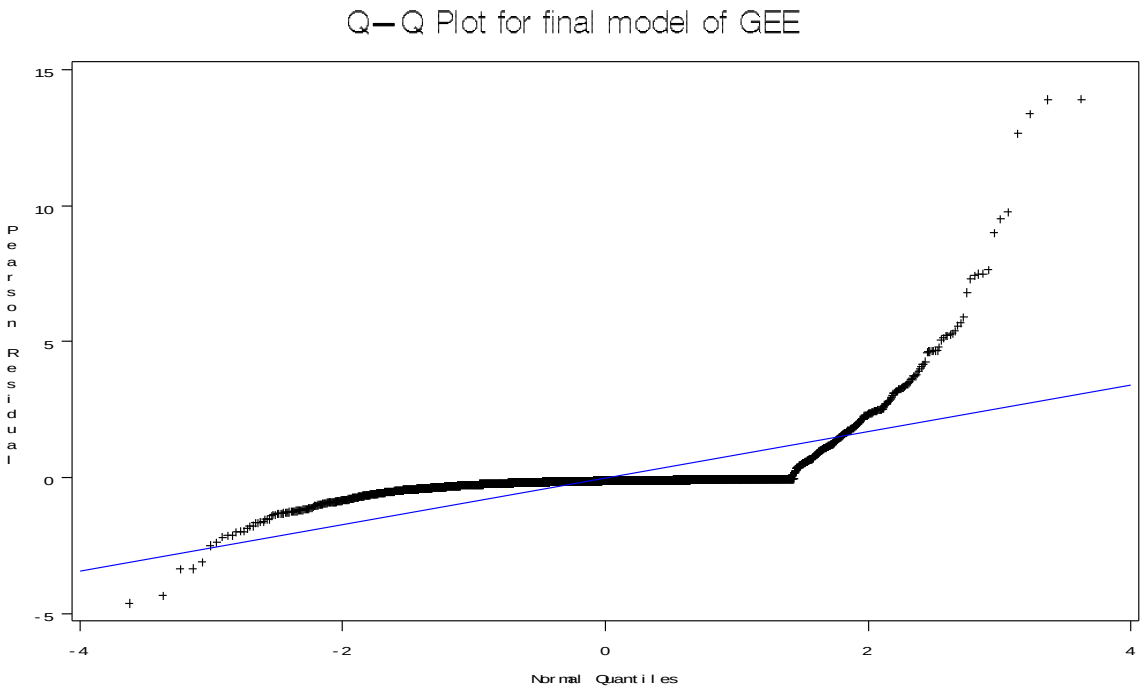



**Figure 3.6 Q-Q Plot for Final Model of GEE**

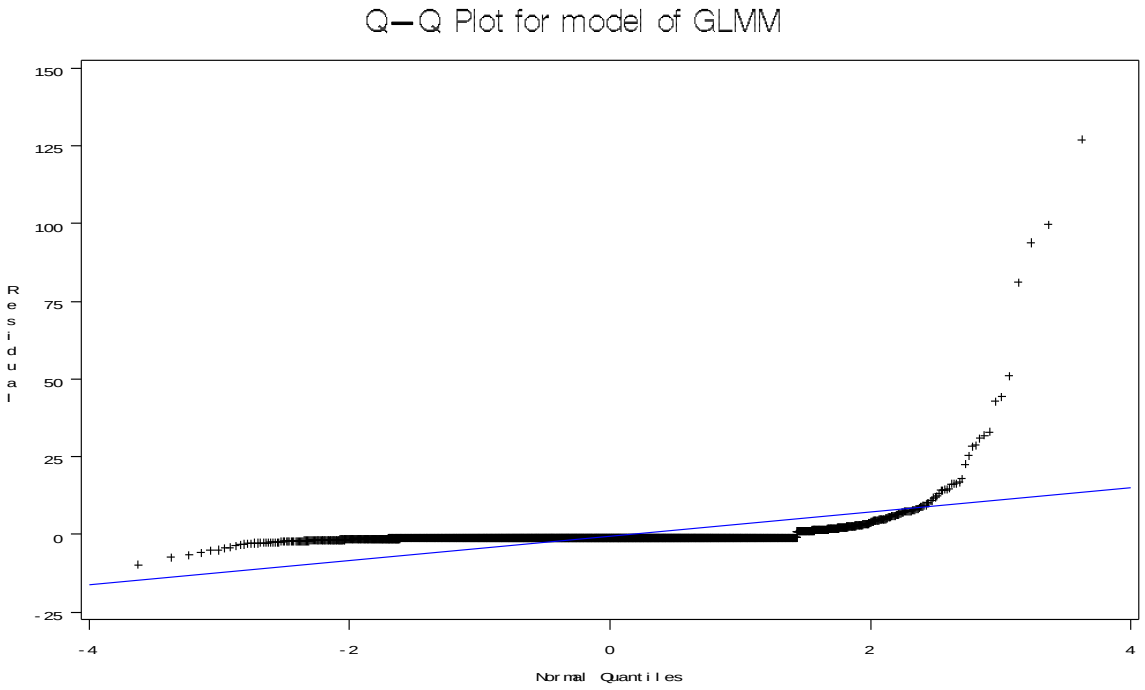



**Figure 3.7 Q-Q Plot for Final Model of GLMM**

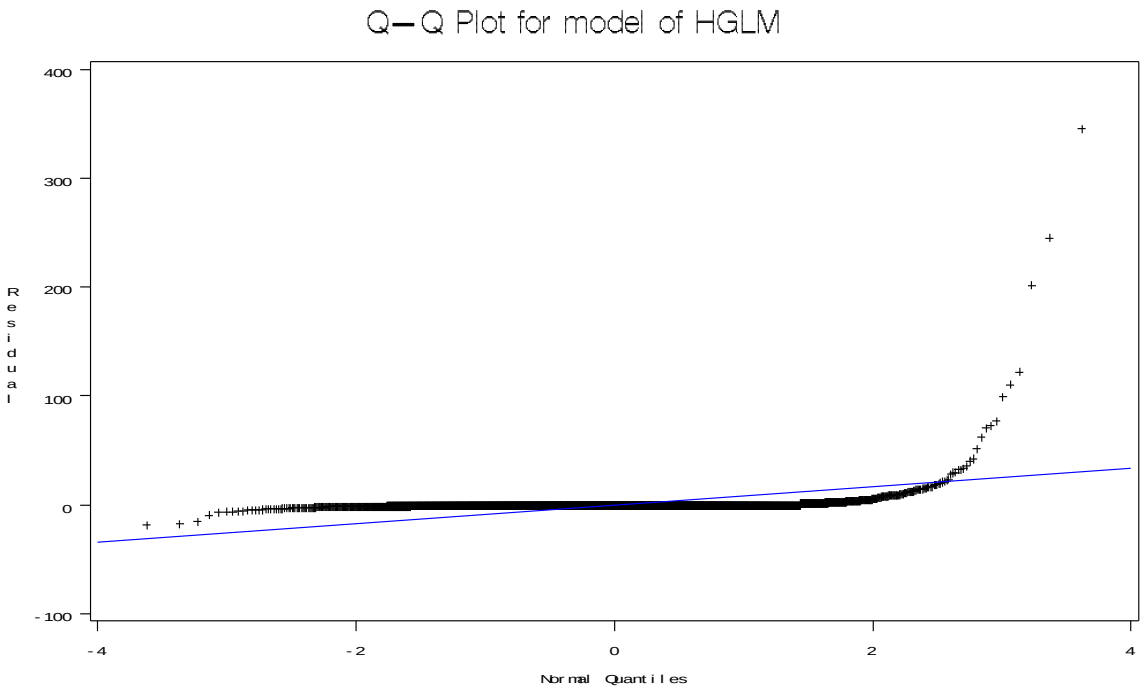



**Figure 3.8 Q-Q Plot for Final Model of HGLM**

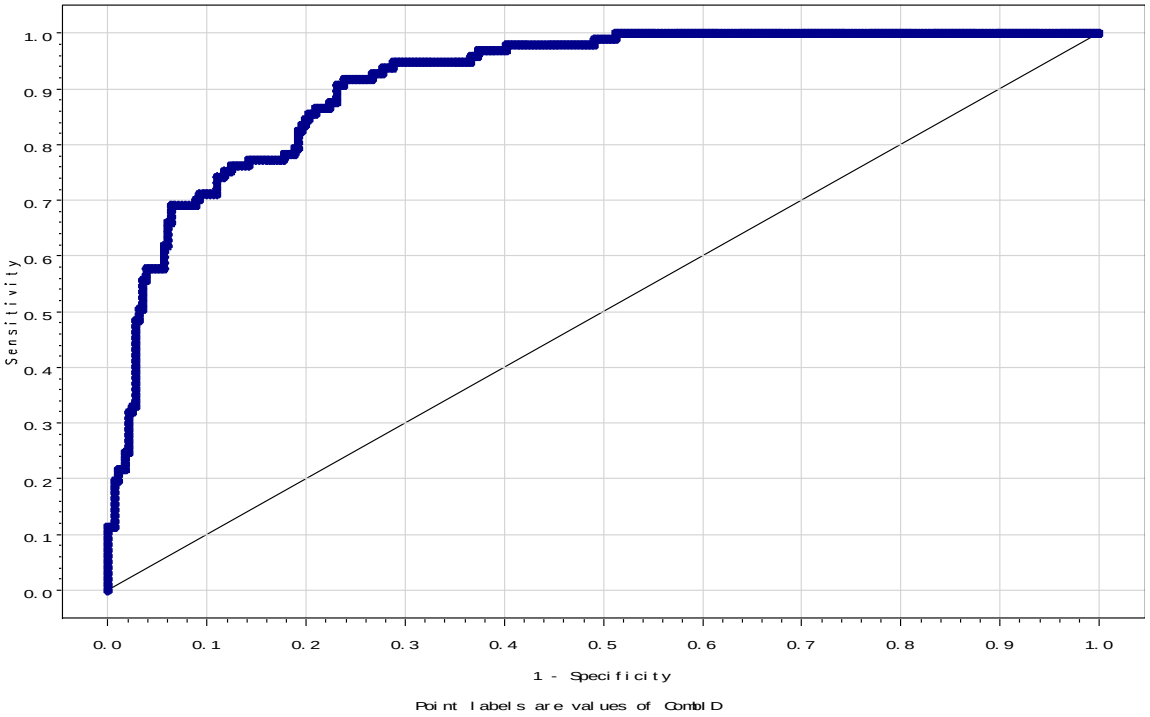



**Figure 3.9 ROC for Final Selected Variables of Maternal Health**

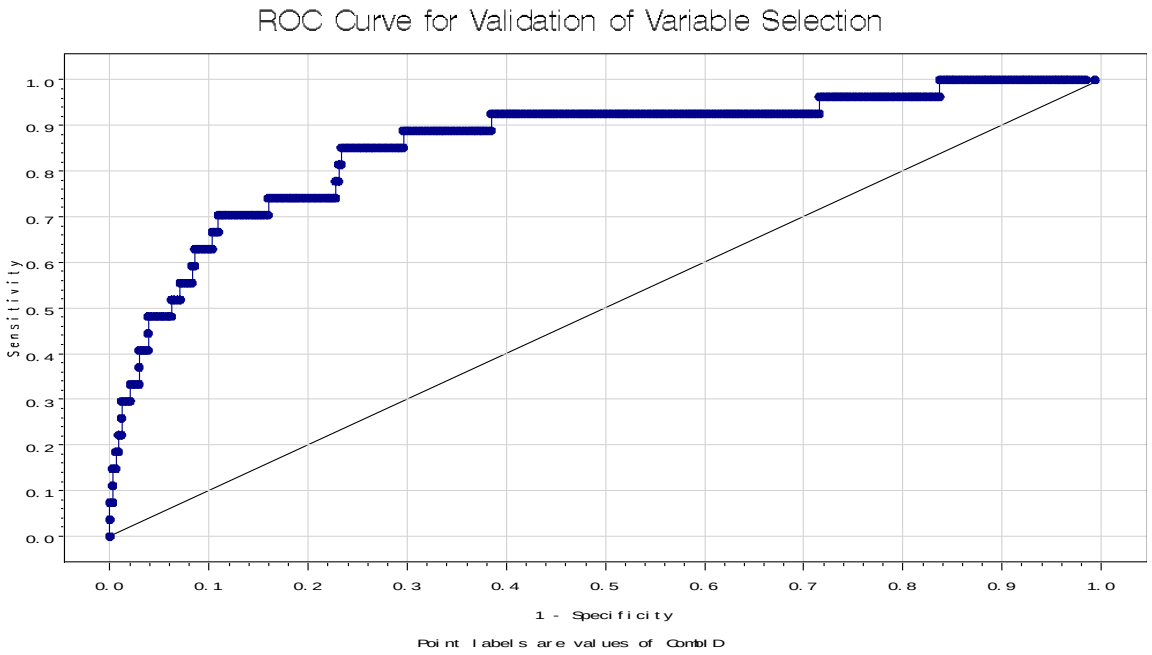



**Figure 3.10 ROC for Final Selected Variables of Infant Health**

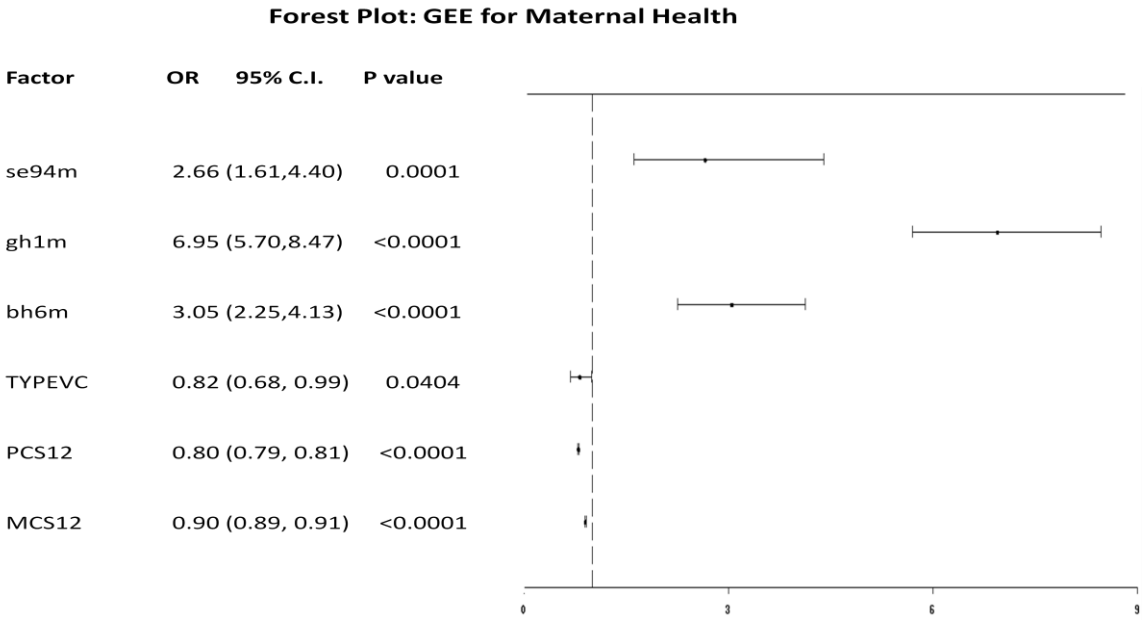



**Figure 3.11Forest Plot of GEE for Maternal Health**

**Forest Plot: GEE for Infant Health**

| Factor | OR | 95% C.I. | P value |
|---|---|---|---|
| se90m | 2.19 | (1.55, 3.08) | <0.0001 |
| se89m | 1.48 | (1.12, 1.97) | 0.0066 |
| pp30m | 1.42 | (1.02, 1.98) | 0.0065 |
| pp28m | 2.02 | (1.42, 2.87) | <0.0001 |
| gh1m | 1.58 | (1.21, 2.06) | 0.0008 |
| gh1_2m | 3.00 | (2.29, 3.93) | <0.0001 |
| TYPEVC | 0.97 | (0.75, 1.24) | 0.8004 |
| MCS12 | 0.98 | (0.96, 0.99) | 0.0003 |

