Bayesian Inference of Joint Multiple Longitudinal Outcomes: Analysis of Hypertension Data

By

Azeez Adeboye

Department of Statistics (Biostatistics), University of Fort Hare, Alice, Eastern Cape, South Africa

<u>Join work with</u>

Prof Yongsong Qin,

Dr James Ndege

Dr Ruffin Mutambayi

40th Annual Conference of the International Society for Clinical Biostatistics (ISCB40), Leuven, Belgium from14-18 July 2019

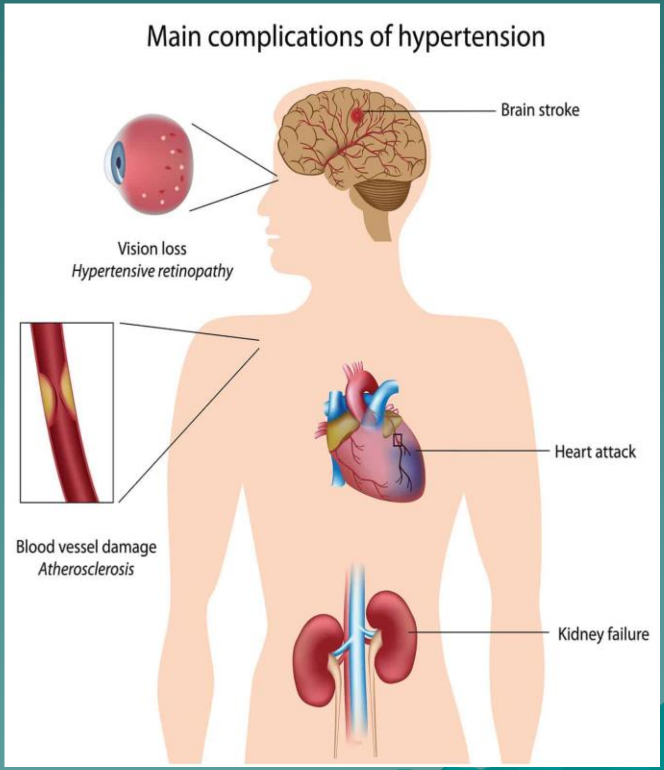
Contents

Assumed knowledge

- Longitudinal data analysis
- Models for Longitudinal data
  - ✓ Linear mixed models
  - ✓ Generalized linear mixed models
  - ✓ Generalized Estimating equation
- Fundamental Assumptions

Background study

- ❖ Hypertension is an abnormally high blood pressure
- ❖ Force of the blood against the artery walls is too high
- ❖ Causes the heart to work harder than normal



Prevalence of hypertension ranges from:

- ❖ 15% in the west of southern Africa
- ❖ 25% in the east of southern Africa
- ❖ between 42% and 54% in South Africa

Introduction

- Repeated measurements of different types of outcomes may be collected from the same patient.
- Such as outcomes of continuous, binary and ordinal data.
- Many well-known statistical inference and methods have been established to model such repeated outcomes separately.
- Within-subject correlation of such data is important.
- Failure to do so may cause parameter estimation problems in the models.

Introduction Cont.

- Bayesian analysis methods for its statistical inferences.
- MCMC algorithm for the posterior conditional distribution for each of the parameters
- Gibbs sampler and WinBUGS for model implementation.

Challenges

- Pairwise approach for the joint model, but leads to a loss of efficiency for the estimates of parameters.
- Also PROC NLMIXED routines in SAS to calculate the estimates of parameters but stops and returns error messages due to non-convergence problem.

Study objective

The objective of this study are:

- ❖ To examine the joint models and association of three different types of longitudinal outcomes of hypertensive patients
- ❖ Identify possible risk factors of the trivariate variables that contribute to the problem of hypertension.

Methodology: Model Specification

The model is motivated by our dataset focusing on the three mixed longitudinal outcomes. For the within-subjects correlation and the association between the mixed outcomes, we propose our joint linear mixed effects models as

$$
\begin{aligned}
& Y_{ij} = X_{1ij}^{T}\beta_1 + Z_{1ij}^{T} b_{1i} + \varepsilon_{ij}, \\
& \Pr\left(R_{ij} = 1\right) = L_1\left(X_{2ij}^{T}\beta_2 + Z_{2ij}^{T} b_{2i}\right), \qquad (1) \\
& \Pr\left(Q_{ij} \le k\right) = L_2\left(X_{3ij}^{T}\beta_3 + Z_{3ij}^{T} b_{3i} + \alpha_k\right),
\end{aligned}
$$

where $j = 1, \ldots, m_i; i = 1, \ldots, n; k = 1, \ldots, K-1, \varepsilon_{ij} \overset{iid}{=\sim} N\left(0, \sigma_{\varepsilon}^{2}\right)$,

$b_i = \left(b_{1i}, b_{2i}, b_{3i}\right)^T \overset{iid}{\sim} N\left(0, H\right), 0 = \alpha_1 < \alpha_2 < \ldots < \alpha_{K-1}$,

$\beta = \left(\beta_1, \beta_2, \beta_3\right)^T, L_1\left(.\right)$ and $L_2\left(.\right)$

Methodology: Statistical Methods

- ❖ Likelihood-based methods may be used for parameter estimation but are
  - ✓ Analytically challenging and computationally demanding.
  - ✓ If random effects are high dimensional, integral problem becomes extremely difficult.
- ❖ Propose quasi-likelihood function and use Bayesian Sampling Methods for the statistical inference.
- ❖ For the required complete conditional distributions of all parameters, we use data augmentation approach.
- ❖ Gelman and Rubin’s diagnostic test for assessing the convergence.
- ❖ The prior distribution of the coefficients follows a Multivariate skew-normal distribution.

Model Comparison and Selection

We use deviance information criterion (DIC) that were obtained from the Markov chain Monte Carlo (MCMC) analysis to select the best model and evaluate the proposed model performance.

- Deviance Information Criterion (DIC)

$$DIC_m = D_m(E_{\theta|z}(\theta)) + 2p_D = \overline{D(\theta)} + p_D$$

- Expected Akaike Information Criteria (EAIC)
- Expected Bayesian information criteria (EBIC)
- Log-pseudo Marginal Likelihood (LPML)
- Conditional predictive ordinate (CPOS)

Motivating Study data

- 376 hypertensive patients with age greater than 18 years
- Trivariate longitudinal responses consist of BMI as continuous outcome, Depression (Yes/No) as binary outcome, and Stress interference (scale 1-4) as ordinal outcomes
- The covariates include Age, Alcohol, Sex and Time to heart attack
- The patients are aged 18 years and over, and their trivariate responses were collected at baseline at Time$\in \{0, 3, 6\}$
- We use Age2=Age+Time-18 as a time-dependent covariate
- Alcohol and Income were collected as covariates of ordinal categories

The Results

In this real data analysis, we run a MCMC chain of length 6000 with burn-in number 5000 in order to ensure the stability of the posterior mean of the parameters. Statistical results is presented in Table 1

$$BMI_{ij} = \beta_{10} + \beta_{11}\mathrm{Age2}_{ij} + \mathrm{Income}_i + b_{i1} + \varepsilon_{ij} \qquad (j = 1, 2, 3, \quad i = 1, 2, \ldots 376)$$

$$\Pr\left(\mathrm{Depression}_{ij} = 1\right) = \Phi\left(\beta_{20} + \beta_{21}\mathrm{Age2}_{