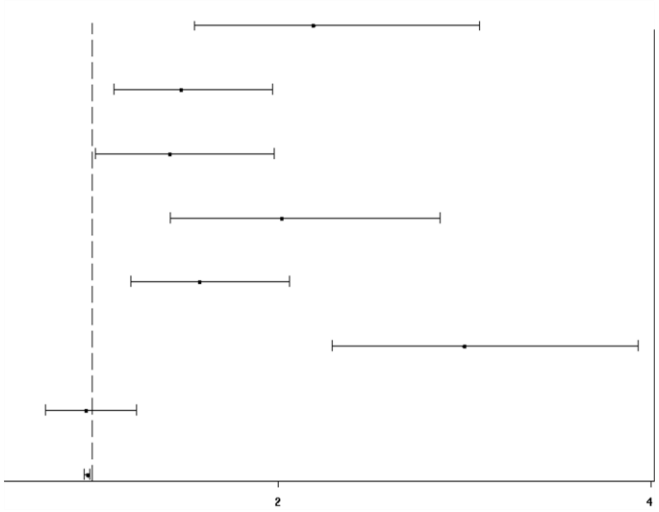



**Figure 3.12 Forest Plot of GEE for Infant Health**

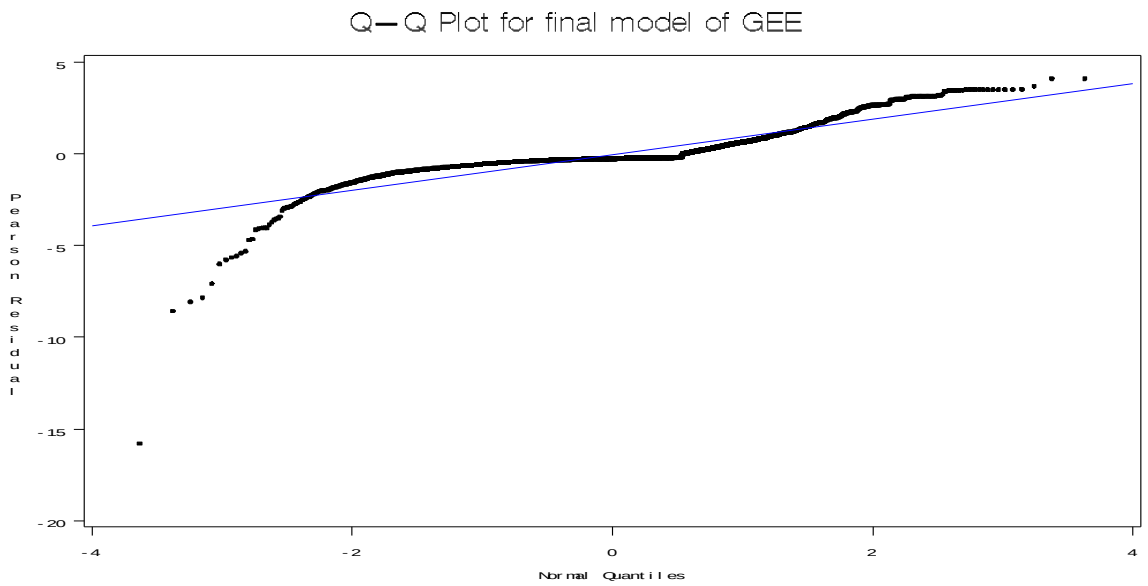



**Figure 3.13 Q-Q Plot of GEE for Maternal Health**

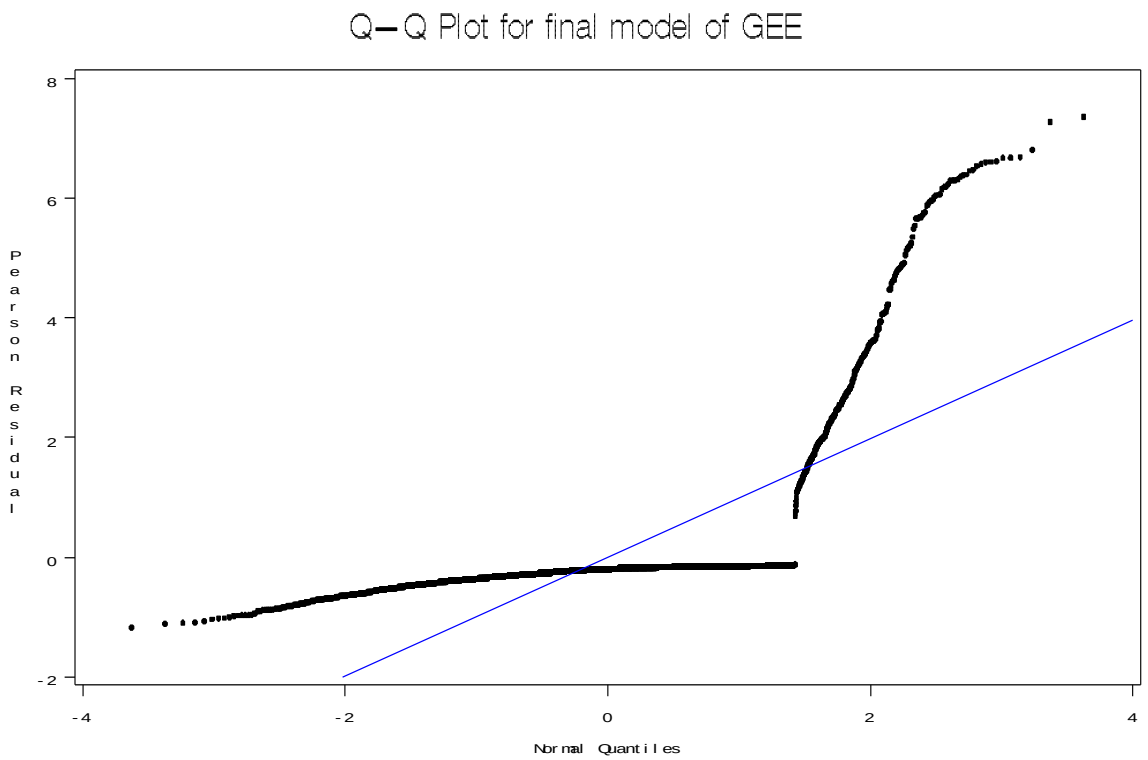



**Figure 3.14 Q-Q Plot of GEE for Infant Health**

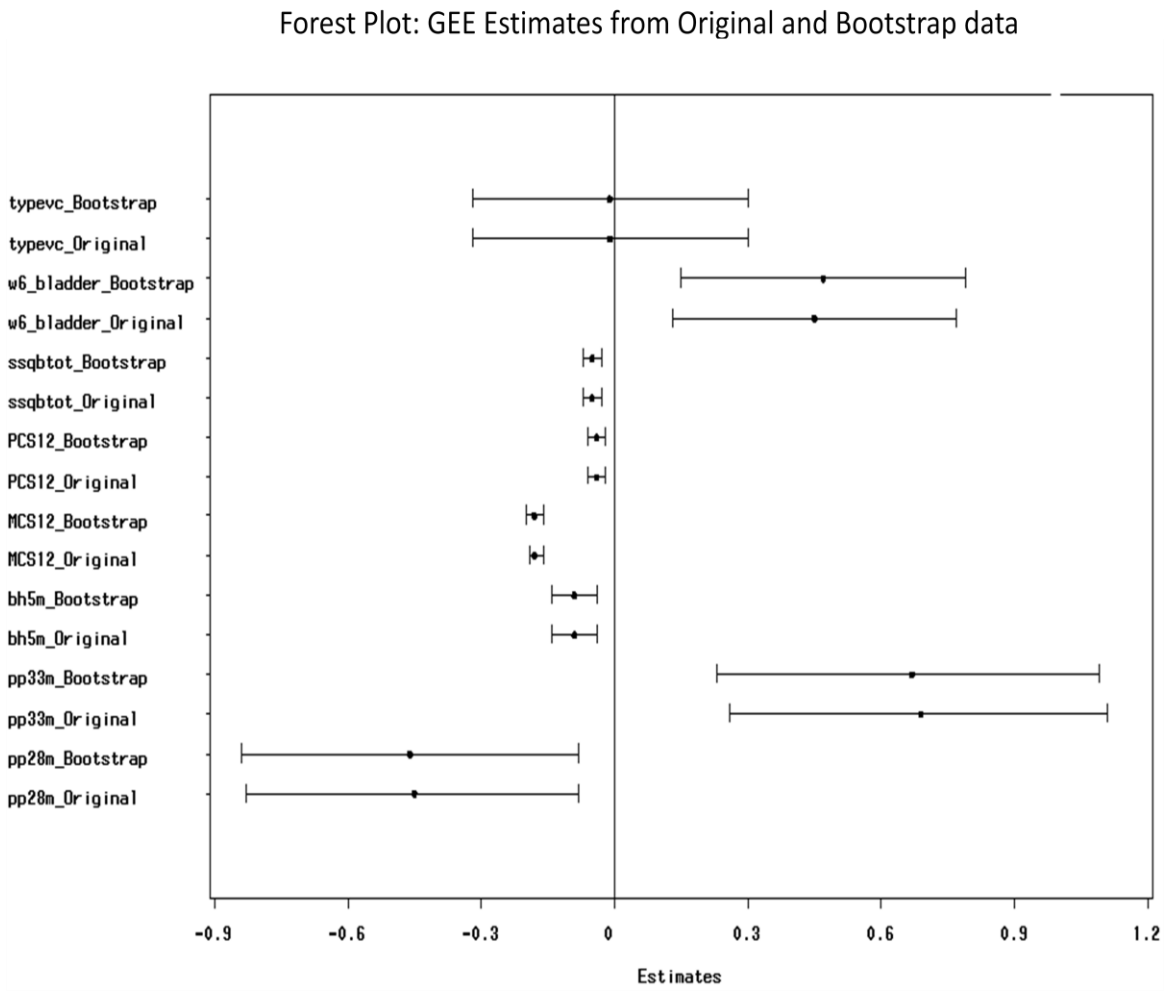



**Figure 3.15 Comparison of GEE Estimates on Original and Bootstrap Data**

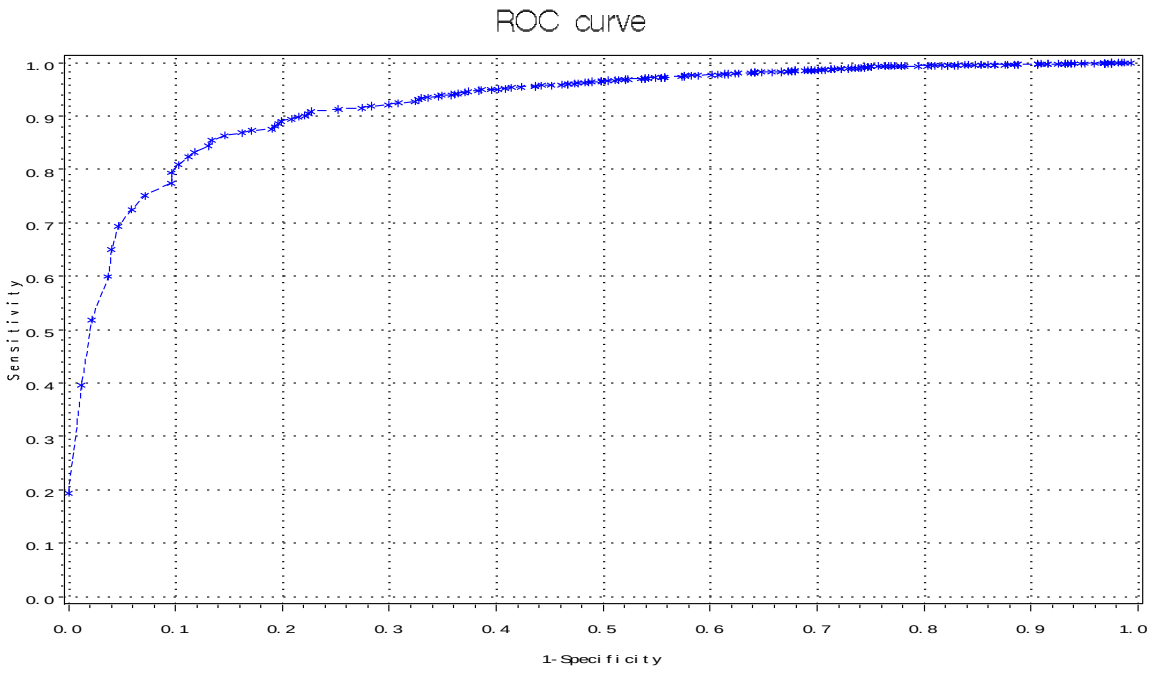



**Figure 3.16 ROC for Final GEE model of Postpartum Depression**

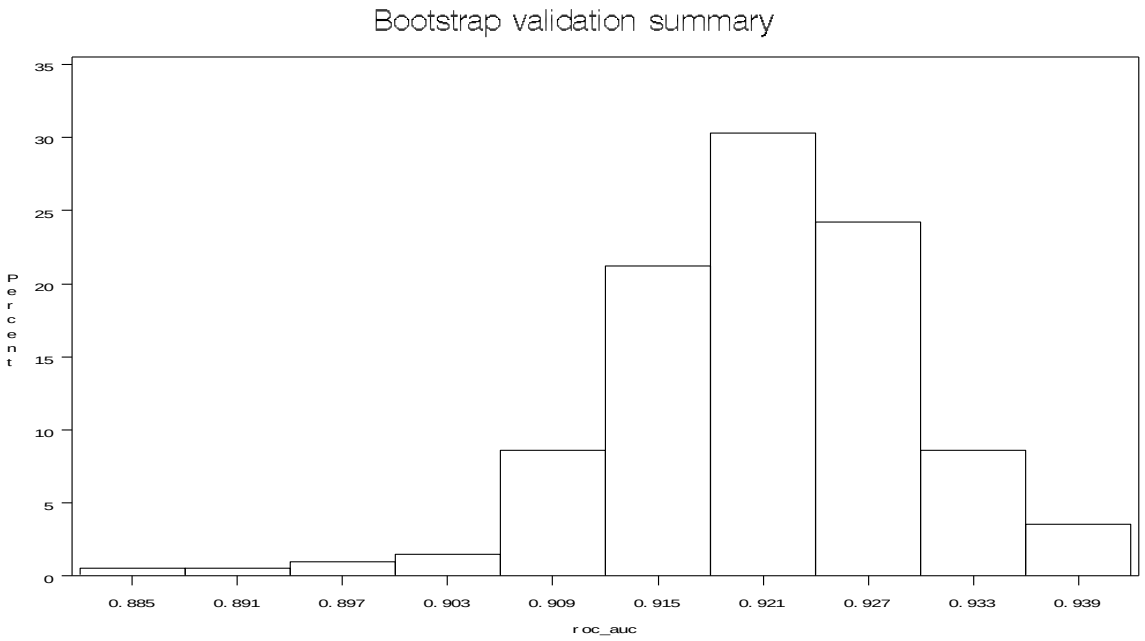



**Figure 3.17 Distribution of AUCs of Bootstrap Models**

**Missing Data Patterns**

| Group | ppd_num | pp28m | pp33m | bh5m | PCS12 | MCS12 | ssqbtot | w6_bladder | TYPEVC | Timepoint |
|---|---|---|---|---|---|---|---|---|---|---|
| 1 | X | X | X | X | X | X | X | X | X | X |
| 2 | X | X | X | X | X | X | X | . | X | X |
| 3 | X | X | X | X | X | X | . | X | X | X |
| 4 | X | X | X | X | . | . | X | X | X | X |
| 5 | X | X | X | . | X | X | X | X | X | X |
| 6 | X | X | X | . | X | X | X | . | X | X |
| 7 | X | X | X | . | . | . | X | X | X | X |
| 8 | X | X | . | X | X | X | X | X | X | X |
| 9 | X | X | . | X | . | . | X | X | X | X |
| 10 | X | X | . | . | X | X | X | X | X | X |
| 11 | X | . | X | X | X | X | X | X | X | X |
| 12 | X | . | . | X | X | X | X | X | X | X |
| 13 | . | X | X | X | X | X | X | X | X | X |
| 14 | . | X | X | X | X | X | X | . | X | X |
| 15 | . | X | X | X | X | X | . | X | X | X |
| 16 | . | X | X | X | X | X | . | . | X | X |
| 17 | . | X | X | X | . | . | X | X | X | X |
| 18 | . | X | X | X | . | . | X | . | X | X |
| 19 | . | X | X | X | . | . | . | . | X | X |
| 20 | . | X | X | . | X | X | X | . | X | X |
| 21 | . | X | X | . | . | . | . | . | X | X |
| 22 | . | X | . | X | . | . | . | . | X | X |
| 23 | . | X | . | . | . | . | . | . | X | X |
| 24 | . | . | X | X | . | . | . | . | X | X |
| 25 | . | . | . | X | . | . | . | . | X | X |
| 26 | . | . | . | . | . | . | . | . | X | X |

**Figure 3.18 Missing Patterns for Outcome of Postpartum Depression**

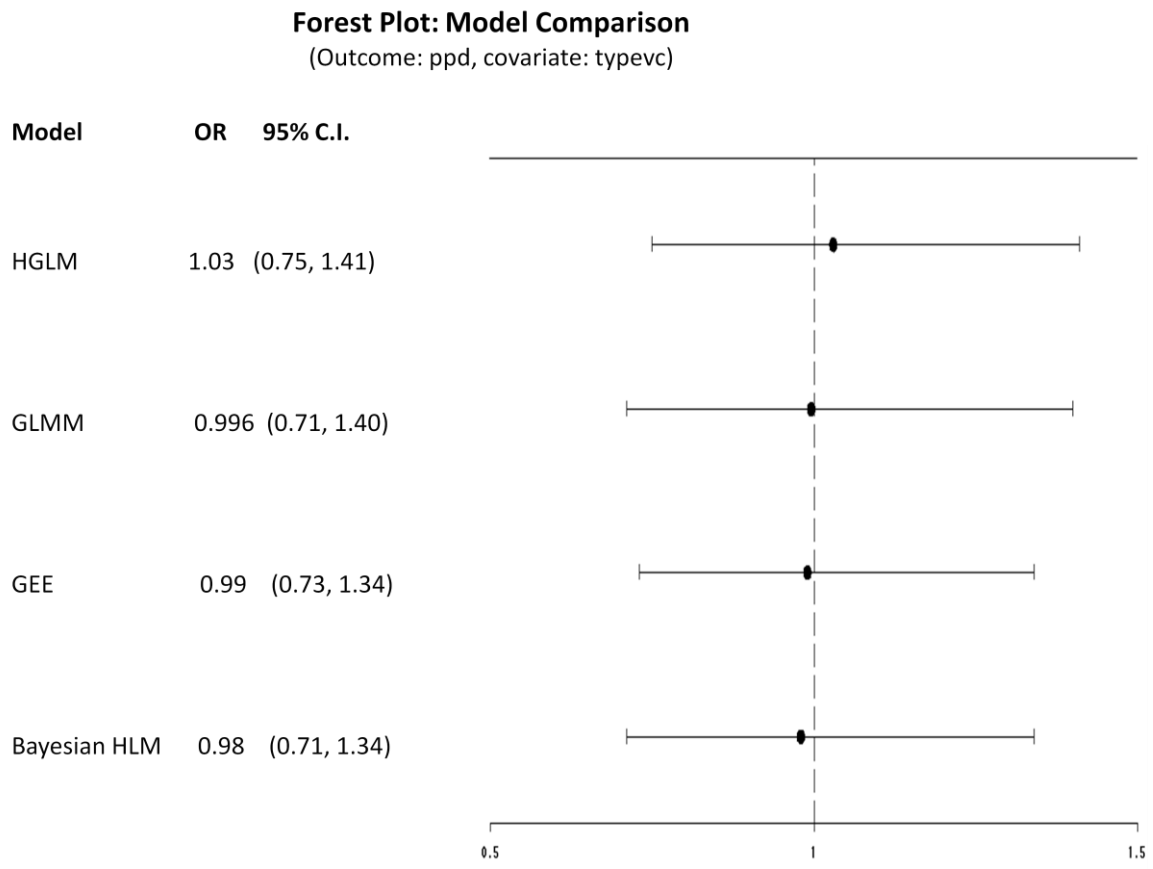



**Figure 3.19 Forest Plot for Modeling Comparisons on Postpartum Depression**

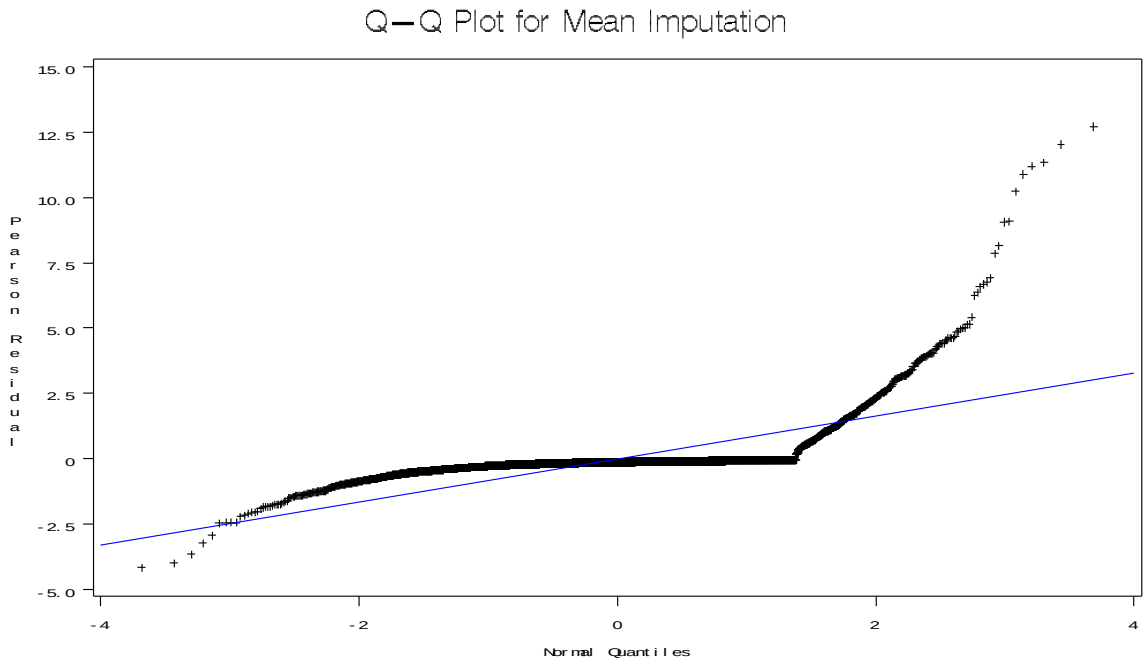



**Figure 4.1 Q-Q Plot for GEE on Mean Imputation Data**

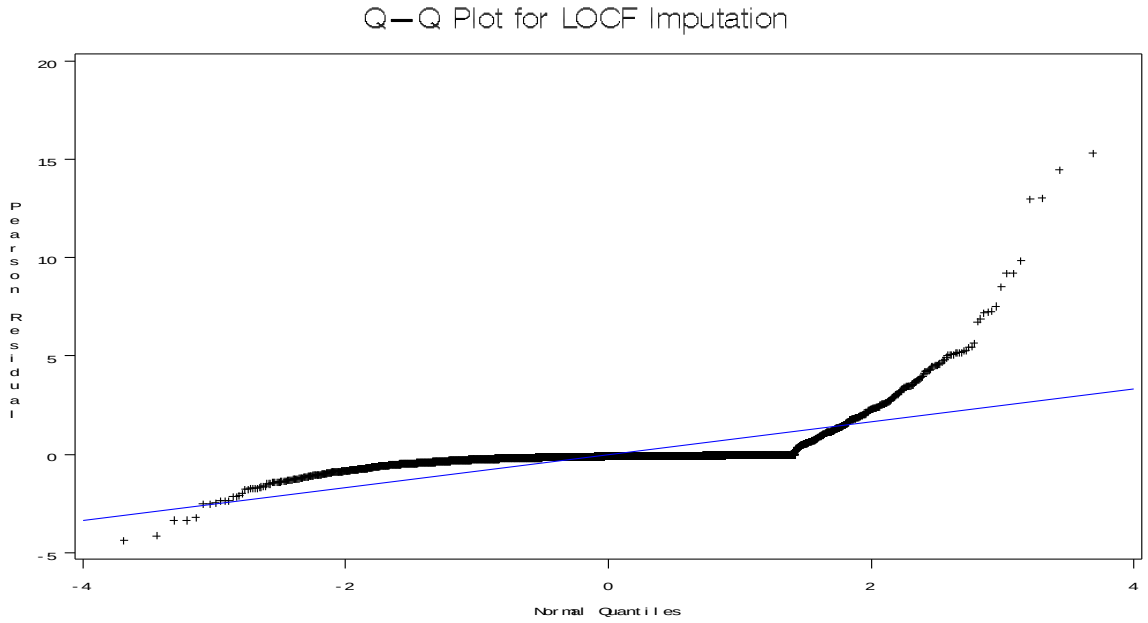



**Figure 4.2 Q-Q Plot for GEE on LOCF Imputation Data**

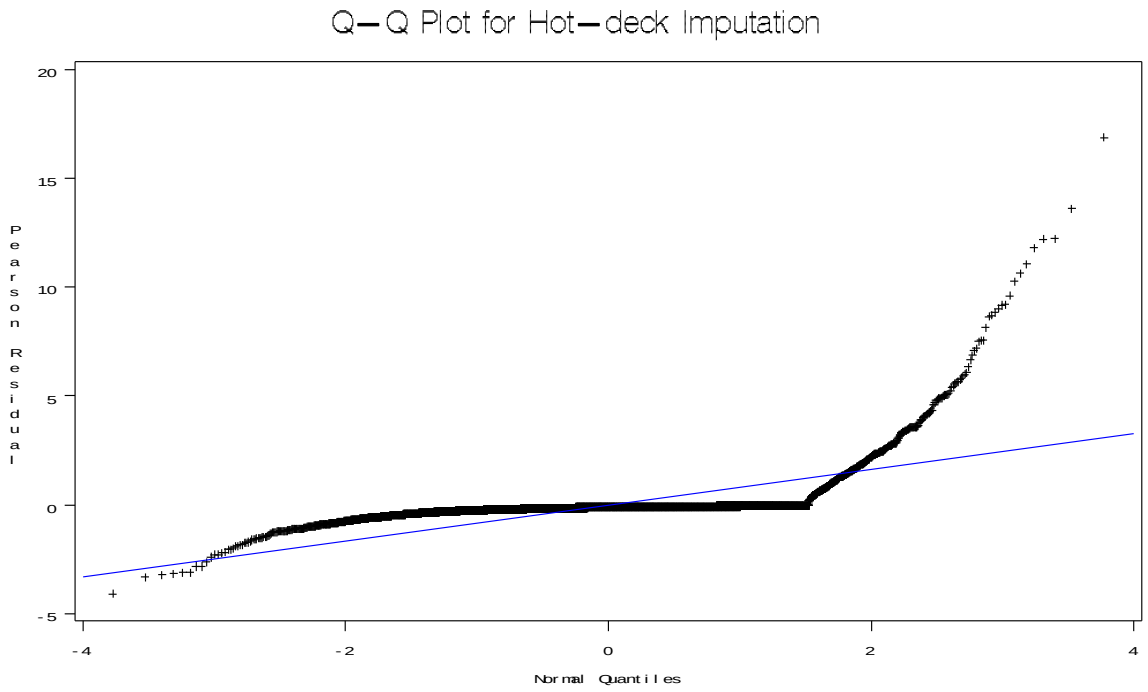



**Figure 4.3 Q-Q Plot for GEE on Hot-deck Imputation Data**

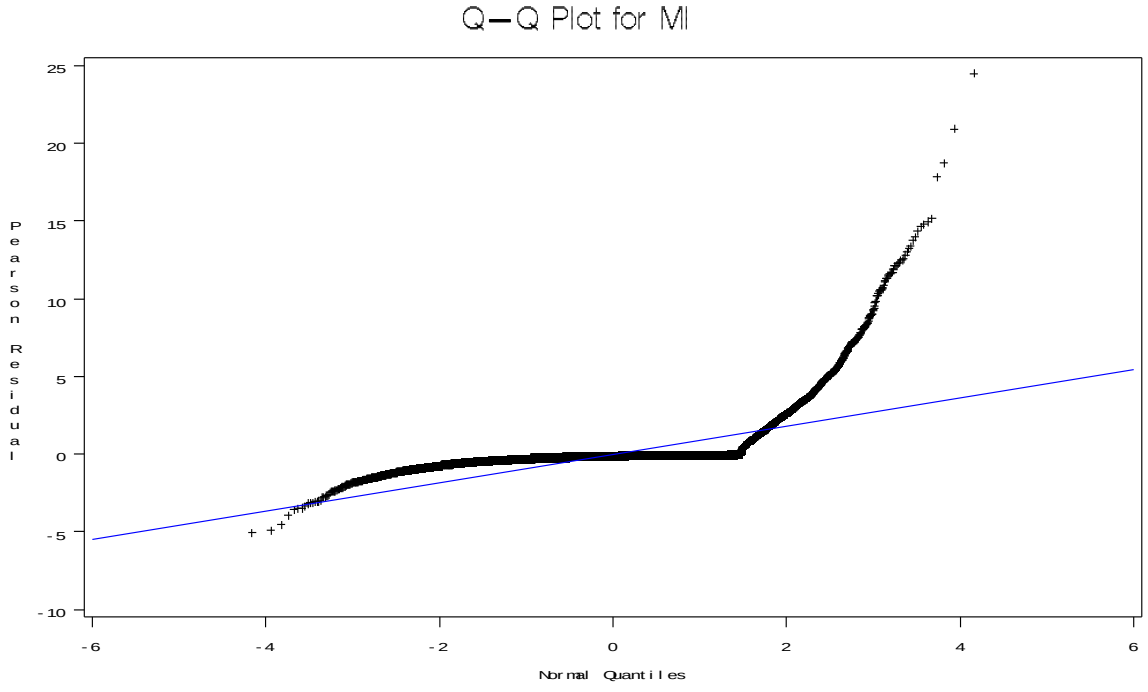
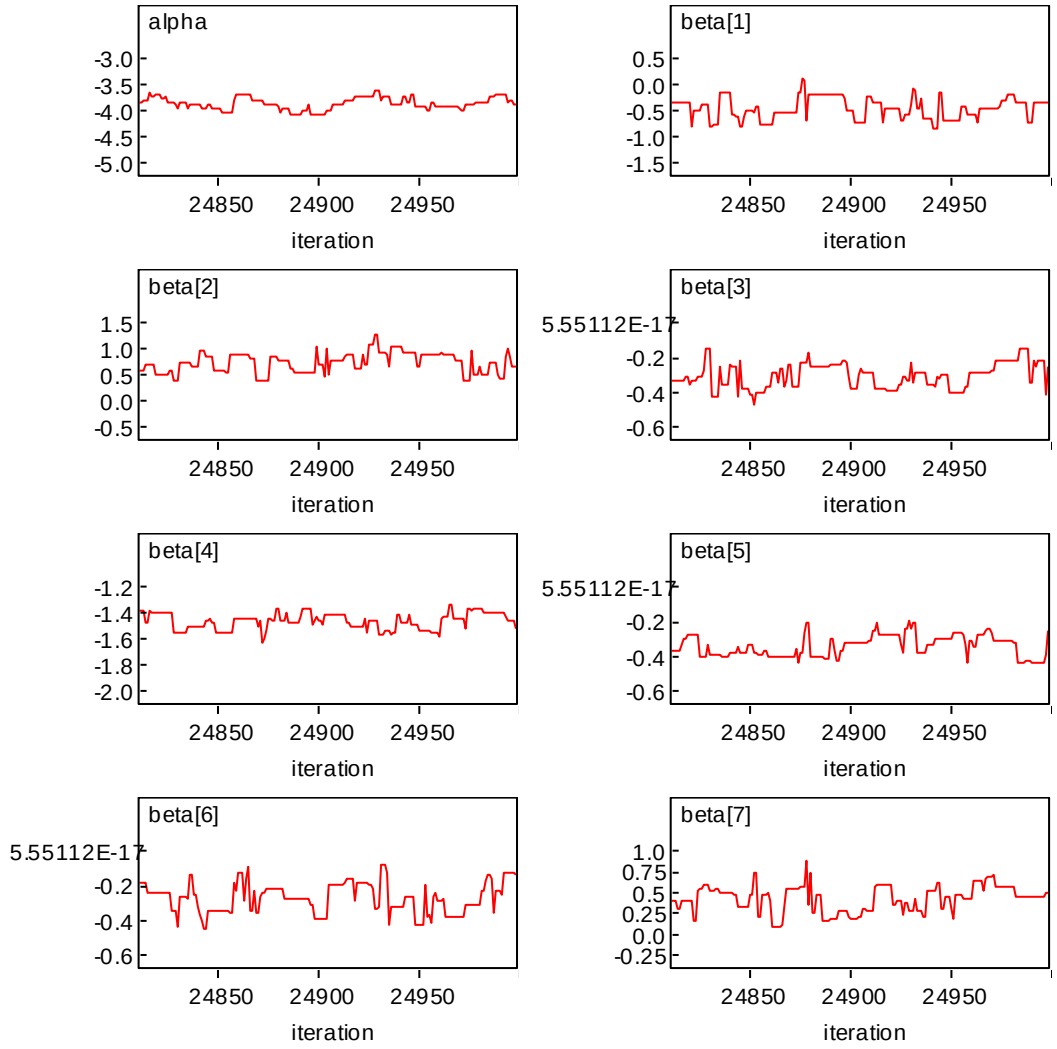
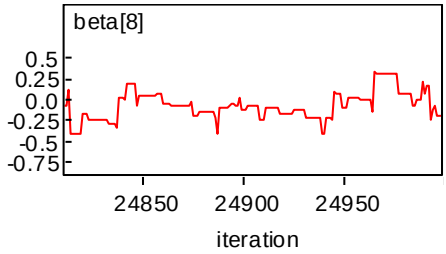


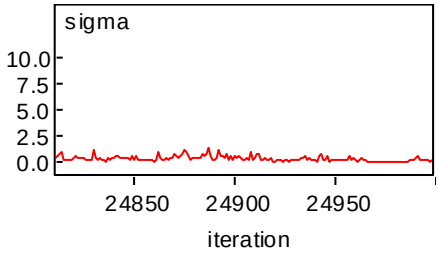

**Figure 4.4 Q-Q Plot for GEE on Multiple Imputation Data**

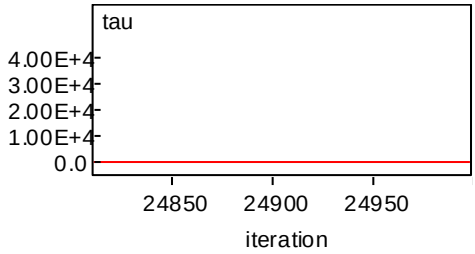
**Figure 4.5 Diagnosis Plot for Bayesian Analysis**

Dynamic trace

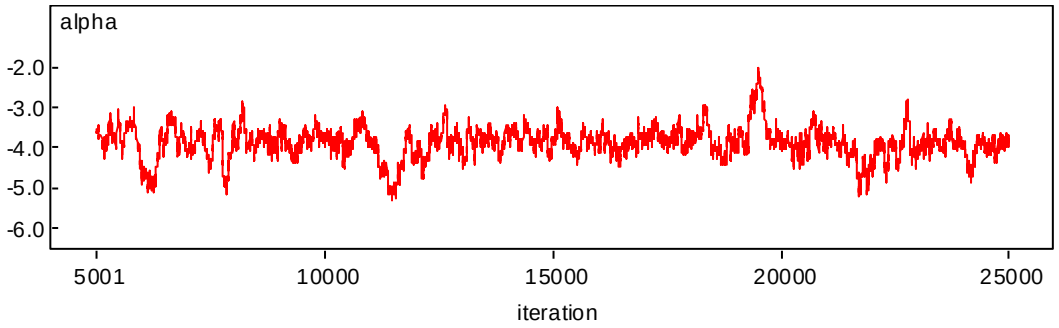


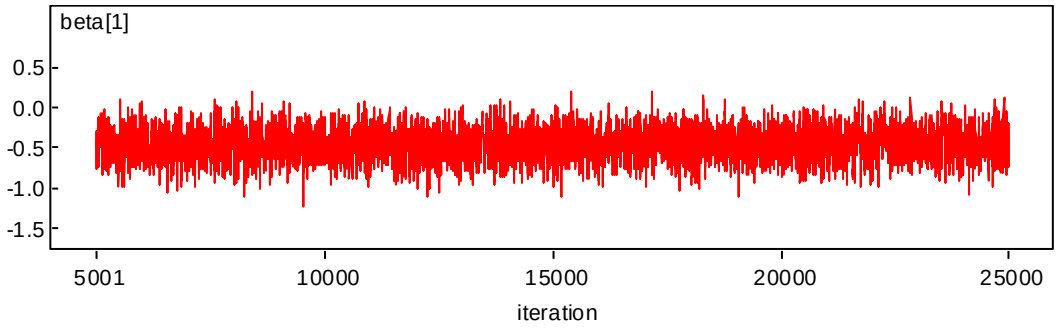

beta[8]
0.5
0.25
0.0
-0.25
-0.5
-0.75
24850
24900
24950
iteration

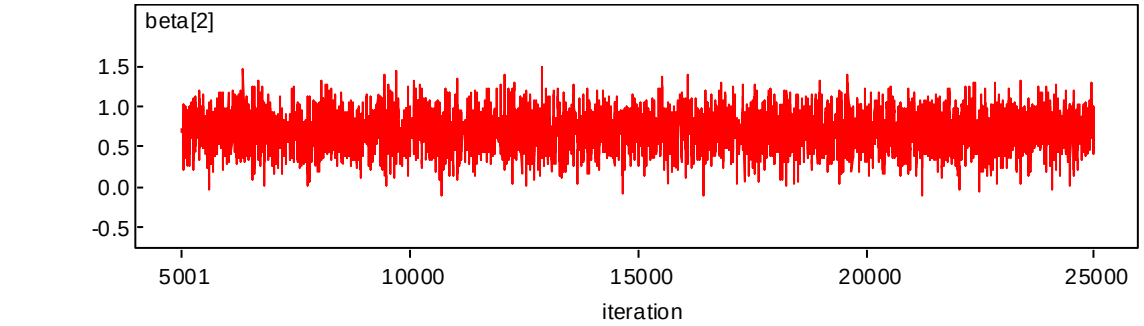

sigma
10.0
7.5
5.0
2.5
0.0
24850
24900
24950
iteration

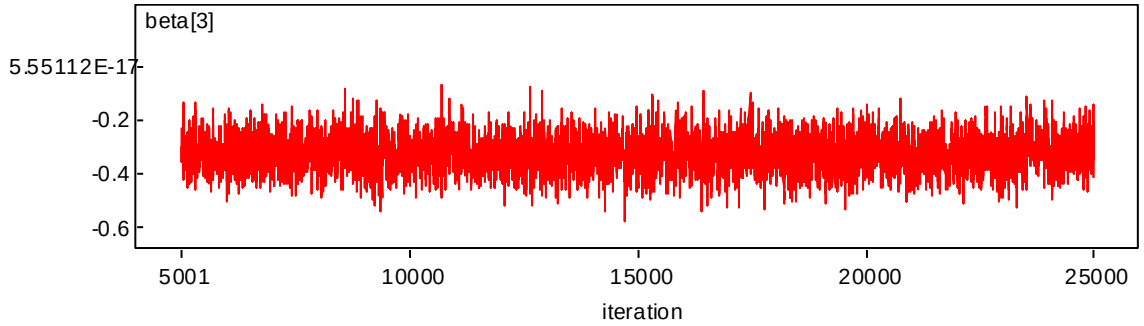

tau
4.00E+4
3.00E+4
2.00E+4
1.00E+4
0.0
24850
24900
24950
iteration

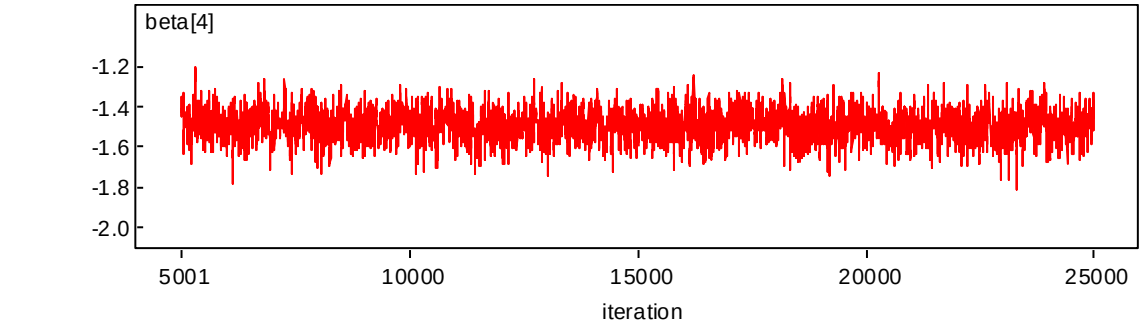
History

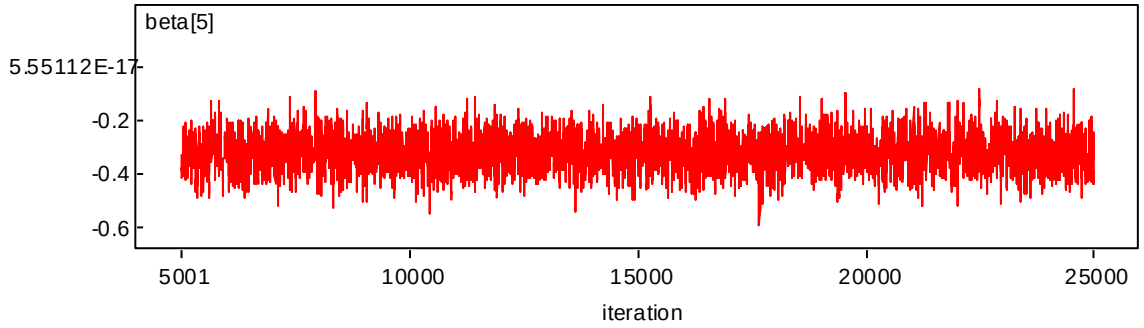

alpha
-2.0
-3.0
-4.0
-5.0
-6.0
5001
10000
15000
20000
25000
iteration

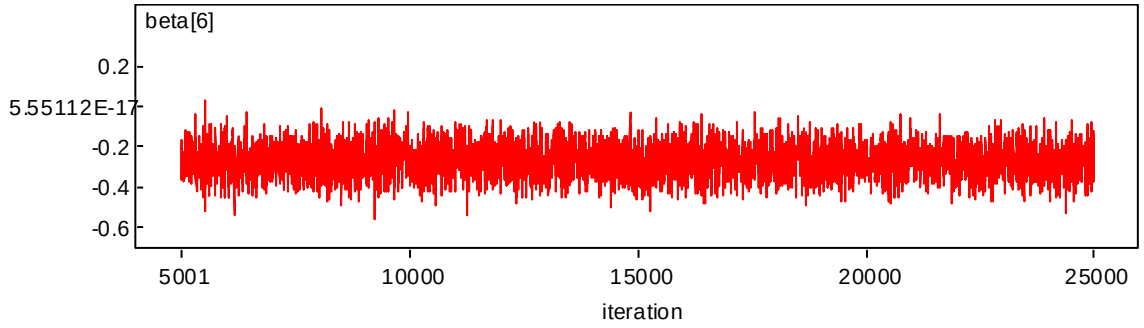

beta[1]
0.5
0.0
-0.5
-1.0
-1.5
5001
10000
15000
20000
25000
iteration

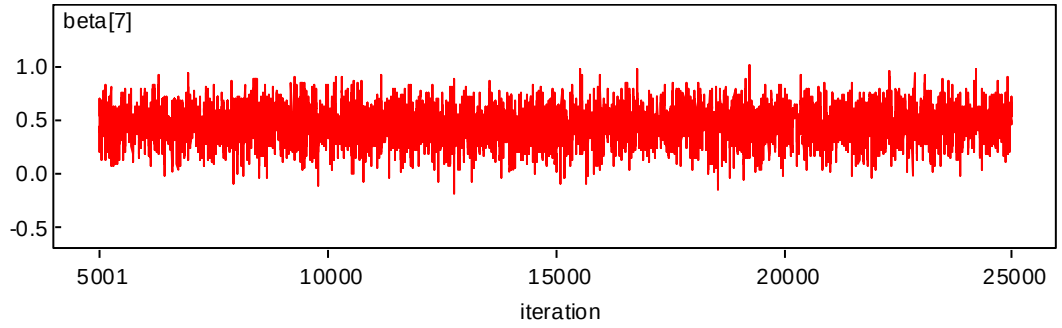

beta[2]
1.5
1.0
0.5
0.0
-0.5
5001
10000
15000
20000
25000
iteration

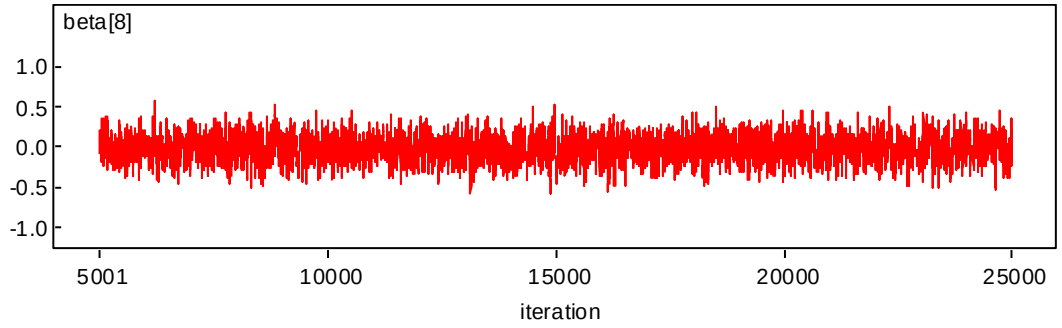

beta[3]
5.55112E-17
-0.2
-0.4
-0.6
5001
10000
15000
20000
25000
iteration

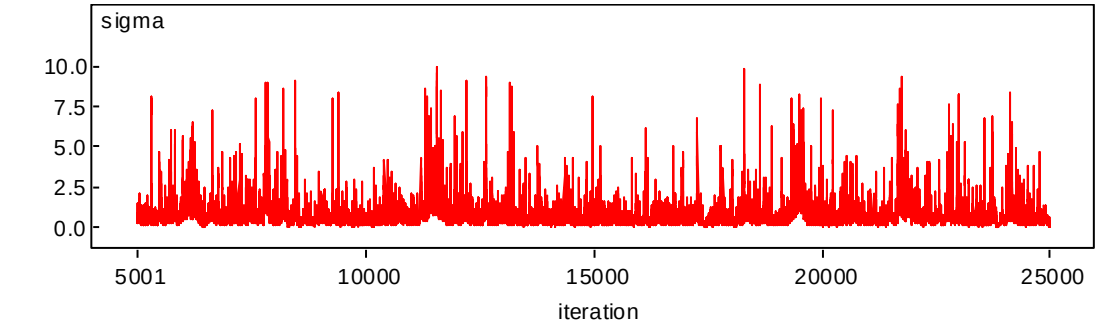

beta[4]
-1.2
-1.4
-1.6
-1.8
-2.0
5001
10000
15000
20000
25000
iteration

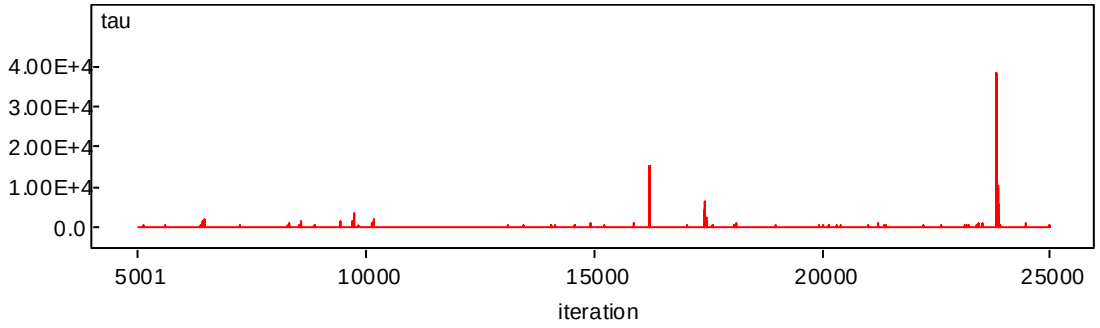

beta[5]
5.55112E-17
-0.2
-0.4
-0.6
5001
10000
15000
20000
25000
iteration


beta[6]
0.2
5.55112E-17
-0.2
-0.4
-0.6
5001
10000
15000
20000
25000
iteration

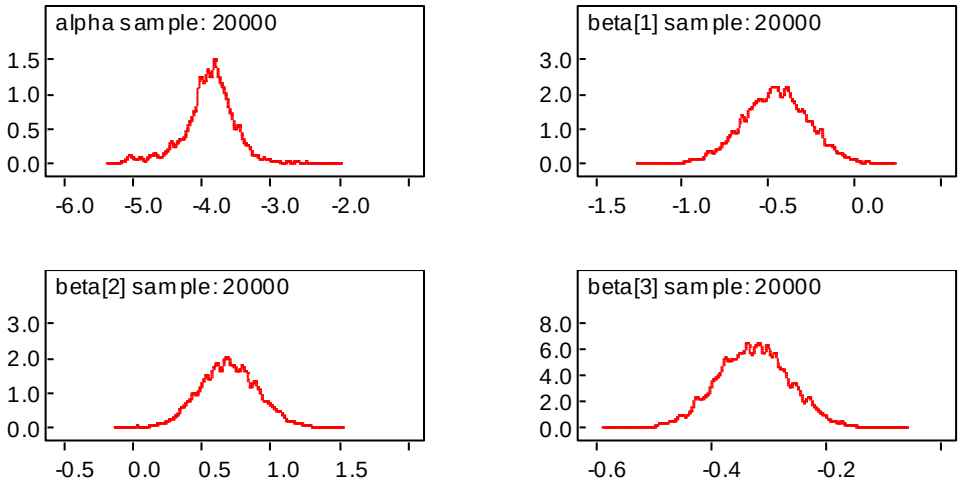

beta[7]
1.0
0.5
0.0
-0.5
5001
10000
15000
20000
25000
iteration

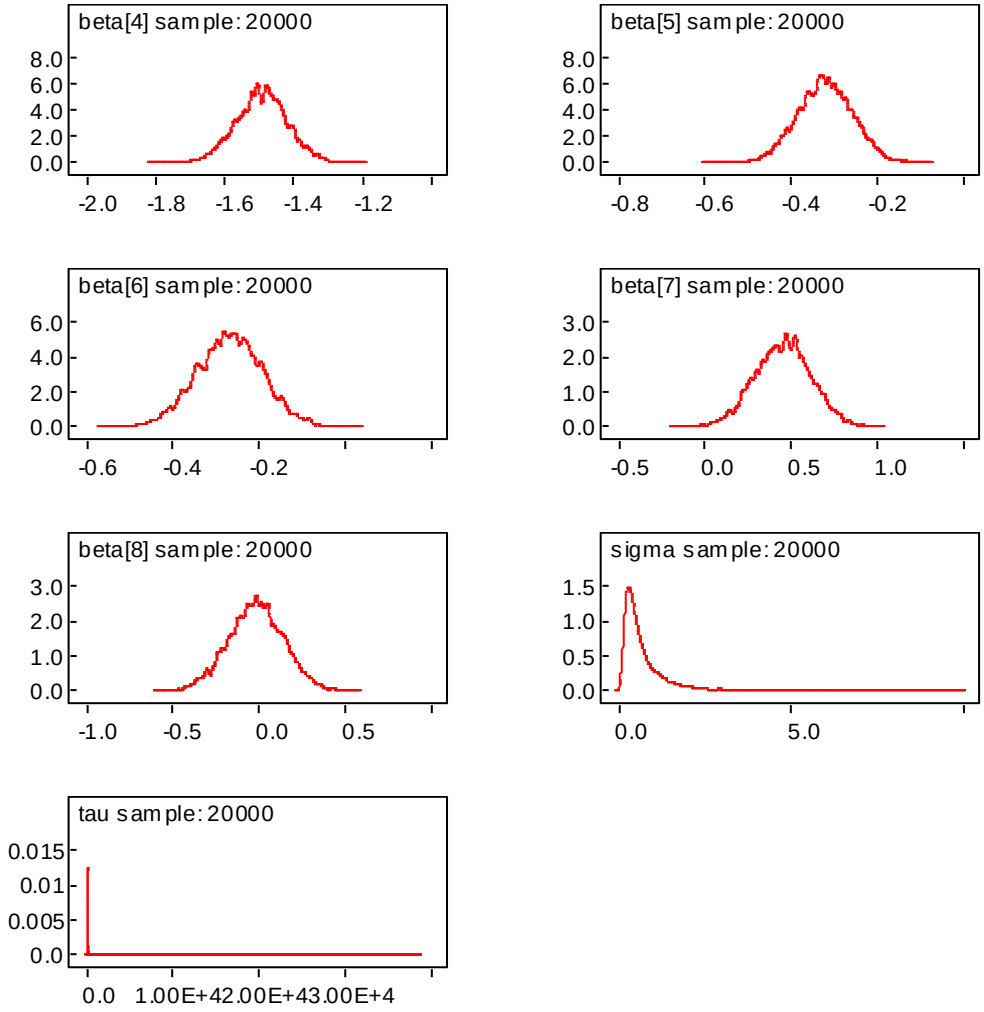

beta[8]
1.0
0.5
0.0
-0.5
-1.0
5001
10000
15000
20000
25000
iteration


sigma
10.0
7.5
5.0
2.5
0.0
5001
10000
15000
20000
25000
iteration

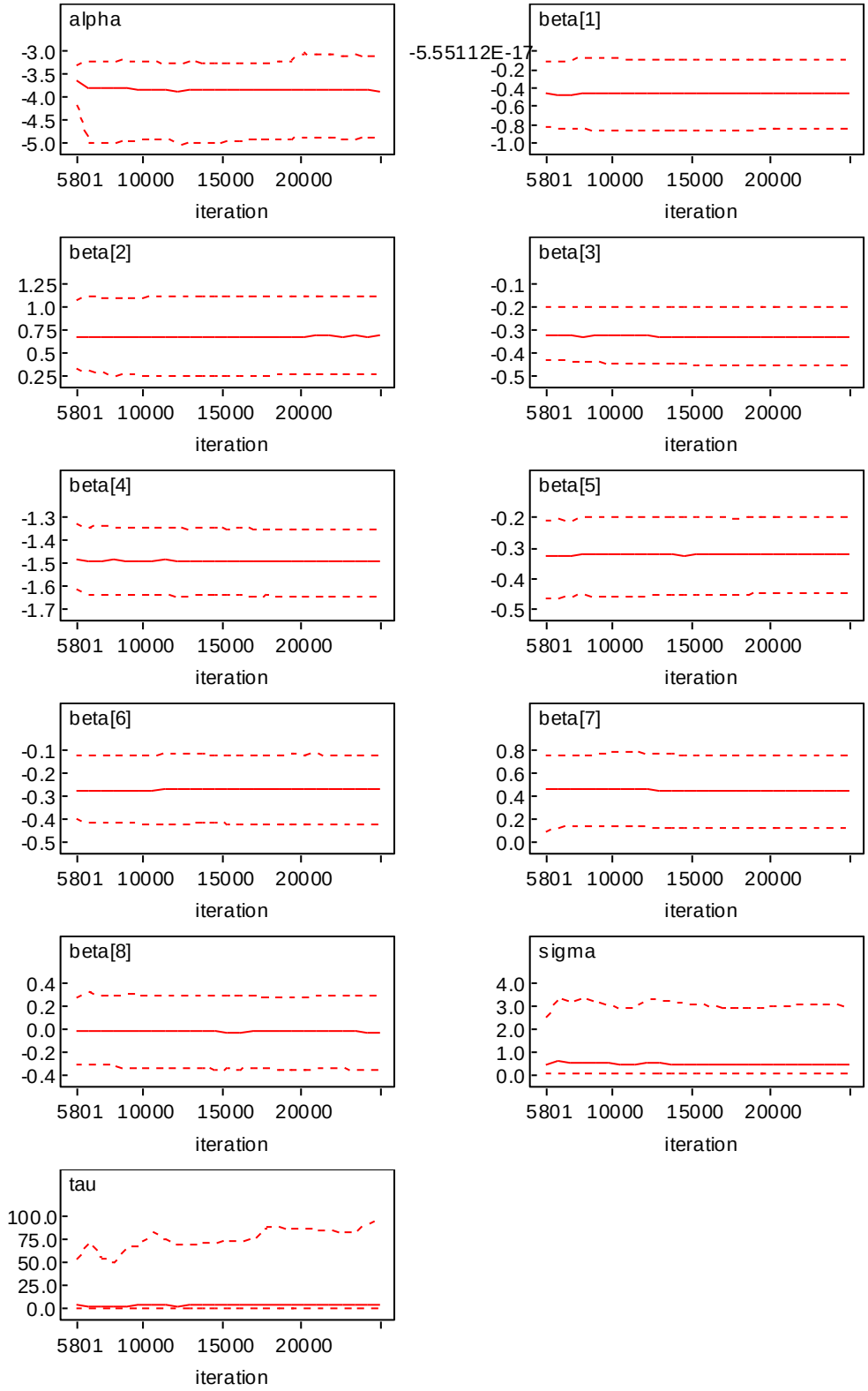

tau
4.00E+4
3.00E+4
2.00E+4
1.00E+4
0.0
5001
10000
15000
20000
25000
iteration

Density

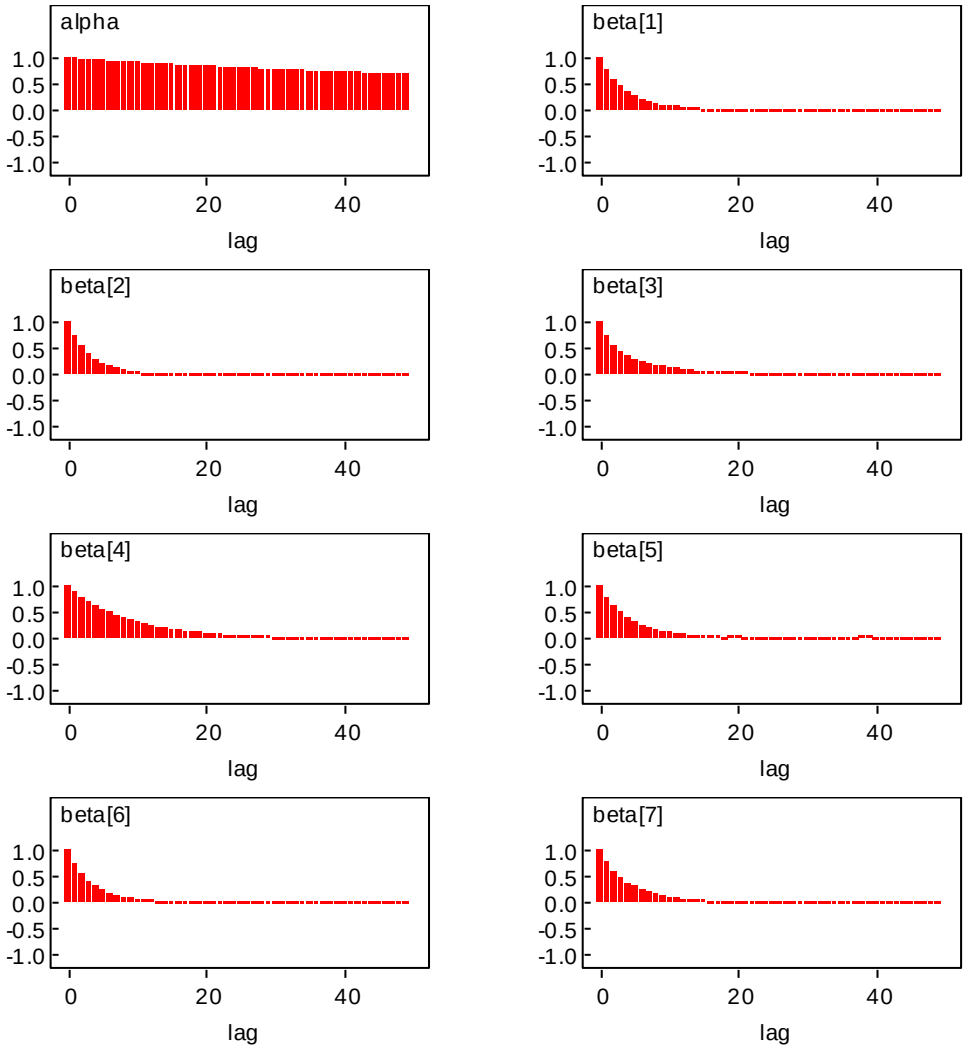

alpha sample: 20000
1.5
1.0
0.5
0.0
-6.0
-5.0
-4.0
-3.0
-2.0
beta[1] sample: 20000
3.0
2.0
1.0
0.0
-1.5
-1.0
-0.5
0.0
beta[2] sample: 20000
3.0
2.0
1.0
0.0
-0.5
0.0
0.5
1.0
1.5
beta[3] sample: 20000
8.0
6.0
4.0
2.0
0.0
-0.6
-0.4
-0.2

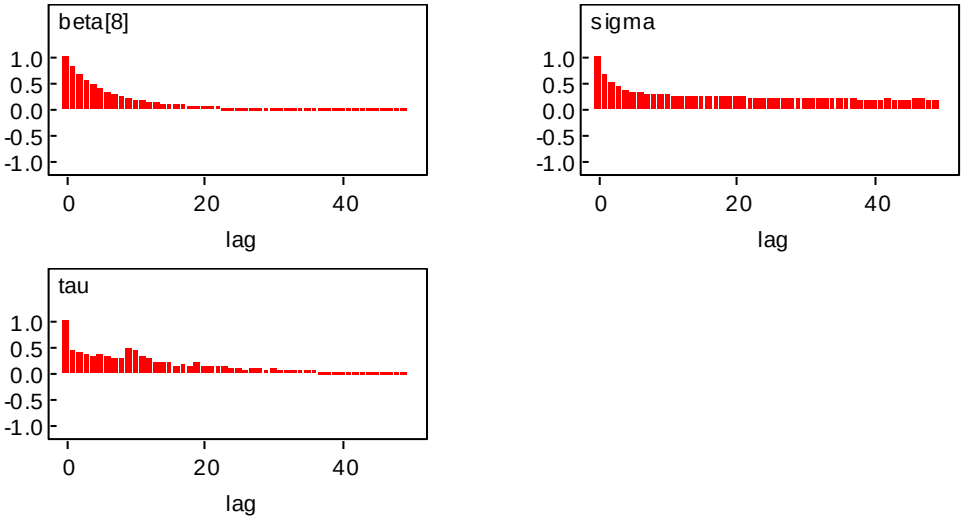

beta[4] sample: 20000
8.0
6.0
4.0
2.0
0.0
-2.0
-1.8
-1.6
-1.4
-1.2
beta[5] sample: 20000
8.0
6.0
4.0
2.0
0.0
-0.8
-0.6
-0.4
-0.2
beta[6] sample: 20000
6.0
4.0
2.0
0.0
-0.6
-0.4
-0.2
beta[7] sample: 20000
3.0
2.0
1.0
0.0
-0.5
0.0
0.5
1.0
beta[8] sample: 20000
3.0
2.0
1.0
0.0
-1.0
-0.5
0.0
0.5
sigma sample: 20000
1.5
1.0
0.5
0.0
0.0
5.0
tau sample: 20000
0.015
0.01
0.005
0.0
0.0
1.00E+4
2.00E+4
3.00E+4

Quantile


alpha
-3.0
-3.5
-4.0
-4.5
-5.0
5801 10000 15000 20000
iteration
beta[1]
-5.55112E-17
-0.2
-0.4
-0.6
-0.8
-1.0
5801 10000 15000 20000
iteration
beta[2]
1.25
1.0
0.75
0.5
0.25
5801 10000 15000 20000
iteration
beta[3]
-0.1
-0.2
-0.3
-0.4
-0.5
5801 10000 15000 20000
iteration
beta[4]
-1.3
-1.4
-1.5
-1.6
-1.7
5801 10000 15000 20000
iteration
beta[5]
-0.2
-0.3
-0.4
-0.5
5801 10000 15000 20000
iteration
beta[6]
-0.1
-0.2
-0.3
-0.4
-0.5
5801 10000 15000 20000
iteration
beta[7]
0.8
0.6
0.4
0.2
0.0
5801 10000 15000 20000
iteration
beta[8]
0.4
0.2
0.0
-0.2
-0.4
5801 10000 15000 20000
iteration
sigma
4.0
3.0
2.0
1.0
0.0
5801 10000 15000 20000
iteration
tau
100.0
75.0
50.0
25.0
0.0
5801 10000 15000 20000
iteration

Autocorrelation




beta[8]
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag
sigma
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag
tau
1.0
0.5
0.0
-0.5
-1.0
0
20
40
lag

# Appendix C Code

## C1. Code for Data Manipulations

```
%LET PATH=C:\Project\Dataset\Qing;
LIBNAME TOMIS3 "C:\Project\Dataset\Qing";
OPTIONS FMTSEARCH=(TOMIS3);

DATA TOMIS3.mq_overall_chart_cat; SET mq_overall_chart_cat;

      /*mother questionaire in hospital*/

      IF pp24 IN (1, 2)        THEN pp24m=0;
      IF pp24 IN (3, 4, 5)     THEN pp24m=1;

      IF pp25 IN (1, 2)        THEN pp25m=0;
      IF pp25 IN (3, 4, 5)     THEN pp25m=1;

      pp26m=SUM(of pp26_0 - pp26_12);

      IF pp28 IN (1, 2)        THEN pp28m=0;
      IF 3 LE pp28 LE 16       THEN pp28m=1;

      IF pp31 IN (1, 2, 3)     THEN pp31m=0;
      IF pp31 IN (4, 5, 6)     THEN pp31m=1;

      IF pp33 IN (1, 2, 3)     THEN pp33m=1;
      IF pp33 IN (4, 5, 6, 7)  THEN pp33m=0;

      IF pp34 IN (1, 2, 3)     THEN pp34m=0;
      IF 4 LE pp34 LE 8        THEN pp34m=1;

      /*6w 6m 12m phone interview questions*/

      IF gh1 IN (1, 2)         THEN gh1m=0;
      IF gh1 IN (3, 4, 5)      THEN gh1m=1;

      IF hs3 IN (1, 2)         THEN hs3m=0;
      IF hs3 IN (3, 4)         THEN hs3m=1;

      Preg_depression=eh46_2;


      IF pp8_3=1 OR eh46_1=1 OR eh46_2=1   THEN anypre_depression=1;
      IF pp8_3=0 AND eh46_1=0 AND eh46_2=0 THEN anypre_depression=0;

      IF pp8_3=1 OR eh46_1=1  THEN  hist_depression=1;
```

```
IF pp8_3=0 AND eh46_1=0 THEN  hist_depression=0;

IF bh6 IN (1, 2)        THEN bh6m=0;
IF bh6 IN (3, 4, 5)     THEN bh6m=1;

IF bh7 IN (1, 2)        THEN bh7m=0;
IF bh7 IN (3, 4)        THEN bh7m=1;

IF gh1_2 IN (1, 2)      THEN gh1_2m=0;
IF gh1_2 IN (3, 4, 5)   THEN gh1_2m=1;

wb25m=SUM(of wb25_1 - wb25_20);

IF bl47=1 AND bl48=2 AND bl49 =2          THEN w6_bladder = 0;
IF bl47 IN (2,3,4)OR bl48=1 OR bl49=1     THEN w6_bladder = 1;

IF pp14 IN (1, 2)       THEN pp14m=0;
IF pp14 IN (3, 4, 5)    THEN pp14m=1;

IF se89 IN (1, 2)       THEN se89m=0;
IF se89 IN (3, 4)       THEN se89m=1;

IF ppd="Yes"            THEN ppd_num=1;
IF ppd="No"             THEN ppd_num=0;

IF hs2 = 1              THEN hs2m = 0;
IF hs2 = 2 OR hs2 = 3   THEN hs2m = 1;

IF hs3 = 1 OR hs3 = 2   THEN hs3m = 0;
IF hs3 = 3 OR hs3 = 4   THEN hs3m = 1;

IF typevc=2             THEN typevc=0;
IF typevc=1             THEN typevc=1;

IF pp6=1                THEN pp6m=0;
IF pp6=2                THEN pp6m=1;

IF pp30=1               THEN pp30m=0;
IF pp30=2               THEN pp30m=1;

IF bh9a=1               THEN bh9am=0;
IF bh9a=2               THEN bh9am=1;

IF wb10=1               THEN wb10m=1;
IF wb10=2               THEN wb10m=0;

IF wb11 IN (1, 2, 4, 5)       THEN wb11m=0;
IF wb11 IN (3, 6, 7, 8, 9)    THEN wb11m=1;

IF se90=1                     THEN se90m=1;
IF se90=2                     THEN se90m=0;

IF se92=1                     THEN se92m=1;
IF se92=2                     THEN se92m=0;
```

```
        IF se94=1                       THEN se94m=1;
        IF se94=2                       THEN se94m=0;

        sx81m=sx81; hs1m=hs1; hs4m=hs4;

RUN;
```

## C2. Code for Primary Analysis for Depression

```
/*Demographic Statistics*/
PROC FREQ DATA= TOMIS3.mq_overall_chart_cat;
      TABLE mage*typevc pp31m*typevc pp30m*typevc
            pp28m*typevc pp34m*typevc pp33m*typevc
            pp6m*typevc/ NOROW NOCUM ;
      WHERE Timepoint=1;
RUN;

/* Multicollinearity diagnostics */
/* 1) Assess the pairwise correlations using Pearson correlation
      - same as infant health analysis */
/*2) Fit a regression model using all possible predictors and
      examine VIF, TOL, and COLLIN (in SAS)*/
PROC REG DATA=TOMIS3.mq_overall_chart_cat;
      MODEL ppd_num=mom_age pp24m pp25m pp26m pp28m pp30m pp31m
                    pp33m pp34m pp35 bh5m bh6m bh7m bh9am
                    wb10m gh1m gh1_2m pcs12 mcs12 wb25m affect_s
                    confidant_s instr_s ssqbtot w6_bladder se92m se94m
                    pp14m hs3m se89m hist_depression preg_depression
                    anypre_depression typevc  /VIF TOL COLLINOINT;
RUN; QUIT;

/*3) Remove the highly correlated variables prior to GEE analysis*/

PROC REG DATA=TOMIS3.mq_overall_chart_cat;
      MODEL ppd_num=mom_age pp24m pp25m pp26m pp28m pp30m pp31m pp33m
                    pp34m pp35 bh5m bh6m bh7m bh9am wb10m gh1m gh1_2m
                    pcs12 mcs12 wb25m ssqbtot w6_bladder se92m se94m
                    pp14m hs3m se89m hist_depression preg_depression
                    typevc    /VIF TOL COLLINOINT;
RUN; QUIT;
```

```
/*Fit GEE to evaluate main effects on depression*/

PROC GENMOD DATA=TOMIS3.mq_overall_chart_cat DESCENDING;
      CLASS CombID ;
      MODEL ppd_num=mom_age pp24m pp25m pp26m pp28m pp30m pp31m pp33m
                    pp34m pp35 bh5m bh6m bh7m bh9am w6_bladder se92m
                    wb10m gh1m gh1_2m pcs12 mcs12 wb25m ssqbtot se94m
                    pp14m hs3m se89m hist_depression preg_depression
                    typevc    / DIST=bin LINK=logit ;
      REPEATED SUBJECT=CombID /CORRW TYPE=exch ;
      OUTPUT OUT=GEE_output PRED=Predicted_value RESCHI=Residual_chi;
RUN;

/*test goodness-of-fit for full GEE model*/
/*checking Pearson chi-square and p value*/
/*QQ plot*/
PROC UNIVARIATE DATA=GEE_output;
      QQPLOT Residual_chi /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for full model of GEE";
RUN;

/*Model finalization: backward to drop item with maximum p value*/
PROC GENMOD DATA=TOMIS3.mq_overall_chart_cat DESCENDING;
      CLASS CombID ;
      MODEL ppd_num= pp33m pp28m bh5m pcs12 mcs12 ssqbtot w6_bladder
                     typevc   / DIST=bin LINK=logit LRCI ;
      REPEATED SUBJECT=CombID /CORRW TYPE=exch;
      OUTPUT OUT=GEE_final_output PRED=Predicted_value
                 RESCHI=Residual_chi;
RUN;

/*calculate marginal R2 QIC QICu for GEE*/
%INC "C:\Project\Dataset\SAS Code\SelectGEE.sas";
%SelectGEE(/*Dataset:*/ TOMIS3.mq_overall_chart_cat,
           /*Cluster:*/ CombID,
           /*Working Matrix Structure:*/ exch,
```

```
            /*Dependent Variable:*/ ppd_num,
            /*Independent Variables:*/ pp33m pp28m bh5m pcs12 mcs12
                                       ssqbtot w6_bladder typevc,
            /*Series Number:*/ 1);

 /*test goodness of fit for GEE final model*/
/*checking Pearson chi-square and p value*/
/*QQ plot*/
PROC UNIVARIATE DATA=GEE_final_output;
      QQPLOT Residual_chi /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for final model of GEE";
RUN;

/*Comparison of different var-cov structures for GEE*/

/*Unstructured*/

/*calculate QIC using macro*/

%INC "C:\Project\Dataset\SAS Code\QIC.sas";

/*Autoregrissive matrix: AR(1)*/

%QIC(CLASS=CombID,
              RESPONSE=ppd_num,
              DIST=binomial,
              SUBJECT=CombID,
              TYPE= AR(1),
              DATA=TOMIS3.mq_overall_chart_cat,
              MODEL=pp28m pp33m bh5m pcs12 mcs12 ssqbtot w6_bladder
                    typevc,
              P=pred,
              QICoptions=noprint,
              APPENDTO=summary);

/*Exchangeable matrix: Exch*/
%QIC(CLASS=CombID,
              RESPONSE=ppd_num,
              DIST=binomial,
              SUBJECT=CombID,
```

```
                TYPE= exch,
                DATA=TOMIS3.mq_overall_chart_cat,
                MODEL=pp28m pp33m bh5m pcs12 mcs12 ssqbtot w6_bladder
                      typevc,
                P=pred,
                QICoptions=noprint,
                APPENDTO=summary);

/*Unstructured matrix: un*/
%QIC(CLASS=CombID,
                RESPONSE=ppd_num,
                DIST=binomial,
                SUBJECT=CombID,
                TYPE= un,
                DATA=TOMIS3.mq_overall_chart_cat,
                MODEL=pp28m pp33m bh5m pcs12 mcs12 ssqbtot w6_bladder
                      typevc,
                P=pred,
                QICoptions=noprint,
                APPENDTO=summary);

/*Independent matrix: inde*/
%QIC(CLASS=CombID,
                RESPONSE=ppd_num,
                DIST=binomial,
                SUBJECT=CombID,
                TYPE= ind,
                DATA=TOMIS3.mq_overall_chart_cat,
                MODEL=pp28m pp33m bh5m pcs12 mcs12 ssqbtot w6_bladder
                      typevc,
                P=pred,
                QICoptions=noprint,
                APPENDTO=summary);

PROC SORT DATA=Summary;
      BY QIC;
```

```
RUN;

PROC PRINT DATA=Summary NOOBS;
      VAR LABEL QIC;
RUN;

/* Fit GLMM model*/
/*Model not converge, we use ABSPCONV=.01 to stop process*/
/*Schukken et al (2010), Correlated time to event data: Modeling
repeated clinical mastitis data from dairy cattle in New York State*/
PROC GLIMMIX DATA=TOMIS3.mq_overall_chart_cat  ABSPCONV=.01 IC=Q;
      CLASS combID ;
      MODEL ppd_num=pp28m pp33m bh5m pcs12 mcs12 ssqbtot w6_bladder
                    typevc  / DIST=binomial LINK=logit SOLUTION CL ;
      RANDOM INT /SUBJECT=combID ;
      OUTPUT OUT=glmm_output PRED=Predicted_value RESID=residual
                 VARIANCE=var;
RUN;
/*model goodness of fit: residual QQ plot*/
PROC UNIVARIATE DATA=glmm_output ;
      QQPLOT Residual /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for model of GLMM";
RUN;

/*Fit Hierarchical generalized linear model*/
PROC GLIMMIX DATA=TOMIS3.mq_overall_chart_cat NOCLPRINT ASYCOV IC=Q;
      CLASS pid1 Timepoint;
      MODEL ppd_num=pp28m pp33m bh5m pcs12 mcs12 ssqbtot w6_bladder
                    Typevc    / SOLUTION DIST=bin LINK=logit CL;
      RANDOM INT Timepoint / SUBJECT=Timepoint (pid1) TYPE=AR(1);
      OUTPUT OUT=hglm_final_output PRED=Predicted_value RESID=residual
                 VARIANCE=var;
RUN;

/*QQ plot for final model*/
PROC UNIVARIATE DATA=hglm_final_output ;
```

```
      QQPLOT Residual /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for model of HGLM";
RUN;

/*Intra-class correlation coefficient calculation*/
/*Calculate ICC using icc9 macro*/
%INC "C:\Project\Reference\ICC\icc9.sas";
%ICC9(DATA=TOMIS3.mq_overall_chart_cat,
                 VARLIST=ppd_num,
                 SUBJECT=pid1,
                 MAXDEC=4,
                 NOPRINT=T,
                 OUTDAT=ICC_output);
```

## C3. Code for Maternal Health

```
/*Checking for Multicollinearity*/
/*Dataset split using same code as infant health analysis*/
PROC REG DATA=TOMIS3.derivation_dataset_simple;
      MODEL gh1_2m=mom_age pp24m pp25m pp26m pp28m pp30m pp31m
                   pp33m pp34m pp35 bh5m bh6m bh7m bh9bm wb10m gh1m
                   ppd_num pcs12 mcs12 wb25m affect_s confidant_s
                   instr_s  ssqbtot  w6_bladder  se92m  se94m  pp14m
                   hs3m se89m hs1m hs2m hs4m bh9am wb11m sx81m
                   se90m hist_depression  preg_depression
                   anypre_depression  typevc  /VIF TOL COLLINOINT;
RUN; QUIT;

/*3) Remove the highly correlated variables prior to GEE analysis*/
PROC REG DATA=TOMIS3.derivation_dataset_simple;
      MODEL gh1_2m=mom_age  pp24m pp25m pp26m pp28m pp30m pp31m pp33m
                   pp34m pp35 bh5m bh6m bh7m bh9bm wb10m gh1m ppd_num
                   pcs12 mcs12 wb25m ssqbtot w6_bladder se92m se94m
                   pp14m hs3m se89m hs1m hs2m hs4m wb11m se90m sx81m
                   hist_depression preg_depression typevc
                   /VIF TOL COLLINOINT;
RUN; QUIT;


/*Bootstrapping procedure:
1) take a random sample, with replacement using SURVEYSELECT procedure .
2) estimate the parameters of a specified model using this resample.
3) save the parameters estimates from the resample model in a new
   dataset (model_sigs).
4) Summarize estimated dataset parms*/

%LET iter=1000;
%LET dv=gh1_2m;
%LET ivs=mom_age pp24m pp25m pp26m pp28m pp30m pp31m pp33m
```

```
         pp34m pp35 bh5m bh6m bh7m bh9bm wb10m gh1m ppd_num pcs12
         mcs12 wb25m ssqbtot w6_bladder se92m se94m pp14m hs3m se89m
         hs1m hs2m hs4m wb11m se90m sx81m  hist_depression
         preg_depression typevc;

SASFILE TOMIS3.derivation_dataset_simple LOAD;
PROC SURVEYSELECT DATA=TOMIS3.derivation_dataset_simple OUT=outdata
                  SEED=0500485 REP=&iter METHOD=URS SAMPRATE=1 OUTHITS;
RUN;
SASFILE TOMIS3.derivation_dataset_simple CLOSE;

ODS LISTING CLOSE;
ODS OUTPUT TYPE3=model_sigs;
PROC GENMOD DATA=outdata DESCENDING;
   BY replicate;
   CLASS CombID;
   MODEL &dv=&ivs / DIST=bin LINK=logit TYPE3;
RUN;
ODS OUTPUT CLOSE;
ODS LISTING;

DATA sigs;
       SET model_sigs;
       sig1=(.    < PROBCHISQ <= 0.05);
       sig2=(0.05 < PROBCHISQ <= 0.15);
       sig3=(0.15  < PROBCHISQ <= 0.25);
       sig4=(PROBCHISQ > 0.25);
       Subtotal=SUM(OF sig1-sig3);
       Proportion=Subtotal/&iter* 100;
RUN;

PROC SUMMARY DATA=sigs NOPRINT NWAY;
      CLASS source;
      OUTPUT OUT=sum_table (DROP=_TYPE_ RENAME=(_FREQ_=COUNT))
      SUM(sig1)= SUM(sig2)= SUM(sig3)= SUM(Subtotal)= SUM(Proportion)= ;
RUN;
```

```
PROC SORT DATA=sum_table;
      BY DESCENDING Subtotal;
RUN;

PROC PRINT DATA=sum_table NOOBS;
     LABEL sig1 ='Frequency of variable with p<=0.05'
           sig2 ='Frequency of variable with 0.05<p<=0.15'
           sig3 ='Frequency of variable with 0.15<p<=0.25'
           Subtotal='Frequency of variable present in bootstrap models'
           Proportion='Percentage of variable present in bootstrap
                       models';
RUN;

/*Variable selection using MarginalR2, QIC, and QICu*/
%INC "C:\Project\Dataset\SAS Code\SelectGEE.sas";
%SelectGEE(/*Dataset:*/ TOMIS3.derivation_dataset_simple,
            /*Cluster:*/ CombID,
            /*Working Matrix Structure:*/ AR(1),
            /*Dependent Variable:*/ gh1_2m,
            /*Independent Variables:*/ MCS12 PCS12 gh1m se94m pp26m
                                       TYPEVC,
            /*Series Number:*/ 1);

%SelectGEE(/*Dataset:*/ TOMIS3.derivation_dataset_simple,
            /*Cluster:*/ CombID,
            /*Working Matrix Structure:*/ AR(1),
            /*Dependent Variable:*/ gh1_2m,
            /*Independent Variables:*/ MCS12 PCS12 gh1m se94m pp26m bh5m
                                       pp34m se92m se89m hs3m TYPEVC,
                              /*Series Number:*/ 2);

%SelectGEE(/*Dataset:*/ TOMIS3.derivation_dataset_simple,
            /*Cluster:*/ CombID,
            /*Working Matrix Structure:*/ AR(1),
            /*Dependent Variable:*/ gh1_2m,
```

```
            /*Independent Variables:*/ MCS12 PCS12 gh1m se94m pp26m bh5m
                                       pp34m se92m se89m hs3m bh7m sx81m
                                       wb10m Preg_depression bh6m
                                       hist_depression TYPEVC ,
                                /*Series Number:*/ 3);

%SelectGEE(/*Dataset:*/ TOMIS3.derivation_dataset_simple,
            /*Cluster:*/ CombID,
            /*Working Matrix Structure:*/ AR(1),
            /*Dependent Variable:*/ gh1_2m,
            /*Independent Variables:*/ MCS12 PCS12 gh1m se94m pp26m bh5m
                                       pp34m se92m se89m hs3m bh7m sx81m
                                       wb10m Preg_depression bh6m
                                       hist_depression wb11m pp30m pp31m
                                       pp28m ppd_num mom_age TYPEVC,
            /*Series Number:*/ 4);

%SelectGEE(/*Dataset:*/ TOMIS3.derivation_dataset_simple,
            /*Cluster:*/ CombID,
            /*Working Matrix Structure:*/ AR(1),
            /*Dependent Variable:*/ gh1_2m,
            /*Independent Variables:*/ MCS12 PCS12 gh1m se94m pp26m bh5m
                                       pp34m se92m se89m hs3m bh7m sx81m
                                       wb10m Preg_depression bh6m
                                       hist_depression wb11m pp30m pp31m
                                       pp28m ppd_num mom_age w6_bladder
                                       hs1m bh9bm pp14m pp24m ssqbtot
                                       wb25m TYPEVC,
            /*Series Number:*/ 5);

%SelectGEE(/*Dataset:*/ TOMIS3.derivation_dataset_simple,
            /*Cluster:*/ CombID,
            /*Working Matrix Structure:*/ AR(1),
            /*Dependent Variable:*/ gh1_2m,
            /*Independent Variables:*/ MCS12 PCS12 gh1m se94m pp26m bh5m
                                       pp34m se92m se89m hs3m bh7m sx81m
```

```
                                        wb10m Preg_depression bh6m
                                        hist_depression wb11m pp30m pp31m
                                        pp28m ppd_num mom_age w6_bladder
                                        hs1m bh9bm pp14m pp24m ssqbtot
                                        wb25m pp35 pp33m hs2m pp25m se90m
                                        hs4m TYPEVC,
            /*Series Number:*/ 6);


PROC SORT Data = All;
      BY MarginalR2;

PROC PRINT Data = All Noobs;
      VAR SSE SST Xprint MarginalR2 QIC QICU;
RUN;

/*validate variables using validation_dataset*/
/*remove missing data*/
PROC SQL NOPRINT;
      CREATE TABLE derivation_dataset_simple AS
      SELECT *
      FROM TOMIS3.derivation_dataset_simple
      WHERE gh1_2m IS NOT NULL;

      CREATE TABLE validation_dataset_simple AS
      SELECT *
      FROM TOMIS3.validation_dataset_simple
      WHERE gh1_2m IS NOT NULL;
QUIT;

%INC "C:\Project\Dataset\SAS Code\ROC.sas";
%INC "C:\Project\Dataset\SAS Code\Rocplot.sas";

PROC LOGISTIC DATA=validation_dataset_simple;
      MODEL gh1_2m (EVENT='1')= MCS12 PCS12 gh1m se94m pp26m bh5m pp34m
                                se92m se89m hs3m bh7m sx81m wb10m
```

```
                                Preg_depression bh6m hist_depression
                                wb11m pp30m pp31m pp28m ppd_num mom_age
                                w6_bladder hs1m bh9bm pp14m pp24m
                                ssqbtot wb25m TYPEVC
                                / OUTROC=or ROCEPS=0 ;
      OUTPUT OUT=validation_out Xbeta=vldtn p=phat;
RUN;

PROC LOGISTIC DATA=validation_dataset_simple;
      MODEL gh1_2m (EVENT='1')= ;
      OUTPUT OUT=validation_out1 Xbeta=int;
RUN;

/*Plot ROC curve and calculate AUC*/
%ROCplot (OUT=validation_out,
          OUTROC=or,
          P=phat,
          ID=CombID,
          GRID=yes,
          MINDIST=2);

%ROC (DATA=validation_out validation_out1,
      VAR=vldtn int,
      RESPONSE= gh1_2m);

/*GEE modeling for maternal health analysis*/
PROC SQL;
      CREATE TABLE TOMIS3.maternal_GEE_dataset AS
      SELECT CombID, pid1, pid2, MCS12, PCS12, gh1m,
             se94m, pp26m, bh5m, pp34m, se92m, se89m,
             hs3m, bh7m, sx81m, wb10m, Preg_depression,
             bh6m, hist_depression, wb11m, pp30m, pp31m,
             pp28m, ppd_num, mom_age, w6_bladder, hs1m,
             bh9bm, pp14m, pp24m, ssqbtot, wb25m, TYPEVC,
             gh1_2m
      FROM TOMIS3.mq_overall_chart_cat;
```

```
QUIT;

PROC GENMOD DATA=TOMIS3.maternal_GEE_dataset DESCENDING;
      CLASS CombID;
      MODEL gh1_2m =MCS12 PCS12 gh1m se94m pp26m bh5m pp34m
                    se92m se89m hs3m bh7m sx81m wb10m
                    Preg_depression bh6m hist_depression wb11m pp30m
                    pp31m pp28m ppd_num mom_age w6_bladder hs1m
                    bh9bm pp14m pp24m ssqbtot wb25m TYPEVC
                    / DIST=bin LINK=logit;
      REPEATED SUBJECT=CombID /CORRW TYPE=AR(1);
      OUTPUT OUT=GEE_output PRED=Predicted_value RESCHI=Residual_chi;
RUN;
/*Test goodness of fit for full GEE model*/
/*Checking pearson chi-square and p value*/
/*QQ plot*/
PROC UNIVARIATE DATA=GEE_output;
      QQPLOT Residual_chi /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for full model of GEE";
RUN;

/*finalize GEE model*/
PROC GENMOD DATA=TOMIS3.maternal_GEE_dataset DESCENDING;
      CLASS CombID;
      MODEL gh1_2m =MCS12 PCS12 gh1m se94m bh6m
                    TYPEVC / DIST=bin LINK=logit;
      REPEATED SUBJECT=CombID /CORRW TYPE=AR(1);
      OUTPUT OUT=GEE_final_output PRED=Predicted_value
                                  RESCHI=Residual_chi;
RUN;

PROC UNIVARIATE DATA=GEE_final_output;
      QQPLOT Residual_chi /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for final model of GEE";
RUN;
```

## C4. Code for Infant Health

```
/*Prepare the overall dataset for maternal/infant health analysis*/

DATA t1_moth_inf  t2_moth_inf  t3_moth_inf;
      SET TOMIS3.mq_overall_chart_cat;
      IF Timepoint=1 THEN OUTPUT t1_moth_inf;
      IF Timepoint=2 THEN OUTPUT t2_moth_inf;
      IF Timepoint=3 THEN OUTPUT t3_moth_inf;
RUN;

/*Randomly draw subsets with proportion of 2/3 for derivation dataset
  and 1/3 for validation dataset*/
/*For longitudinal data, need to draw separately from each time point*/
PROC SQL;
      CREATE TABLE PatientID_set AS
      SELECT DISTINCT CombID
      FROM TOMIS3.mq_overall_chart_cat;
QUIT;

DATA derivation_ID validation_ID ;
      SET PatientID_set;
      IF RANUNI(0) LE 2/3  THEN
            OUTPUT  derivation_ID;
      ELSE OUTPUT validation_ID;
RUN;

PROC SQL;
      CREATE TABLE derivation_dataset AS
      SELECT CombID, pid1, pid2, mom_age, pp24m, pp25m, pp26m, pp28m,
             pp30m, pp31m, pp33m, pp34m,pp35, bh5m, bh6m, bh7m, bh9bm,
             wb10m, gh1m, gh1_2m, pcs12, mcs12, wb25m,affect_s,
             confidant_s, instr_s, ssqbtot, w6_bladder, se92m, se94m,
             pp14m, hs3m, se89m, hs1m, hs2m, hs4m, bh9am, wb11m,
             se90m, sx81m, hist_depression, preg_depression,
             anypre_depression, typevc, Timepoint, ppd_num
```

```
      FROM TOMIS3.mq_overall_chart_cat
      WHERE CombID IN (SELECT CombID FROM derivation_ID);

      CREATE TABLE validation_dataset AS
      SELECT CombID, pid1, pid2, mom_age, pp24m, pp25m, pp26m, pp28m,
             pp30m, pp31m, pp33m, pp34m,pp35, bh5m, bh6m, bh7m, bh9bm,
             wb10m, gh1m, gh1_2m, pcs12, mcs12, wb25m,affect_s,
             confidant_s, instr_s, ssqbtot, w6_bladder, se92m, se94m,
             pp14m, hs3m, se89m, hs1m, hs2m, hs4m, bh9am, wb11m, se90m,
             sx81m, hist_depression, preg_depression, anypre_depression,
             typevc, Timepoint, ppd_num
      FROM TOMIS3.mq_overall_chart_cat
      WHERE CombID IN (SELECT CombID FROM validation_ID);
QUIT;

/*Checking for Multicollinearity*/
/* 1) Assess the pairwise correlations using Pearson correlation
   2) Fit a regression model using all possible predictors and examine
      VIF, TOL, and COLLIN (in SAS)
   3) Remove the highly correlated variables prior to GEE analysis*/

/* 1) Assess the pairwise correlations using Pearson correlation*/
PROC CORR DATA=TOMIS3.derivation_dataset_simple OUTP=output_corr;
      VAR bh6m mom_age pp24m pp25m pp26m pp28m pp30m pp31m pp33m
          pp34m pp35 bh5m bh7m bh9bm wb10m gh1m gh1_2m ppd_num
          pcs12 mcs12 wb25m affect_s confidant_s instr_s sqbtot
          w6_bladder se92m se94m pp14m hs3m se89m hs1m hs2m hs4m
          bh9am wb11m se90m sx81m hist_depression preg_depression
          anypre_depression typevc ;
RUN;

PROC TEMPLATE;
     EDIT Base.Corr.StackedMatrix;
     COLUMN (RowName RowLabel) (Matrix) * (Matrix2);
     EDIT matrix;
     CELLSTYLE _val_  = -1.00 as {backgroundcolor=CXEEEEEE},
```

```
                _val_ <= -0.75 as {backgroundcolor=red},
                _val_ <= -0.50 as {backgroundcolor=blue},
                _val_ <= -0.25 as {backgroundcolor=cyan},
                _val_ <=  0.25 as {backgroundcolor=white},
                _val_ <=  0.50 as {backgroundcolor=cyan},
                _val_ <=  0.75 as {backgroundcolor=blue},
                _val_ <   1.00 as {backgroundcolor=red},
                _val_  =  1.00 as {backgroundcolor=CXEEEEEE};
            END;
        END;
 RUN;

ODS HTML BODY='corr.html' STYLE=statistical;
ODS LISTING CLOSE;
PROC CORR DATA=TOMIS3.derivation_dataset_simple NOPROB;
    ODS SELECT PearsonCorr;
RUN;

ODS LISTING;
ODS HTML CLOSE;
PROC TEMPLATE;
      DELETE Base.Corr.StackedMatrix;
RUN;

/*2) Fit a regression model using all possible predictors and examine
     VIF, TOL, and COLLIN (in SAS)*/
PROC REG DATA=TOMIS3.derivation_dataset_simple;
      MODEL bh6m=mom_age pp24m pp25m pp26m pp28m pp30m pp31m pp33m
                pp34m pp35 bh5m bh7m bh9bm wb10m gh1m gh1_2m  pd_num
                pcs12 mcs12 wb25m affect_s confidant_s instr_s ssqbtot
                w6_bladder se92m se94m pp14m hs3m se89m hs1m hs2m hs4m
                bh9am wb11m se90m sx81m hist_depression preg_depression
                anypre_depression typevc
                /VIF TOL COLLINOINT;
RUN; QUIT;
```

```
/*3) Remove the highly correlated variables prior to GEE analysis*/
PROC REG DATA=TOMIS3.derivation_dataset_simple;
      MODEL bh6m=mom_age pp24m pp25m pp26m pp28m pp30m pp31m pp33m

                pp34m pp35 bh5m bh7m bh9bm wb10m gh1m gh1_2m ppd_num
                pcs12 mcs12 wb25m ssqbtot w6_bladder se92m se94m pp14m
                hs3m se89m hs1m hs2m hs4m wb11m se90m sx81m
                hist_depression preg_depression typevc
                /VIF TOL COLLINOINT;
RUN; QUIT;


/*Infant health analysis*/
/*Bootstrapping procedure:
1) take a random sample, with replacement using SURVEYSELECT procedure .
2) estimate the parameters of a specified model using this resample.
3) save the parameters estimates from the resample model in a new
   dataset (model_sigs).
4) Summarize estimated dataset parms*/

%LET iter=1000;
%LET dv=bh6m;
%LET ivs=mom_age pp24m pp25m pp26m pp28m pp30m pp31m pp33m pp34m pp35
         bh5m bh7m bh9bm wb10m gh1m gh1_2m ppd_num pcs12 mcs12 wb25m
         ssqbtot  w6_bladder se92m se94m pp14m hs3m se89m hs1m hs2m hs4m
         wb11m se90m  sx81m  hist_depression  preg_depression typevc;

SASFILE TOMIS3.derivation_dataset_simple LOAD;
PROC SURVEYSELECT DATA=TOMIS3.derivation_dataset_simple
      OUT=outdata SEED=0500485
      REP=&iter METHOD=URS SAMPRATE=1 OUTHITS;
RUN;
SASFILE TOMIS3.derivation_dataset_simple CLOSE;

ODS LISTING CLOSE;
ODS OUTPUT TYPE3=model_sigs;
```

```
PROC GENMOD DATA=outdata DESCENDING;
   BY replicate;
   CLASS CombID;
   MODEL &dv=&ivs / DIST=bin LINK=logit TYPE3;
RUN;
ODS OUTPUT CLOSE;
ODS LISTING;

DATA sigs;
       SET model_sigs;
       sig1=(.    < PROBCHISQ <= 0.05);
       sig2=(0.05 < PROBCHISQ <= 0.15);
       sig3=(0.15 < PROBCHISQ <= 0.25);
       sig4=(PROBCHISQ > 0.25);
       Subtotal=SUM(OF sig1-sig3);
       Proportion=Subtotal/&iter* 100;
RUN;

PROC SUMMARY DATA=sigs NOPRINT NWAY;
      CLASS source;
      OUTPUT OUT=sum_table (DROP=_TYPE_ RENAME=(_FREQ_=COUNT))
      SUM(sig1)= SUM(sig2)= SUM(sig3)= SUM(Subtotal)= SUM(Proportion)= ;
RUN;

PROC SORT DATA=sum_table;
      BY DESCENDING Subtotal;
RUN;

PROC PRINT DATA=sum_table NOOBS;
      LABEL sig1 = 'Frequency of variable with p<=0.05'
            sig2 = 'Frequency of variable with 0.05<p<=0.15'
            sig3 = 'Frequency of variable with 0.15<p<=0.25'
            Subtotal= 'Frequency of variable present in bootstrap
                       models'
            Proportion = 'Percentage of variable present in bootstrap
                          models';
```

```
RUN;

/*Variable selection using same macro as maternal health*/
/*please maternal health analysis*/
/*Validate variables using validation_dataset*/

/*Remove missing data*/

PROC SQL NOPRINT;
      CREATE TABLE inf_derivation_dataset_simple AS
      SELECT *
      FROM TOMIS3.derivation_dataset_simple
      WHERE gh1_2m IS NOT NULL;

      CREATE TABLE inf_validation_dataset_simple AS
      SELECT *
      FROM TOMIS3.validation_dataset_simple
      WHERE gh1_2m IS NOT NULL;
QUIT;

%INC "C:\Project\Dataset\SAS Code\ROC.sas";
%INC "C:\Project\Dataset\SAS Code\ROCplot.sas";

PROC LOGISTIC DATA=inf_validation_dataset_simple;
      MODEL bh6m (EVENT='1')= pp25m gh1m se90m pp33m pp24m pp35 hs3m
                              MCS12 pp14m bh7m hs1m pp28m hs2m pp34m
                              hist_depression pp30m se89m gh1_2m bh5m
                              pp26m wb11m mom_age pp31m ppd_num PCS12
                              w6_bladder ssqbtot sx81m Preg_depression
                              bh9bm wb10m TYPEVC
                                                                    /
      OUTROC=or ROCEPS=0 ;
      OUTPUT OUT=inf_validation_out Xbeta=inf_vldtn  p=phat;
RUN;

PROC LOGISTIC DATA=inf_validation_dataset_simple;
```

```
      MODEL gh1_2m (EVENT='1')= ;
      OUTPUT OUT=inf_validation_out1 Xbeta=int;
RUN;

TITLE 'ROC Curve for Validation of Variable Selection';
%ROCplot (OUT=inf_validation_out,
          OUTROC=or,
          P=phat,
          ID=CombID,
          GRID=yes,
            MINDIST=2);

%ROC (DATA=inf_validation_out inf_validation_out1,
      VAR=inf_vldtn int,
      RESPONSE= bh6m);


/*GEE modelling for infant health analysis*/
/*Dataset for GEE modelling*/
PROC SQL;
      CREATE TABLE TOMIS3.Infant_GEE_dataset AS
      SELECT CombID, pid1, pid2, se90m, pp24m, pp25m, gh1m,
             MCS12, pp33m, hs1m, hs3m, pp14m, pp35, bh7m,
             hs2m, pp26m, pp28m, ppd_num, hist_depression,
             bh5m, pp34m, se89m, wb25m, gh1_2m, pp30m, pp31m,
             wb10m, wb11m, Preg_depression, mom_age, se94m,
             ssqbtot, bh9bm, se92m, sx81m, w6_bladder, PCS12,
             hs4m, typevc, bh6m
      FROM TOMIS3.mq_overall_chart_cat;
QUIT;

PROC GENMOD DATA=TOMIS3.Infant_GEE_dataset1 DESCENDING;
      CLASS CombID;
      MODEL bh6m=se90m pp24m pp25m gh1m MCS12 pp33m hs1m hs3m pp14m
                 pp35 bh7m hs2m pp26m pp28m ppd_num hist_depression bh5m
                 pp34m se89m wb25m gh1_2m pp30m pp31m wb10m wb11m
```

```
                Preg_depression mage se94m ssqbtot bh9bm se92m sx81m
                w6_bladder PCS12 hs4m typevc
                / DIST=bin LINK=logit;
      REPEATED SUBJECT=CombID /CORRW TYPE=AR(1);
      OUTPUT OUT=GEE_output PRED=Predicted_value RESCHI=Residual_chi;
RUN;

/*Test goodness of fit for full GEE model*/
/*Checking pearson chi-square and p value*/
/*QQ plot*/
PROC UNIVARIATE DATA=GEE_output;
      QQPLOT Residual_chi /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for full model of GEE";
RUN;

/*Finalize GEE model*/
PROC GENMOD DATA=TOMIS3.Infant_GEE_dataset1 DESCENDING;
      CLASS CombID;
      MODEL bh6m=se90m gh1m MCS12 pp28m se89m gh1_2m pp30m typevc
                  / DIST=bin LINK=logit;
      REPEATED SUBJECT=CombID /CORRW TYPE=cs;
      OUTPUT OUT=GEE_final_output PRED=Predicted_value
                                  RESCHI=Residual_chi;
RUN;

PROC UNIVARIATE DATA=GEE_final_output;
      QQPLOT Residual_chi /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for final model of GEE";
RUN;
```

## C5. Code for Missing Data Imputations

```
/*GEE model for original dataset*/
PROC GENMOD DATA=TOMIS3.mq_overall_chart_cat DESCENDING;
      CLASS CombID;
      MODEL ppd_num=pp28m pp33m bh5m pcs12 mcs12 ssqbtot w6_bladder
                    typevc / DIST=bin LINK=logit;
      REPEATED SUBJECT=CombID /CORRW TYPE=AR(1);
      OUTPUT OUT=GEE_final_output PRED=Predicted_value
                                  RESCHI=Residual_chi;
RUN;

/*test goodness of fit for GEE final model: QQ plot*/
PROC UNIVARIATE DATA=GEE_final_output;
      QQPLOT Residual_chi /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for final model of GEE";
RUN;

/*Single imputation for missing data*/
/*Dataset with variables only in final model*/
PROC SQL;
      CREATE TABLE TOMIS3.mq_overall_chart_FinMod_var AS
      SELECT CombID, pp28m, pp33m, bh5m, pcs12, mcs12, ssqbtot,
             w6_bladder, typevc, Timepoint, ppd_num
      FROM TOMIS3.mq_overall_chart_cat;
QUIT;

/*Summary of missing data*/
/*Check missing status for each variable*/
PROC MEANS data=TOMIS3.mq_overall_chart_FinMod_var
                N NMISS;
RUN;

/*Look at the number of unique values that each variable takes together
```

```
   with number of different types of missing values*/

OPTIONS NOFMTERR NOCENTER NODATE NOLABEL;
PROC FREQ DATA = TOMIS3.mq_overall_chart_FinMod_var NLEVELS;
  TABLES _ALL_ /NOPRINT MISSING;
RUN;

 /*Look at the number of missing values for each variable*/
PROC MEANS DATA = TOMIS3.mq_overall_chart_FinMod_var NMISS N;
      CLASS Timepoint;
      VAR ppd_num pp28m pp33m bh5m pcs12 mcs12 ssqbtot
          w6_bladder typevc;
RUN;

/*Calculate the proportion of missing values for each variable*/
PROC MEANS DATA = TOMIS3.mq_overall_chart_FinMod_var NMISS;
  VAR ppd_num pp28m pp33m bh5m pcs12 mcs12 ssqbtot
      w6_bladder typevc;
  OUTPUT OUT=T (DROP=_TYPE_ _FREQ_) NMISS=/AUTONAME;
RUN;

PROC TRANSPOSE DATA = T PREFIX=NMISS OUT=S1;
  VAR _NUMERIC_;
RUN;

/*Calculate missing proportion*/
DATA S2; SET S1; PMISS = NMISS1/7680*100; RUN;
PROC PRINT DATA = S2; RUN;

/*Explore the pattern of missing values*/
ODS SELECT MISSPATTERN;
ODS Listing Close;
ODS HTML body='ods-body.htm';
PROC MI DATA = TOMIS3.mq_overall_chart_FinMod_var NIMPUTE=0;
  VAR ppd_num pp28m pp33m bh5m pcs12 mcs12 ssqbtot
      w6_bladder typevc Timepoint ;
```

```
RUN;
ODS HTML Close;
ODS Listing;

/*LOCF imputation*/
%INC "C \Project\Dataset\SAS Code\Missing imputation\locf.sas";
%INC "C:\Project\Dataset\SAS Code\Missing imputation\words.sas";
%INC "C:\Project\Dataset\SAS Code\Missing imputation\commas.sas";
%INC "C:\Project\Dataset\SAS Code\Missing imputation\vartype.sas";
%INC "C:\Project\Dataset\SAS Code\Missing imputation\quotelst.sas";
%INC "C:\Project\Dataset\SAS Code\Missing imputation\attrv.sas";

%LOCF(DSIN=TOMIS3.mq_overall_chart_FinMod_var,
      DSOUT=LOCF_FinMod_var,
      VARS=pp28m pp33m bh5m pcs12 mcs12 ssqbtot
      w6_bladder typevc ppd_num,
      BYGROUP=CombID,
      VISITVARS=Timepoint);

PROC MEANS data=LOCF_FinMod_var N NMISS;
RUN;

/*GEE model for LOCF imputation dataset*/
PROC GENMOD DATA=LOCF_FinMod_var  DESCENDING;
      CLASS CombID;
      MODEL ppd_num=pp28m pp33m bh5m pcs12 mcs12 ssqbtot
                    w6_bladder typevc / DIST=bin LINK=logit;
      REPEATED SUBJECT=CombID /CORRW TYPE=AR(1);
      OUTPUT OUT=LOCF_GEE_final_output PRED=Predicted_value
                                       RESCHI=Residual_chi;
RUN;

PROC UNIVARIATE DATA=LOCF_GEE_final_output;
      QQPLOT Residual_chi /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for LOCF Imputation";
RUN;
```

```
/*Mean imputation*/
PROC SQL;
      CREATE TABLE Mean_FinMod_var AS
      SELECT CombID, Timepoint,
            CASE  pp28m
            WHEN . THEN ROUND(MEAN(pp28m)) ELSE pp28m
            END AS pp28m,

            CASE pp33m
            WHEN . THEN ROUND(MEAN(pp33m)) ELSE pp33m
            END AS pp33m,

            CASE w6_bladder
            WHEN . THEN ROUND(MEAN(w6_bladder)) ELSE w6_bladder
            END AS w6_bladder,

            CASE TYPEVC
            WHEN . THEN ROUND(MEAN(TYPEVC)) ELSE TYPEVC
            END AS TYPEVC,

            CASE ppd_num
            WHEN . THEN ROUND(MEAN(ppd_num)) ELSE ppd_num
            END AS ppd_num,

            CASE bh5m
            WHEN . THEN MEAN(bh5m) ELSE bh5m
            END AS bh5m,

            CASE pcs12
            WHEN . THEN MEAN(pcs12) ELSE pcs12
            END AS pcs12,

            CASE mcs12
            WHEN . THEN MEAN(mcs12) ELSE mcs12
            END AS mcs12,
```

```
            CASE ssqbtot
            WHEN . THEN MEAN(ssqbtot) ELSE ssqbtot
            END AS ssqbtot

      FROM TOMIS3.mq_overall_chart_FinMod_var
      GROUP BY CombID;
QUIT;

PROC MEANS data=Mean_FinMod_var N NMISS;
RUN;
PROC MEANS data=TOMIS3.mq_overall_chart_FinMod_var N NMISS;
RUN;

/*GEE model for mean imputation dataset*/
PROC GENMOD DATA=Mean_FinMod_var  DESCENDING;
      CLASS CombID;
      MODEL ppd_num=pp28m pp33m bh5m pcs12 mcs12 ssqbtot
                  w6_bladder typevc / DIST=bin LINK=logit;
      REPEATED SUBJECT=CombID /CORRW TYPE=AR(1);
      OUTPUT OUT=Mean_GEE_final_output PRED=Predicted_value
                                       RESCHI=Residual_chi;
RUN;

PROC UNIVARIATE DATA=Mean_GEE_final_output;
      QQPLOT Residual_chi /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for Mean Imputation";
RUN;


/* Hot-deck imputation*/
DATA INFILE; SET TOMIS3.mq_overall_chart_FinMod_var; RUN;

%INC 'C:\Project\Dataset\SAS Code\Missing imputation\hotdeck.sas';

%LET INFILE =mq_overall_chart_FinMod_var;
```

```
%LET OUTFILE =Hotdeck_FinMod_var;
%LET ID = CombID;
%LET RESPONSE =NONE;

%HOTDECK(
/*place variable(s) to be imputed here===>*/ pp28m pp33m bh5m pcs12
                              mcs12 ssqbtot w6_bladder typevc ppd_num,
/*CLASSING or POST-STRATA variable(s)=>*/ Timepoint ,
/*place sorting variable(s) here========>*/ CombID ,
/*select method: PREV, NEXT, or BOTH===>*/ prev,
/*list value(s) to treat as missing here====>*/ . ,
/*put 1 to impute all, 0 just the missing===>*/ 0
);
DATA Hotdeck_FinMod_var;
      SET LIB.Hotdeck_FinMod_var;
      KEEP CombID pp28m pp33m bh5m pcs12 mcs12 ssqbtot
           w6_bladder typevc ppd_num;

PROC MEANS data=Hotdeck_FinMod_var N NMISS;
RUN;

/*GEE model for hot deck imputation dataset*/
PROC GENMOD DATA=Hotdeck_FinMod_var  DESCENDING;
      CLASS CombID;
      MODEL ppd_num=pp28m pp33m bh5m pcs12 mcs12 ssqbtot
                    w6_bladder typevc  / DIST=bin LINK=logit;
      REPEATED SUBJECT=CombID /CORRW TYPE=AR(1);
      OUTPUT OUT=Hotdeck_GEE_final_output PRED=Predicted_value
                                          RESCHI=Residual_chi;
RUN;

PROC UNIVARIATE DATA=Hotdeck_GEE_final_output;
      QQPLOT Residual_chi /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for Hot-deck Imputation";
RUN;
```

```
/*Multiple imputation*/
PROC MEANS data=TOMIS3.mq_overall_chart_FinMod_var
           N NMISS MIN MAX MEAN STD;
RUN;

/*Format dataset to one subject in one row*/
DATA mq_overall_chart_FinMod_var;
      SET TOMIS3.mq_overall_chart_FinMod_var;
RUN;

PROC SORT DATA =mq_overall_chart_FinMod_var;
  BY CombID Timepoint;
RUN;

DATA Formatted_data;
  SET mq_overall_chart_FinMod_var;
      ARRAY ppd_num_t(3);     ARRAY pp28m_t(3);       ARRAY pp33m_t(3);
      ARRAY bh5m_t(3);        ARRAY pcs12_t(3);       ARRAY mcs12_t(3);
      ARRAY ssqbtot_t(3);     ARRAY w6_bladder_t(3);  ARRAY typevc_t(3);
  BY CombID;

RETAIN ppd_num_t pp28m_t pp33m_t bh5m_t pcs12_t mcs12_t
         ssqbtot_t w6_bladder_t typevc_t;

IF FIRST.CombID THEN
DO i = 1 TO 3;
      ppd_num_t(i)=.;   pp28m_t(i)=.; pp33m_t(i)=.;   bh5m_t(i)=.;
      pcs12_t(i)=.;     mcs12_t(i)=.; ssqbtot_t(i)=.; w6_bladder_t(i)=.;
      typevc_t(i)=.;
END;

ppd_num_t(Timepoint)=ppd_num; pp28m_t(Timepoint)=pp28m;
pp33m_t(Timepoint)=pp33m;     bh5m_t(Timepoint)=bh5m;
pcs12_t(Timepoint)=pcs12;     mcs12_t(Timepoint)=mcs12;
ssqbtot_t(Timepoint)=ssqbtot; w6_bladder_t(Timepoint)=w6_bladder;
typevc_t(Timepoint)=typevc;
```

```
IF LAST.CombID;
  DROP ppd_num pcs12 mcs12 bh5m w6_bladder
       pp28m pp33m typevc ssqbtot Timepoint i;
RUN;

PROC PRINT DATA = Formatted_data (OBS=10) NOOBS;
RUN;

/*Explore the pattern of missing values*/
ODS SELECT MISSPATTERN;
PROC MI DATA = Formatted_data OUT=Formatted_data_imputed;
  VAR ppd_num_t: pp28m_t: pp33m_t: bh5m_t: pcs12_t: mcs12_t:
      ssqbtot_t: w6_bladder_t: typevc_t: ;
RUN;

/*Multiple imputation*/
PROC MI DATA = Formatted_data OUT=Formatted_data_imputed_out
               NIMPUTE=5
               SEED=0500485;

  VAR ppd_num_t:  pp28m_t:    pp33m_t: bh5m_t:  pcs12_t:
      mcs12_t:    ssqbtot_t:  w6_bladder_t:     typevc_t:;
  MCMC ACFPLOT NBITER=1000;
RUN;

/*Turn dataset back to longitudinal format*/
DATA Long_format_imputed;
      SET Formatted_data_imputed_out;

      ARRAY ppd_num_t(3) ppd_num_t:;ARRAY pp28m_t(3) pp28m_t:;
      ARRAY pp33m_t(3) pp33m_t:;    ARRAY pcs12_t(3) pcs12_t:;
      ARRAY mcs12_t(3) mcs12_t:;    ARRAY ssqbtot_t(3) ssqbtot_t:;
      ARRAY bh5m_t(3) bh5m_t:;      ARRAY w6_bladder_t(3) w6_bladder_t:;
      ARRAY typevc_t (3) typevc_t:;
```

```
      DO Timepoint = 1 TO 3;
         ppd_num = ppd_num_t(Timepoint);  pp28m = pp28m_t(Timepoint);
         pp33m = pp33m_t(Timepoint);      bh5m= bh5m_t(Timepoint);
         pcs12 = pcs12_t(Timepoint);      mcs12 = mcs12_t(Timepoint);
         ssqbtot=ssqbtot_t(Timepoint);    typevc = typevc_t(Timepoint);
         w6_bladder= w6_bladder_t(Timepoint);
         OUTPUT;
      END;

      DROP ppd_num_t: pcs12_t:   mcs12_t:   pp28m_t:
           pp33m_t:   typevc_t : bh5m_t: ;  w6_bladder_t:
RUN;

DATA Long_format_imputed;
      SET Long_format_imputed;
      IF ppd_num LT 0.5 THEN ppd_num = 0;
      ELSE ppd_num = 1;
RUN;

PROC SORT DATA=Long_format_imputed;
      BY _imputation_;
RUN;

/*fit GEE for complete dataset BY imputation*/
PROC GENMOD DATA=Long_format_imputed DESCENDING;
      BY _imputation_;
      CLASS CombID;
      MODEL ppd_num=pp28m pp33m bh5m pcs12 mcs12 ssqbtot w6_bladder
                    typevc / DIST=bin LINK=logit;
      REPEATED SUBJECT=CombID /MCovB TYPE=AR(1);
      ODS OUTPUT ParameterEstimates=gmparms
                 ParmInfo=gmpinfo
                 GEENCov=gmcovb;
      OUTPUT OUT=MI_GEE_final_output PRED=Predicted_value
                                     RESCHI=Residual_chi;
RUN;
```

```
/*Combine analysis results*/
PROC MIANALYZE PARMS=gmparms COVB=gmcovb PARMINFO=gmpinfo;
     MODELEFFECTS intercept pp28m pp33m bh5m pcs12 mcs12 ssqbtot
                  w6_bladder typevc;
RUN;

/*Q-Q plot for MI*/
PROC UNIVARIATE DATA=MI_GEE_final_output;
      QQPLOT Residual_chi /NORMAL (MU=est SIGMA=est COLOR=blue);
      TITLE "Q-Q Plot for MI";
RUN;
```

## C6. Code for Bootstrap Model Validation

```
/*Model validation using bootstrap method*/
/*The procedures are as follows
1) Draw B samples x'(1), x'(2), . . . ,x'(B) with replacement from
   original data x
2) Fit regression model to each bootstrap sample x' and examine
   distribution of estimates of parameters
3) Average estimates, draw ROC and calculate AUC */


/*Prepare dataset for validation*/
PROC SQL;
      CREATE TABLE TOMIS3.Bootstrap_validation_dataset AS
      SELECT CombID, mom_age, pp24m, pp25m, pp26m, pp28m, pp30m, pp31m,
             pp33m, pp34m, pp35, bh5m, bh6m, bh7m, bh9am, wb10m, gh1m,
             gh1_2m, pcs12, mcs12, wb25m,ssqbtot, w6_bladder, se92m,
             se94m, pp14m, hs3m, se89m, hist_depression,
             preg_depression, typevc, ppd_num
      FROM TOMIS3.mq_overall_chart_cat;
QUIT;

/*Bootstrapping: draw 200 resempling datasets with replacement*/
%LET rep=200;
%LET dv=ppd_num;
%LET ivs=pp28m pp33m bh5m pcs12 mcs12 ssqbtot w6_bladder;

SASFILE TOMIS3.Bootstrap_validation_dataset LOAD;
PROC SURVEYSELECT DATA=TOMIS3.Bootstrap_validation_dataset
                  OUT=bootstrap_outdata SEED=0500485
                  REP=&rep METHOD=URS SAMPRATE=1 OUTHITS;
RUN;
SASFILE TOMIS3.Bootstrap_validation_dataset CLOSE;
```

```
ODS LISTING CLOSE;

/*GEE for bootstrapped dataset*/
PROC GENMOD DATA=bootstrap_outdata DESCENDING;
   BY replicate;
   CLASS CombID;
   MODEL &dv=&ivs / DIST=bin LINK=logit ;
   REPEATED SUBJECT=CombID /MCovB TYPE=AR(1);
   ODS OUTPUT ParameterEstimates=gmparms;
RUN;

/*GEE for original dataset*/
PROC GENMOD DATA=TOMIS3.Bootstrap_validation_dataset DESCENDING;
   CLASS CombID;
   MODEL &dv=&ivs / DIST=bin LINK=logit ;
   REPEATED SUBJECT=CombID /CORRW TYPE=AR(1);
   OUTPUT OUT=GEE_output_or P=p_hat ;
RUN;

PROC SQL;
      CREATE TABLE GEE_Bootvalid AS
      SELECT *
      FROM TOMIS3.gmparms
      WHERE Parameter NOT IN ("Intercept", "Scale");
QUIT;

/* Shapiro-Wilk W test for normality*/
PROC UNIVARIATE DATA=GEE_Bootvalid NORMAL PLOT;
      CLASS Parameter;
      QQPLOT Estimate /NORMAL(MU=EST SIGMA=EST COLOR=RED L=1);
RUN;
PROC NPAR1WAY DATA=GEE_Bootvalid wilcoxon edf ;
      CLASS Parameter;
      VAR Estimate;
      Exact;
```

```
 run;


/********GEE output=>ROC*********/

PROC SQL;
      CREATE TABLE Auc
      LIKE TOMIS3.Auc_boot_valid;
QUIT;

%MACRO Auc_calc;

%DO j=1 %TO &iter;
      PROC SQL;
            CREATE TABLE pred&j AS
                  SELECT *, ppd_num AS OBS
                  FROM GEE_output_or
                  WHERE Replicate = &j;

            CREATE TABLE data0&j AS
                  SELECT *
                  FROM pred&j
                  WHERE obs=0;
            CREATE TABLE data1&j AS
                  SELECT *, 0 AS Rsum1
                  FROM pred&j
                  WHERE obs=1;

            CREATE TABLE temp&j AS
                  SELECT *, 0 AS ROC1
                  FROM pred&j;

            CREATE TABLE out&j AS
                  SELECT ROC1
                  FROM temp&j;
      QUIT;
```

```
/*********Get ROC curve*************/;
      DATA yy&j;
      SET pred&j;
            DO i=1 TO 200;
                  IF p_hat>0.005*i THEN y=1;
                  ELSE IF .z<p_hat<0.005*i THEN y=0;
                  ELSE y=.;
                  OUTPUT;
            END;
      RUN;
      PROC SORT DATA=yy&j; BY i; RUN;

/********Get sensitivity and specificity*************/;
      PROC FREQ DATA=yy&j;
            TABLES y*obs/NOPRINT OUT=pct&j OUTPCT;
            BY i;
      RUN;

      DATA sen&j;
            SET pct&j;
            IF y=0 and obs=0;
                        sensi=PCT_COL;
                        cut=0.005*i;

      DATA spc&j;
            SET pct&j;
            IF y=1 and obs=1;
                        speci=PCT_COL;
                        cut=0.005*i;
      RUN;

      DATA curve1&j;
            MERGE sen&j spc&j;
            BY cut;
            _spc=100-speci;
```

```
      RUN;

      DATA curve&j;
            SET curve1&j;
            sensi1=sensi/100;
            _spc1=_spc/100;
      RUN;
      PROC SORT DATA=curve&j;
      BY _spc1;
      RUN;

/*Area under ROC curve*/
      DATA auc&j;
      SET curve&j end=eof;
                  Replicate=&j; DROP j;
                  lagx=lag(_spc1);
                  lagy=lag(sensi1);
            IF order=1 THEN DO;
                  lagx=0;
                  lagy=0;
            END;
      tpzd=(_spc1-lagx)*(sensi1+lagy)/2;
      sumtpz+tpzd;

      IF eof THEN DO;
            roc_auc=sumtpz+(1-_spc1)*(sensi1+1)/2;;
            OUTPUT;
      END;
      RUN;

      PROC APPEND DATA=auc&j BASE=Auc; RUN;
%END;
%MEND;

%Auc_calc;
```

```
DATA Auc_bootstrap;
      SET TOMIS3.auc;
      KEEP Replicate Roc_auc;
      IF Replicate IN (101,198) THEN DELETE;

      IF 0.9=<Roc_auc<1        THEN Class='Excellent';
      IF 0.8=<Roc_auc<0.9      THEN Class='Good';
      IF 0.7=<Roc_auc<0.8      THEN Class='Worthless';
      IF Roc_auc<0.7           THEN Class='Bad';

RUN;

PROC MEANS DATA=Auc_bootstrap N MEAN MIN MAX CLM;
      VAR roc_auc;
RUN;

TITLE1 'Bootstrap validation summary';

PROC UNIVARIATE DATA=Auc_bootstrap;
      VAR Roc_auc;
      HISTOGRAM ;
RUN;
```

## C7. Code for Forest Plots

```
/* SET THE GRAPHICS ENVIRONMENT */
GOPTIONS RESET=all CBACK=white BORDER HTITLE=12pt HTEXT=10pt;

/*Missing data*/
%LET dataset=missing_forest_plot;
%LET a=0;
%LET b=2;
%LET c=0.2;
%LET variable=Missing_Data

/*Different models*/
%LET dataset=PPD_GEE_forest_plot;
%LET a=0;
%LET b=3.5;
%LET c=0.5;
%LET variable=GEE;

%LET dataset=PPD_GLMM_forest_plot;
%LET a=0;
%LET b=3.5;
%LET c=0.5;
%LET variable=GLMM;

%LET dataset=PPD_HGLM_forest_plot;
%LET a=0;
%LET b=3.5;
%LET c=0.5;
%LET variable=HGLM;

%LET dataset=PPD_Bayesian_forest_plot;
%LET a=0;
%LET b=3.5;
```

```
%LET c=0.5;
%LET variable=Bayesian;

/*Model comparison*/
%LET dataset=PPD_comparison_forest_plot;
%LET a=0.5;
%LET b=1.5;
%LET c=0.5;
%LET variable=Model_Comparisan;

/*Bayesian sensitivity analysis*/
%LET dataset=PPD_Bayesian_Sensitivity_forest_plot;
%LET a=0;
%LET b=2;
%LET c=0.5;
%LET variable=Sensitivity;

/*Maternal health analysis*/
%LET dataset=Maternal_GEE_forest_plot;
%LET a=0;
%LET b=9;
%LET c=3;
%LET variable=Maternal;

/*Infant health analysis*/
%LET dataset=Infant_GEE_forest_plot;
%LET a=0;
%LET b=4;
%LET c=2;
%LET variable=Infant;

/*Summary of estimates of variables at different models */
%LET dataset=Summary_forest_plot;
%LET a=0;
%LET b=3.5;
%LET c=0.5;
```

```
%LET variable=pp28m;          %LET variable=pp33m;
%LET variable=bh5m;           %LET variable=PCS12;
%LET variable=MCS12;          %LET variable=ssqbtot;
%LET variable=w6_bladder;     %LET variable=TYPEVC;

 /*Summary of estimates of variables at different imputation methods */
%LET dataset=MiSummary_forest_plot;
%LET a=0;
%LET b=3;
%LET c=0.5;
%LET variable=pp28m;          %LET variable=pp33m;
%LET variable=bh5m;           %LET variable=PCS12;
%LET variable=MCS12;          %LET variable=ssqbtot;
%LET variable=w6_bladder;     %LET variable=TYPEVC;


PROC IMPORT OUT=&dataset REPLACE
              DATAFILE="C:\Forest plot data\&dataset..csv";
              GETNAMES=YES;
RUN;

/*The following DATA SET step is ONLY for Summary of
      estimates of variables at different models*/
%PUT &variable;

DATA  &dataset;
      SET &dataset;
      IF Variable_name="&variable" THEN OUTPUT  &dataset;
RUN;

/* CREATE AN ANNOTATE DATA SET TO DRAW THE LINES. */
DATA ANNO;
   LENGTH FUNCTION STYLE COLOR $8;
   RETAIN XSYS YSYS '2' WHEN 'A';
   SET &dataset;
/* DRAW THE HORIZONTAL LINE FROM LOWER_LIMIT TO UPPER_LIMIT */
```

```
   FUNCTION='MOVE'; XSYS='2'; YSYS='2'; YC=YVAR; X=LOWER_LIMIT;
   COLOR='BLACK'; OUTPUT;
   FUNCTION='DRAW'; X=UPPER_LIMIT; COLOR='BLACK'; SIZE=1; OUTPUT;

/* DRAW THE TICK LINE FOR THE LOWER_LIMIT VALUE */
   FUNCTION='MOVE';XSYS='2'; YSYS='2';YC=YVAR; X=LOWER_LIMIT;
   COLOR='BLACK'; OUTPUT;
   FUNCTION='DRAW';X=LOWER_LIMIT; YSYS='9'; Y=+1; SIZE=1; OUTPUT;
   FUNCTION='DRAW';X=LOWER_LIMIT; Y=-2; SIZE=1;OUTPUT;

   /* DRAW THE TICK LINE FOR THE UPPER_LIMIT VALUE */
   FUNCTION='MOVE';XSYS='2'; YSYS='2'; YC=YVAR; X=UPPER_LIMIT;
   COLOR='BLACK'; OUTPUT;
   FUNCTION='DRAW';X=UPPER_LIMIT; YSYS='9'; Y=+1; SIZE=1; OUTPUT;
   FUNCTION='DRAW';X=UPPER_LIMIT; Y=-2; SIZE=1; OUTPUT;
RUN;

TITLE1 "Forest Plot for &variable";
AXIS1 LABEL=NONE
      MINOR=NONE
      OFFSET=(5,5)
      ORDER=('Bayesian' 'HGLM' 'GLMM' 'GEE'); /*for model comparison*/
      ORDER=('MI' 'Hot-Deck' 'LOCF' 'Mean'); /*for missing imputation*/

AXIS2 ORDER=(&a TO &b BY &c)
      LABEL=('Odds Ratio')
      MINOR=NONE;

SYMBOL1 INTERPOL=NONE COLOR=BLACK VALUE=DOT HEIGHT=0.5;

PROC GPLOT DATA=&dataset;
   PLOT YVAR*RATE / ANNOTATE=ANNO
        NOLEGEND
        VAXIS=AXIS1
        HAXIS=AXIS2
        HREF = 1
```

```
        LHREF = 2;
RUN;
QUIT;
```

## C8. Code for Bayesian Analysis (WinBUGS)

```
/* Output 'TOMIS3.mq_overall_chart_cat' to excel format for WinBUGS*/
DATA TOMIS3.mq_overall_chart_cat_simple;
      SET TOMIS3.mq_overall_chart_cat;
      KEEP pp28m pp33m bh5m pcs12 mcs12 ssqbtot w6_bladder typevc
           Timepoint ppd_num;
RUN;

PROC SQL;
      DROP TABLE TOMIS3.mq_overal_simple_no_missing;
      CREATE TABLE TOMIS3.mq_overal_simple_no_missing AS
      SELECT *
      FROM TOMIS3.mq_overall_chart_cat_simple
      WHERE pp28m is NOT NULL AND         pp33m is NOT NULL AND
            bh5m is NOT NULL  AND         mcs12 is NOT NULL AND
            pcs12 is NOT NULL AND         ssqbtot is NOT NULL AND
            w6_bladder is NOT NULL AND    typevc is NOT NULL AND
            Timepoint is NOT NULL AND     ppd_num is NOT NULL;
QUIT;

/*Standardize continuous variables for WinBUGS dataset*/
PROC STANDARD DATA=TOMIS3.mq_overal_simple_no_missing
              OUT=TOMIS3.mq_overal_simple_no_missing_sd
              MEAN=0 STD=1;
      VAR bh5m pcs12 mcs12 ssqbtot;
RUN;

PROC MEANS DATA=TOMIS3.mq_overal_simple_no_missing_sd;
RUN;
```

**WinBUGS code for Bayesian analysis**

```
model {
            # N observations
            for (i in 1:N) {
                  y[i] ~ dbern(p2[i])
                  logit(p[i]) <- alpha + inprod(beta[],x[i,1:R]) +
                                 u[x[i,R+1]]
                  p2[i]<-max(0.00001,min(0.99999, p[i]))
            }
            # M timepoints
            for (j in 1:M) {u[j] ~ dnorm (0,tau)}

            # R variable numbers
            for (k in 1:R){beta[k] ~ dnorm(0.0,1.0E-6)}

            # Hyperprior
            # tau ~ dgamma(0.001,0.001)
            # tau ~ dgamma(0.01,0.01)
            # tau ~ dgamma(0.1,0.1)

            # Priors
            alpha ~ dnorm(0.0,1.0E-6)
            tau <- 1/(sigma*sigma)
            sigma ~ dunif(0, 10)
            # sigma ~ dunif(0, 5)
            # sigma ~ dunif(0, 15)
            # sigma ~ dunif(0, 20)
            # sigma ~ dunif(0, 25)
            # sigma ~ dunif(0, 50)
}

# initial value
list(alpha=10.57, sigma=1,
     beta=c(-0.05, -0.19, -0.12, 0.51, 0.47, 0.04))
```

```
# data loading
list(N=1956, M=3, R=6)
```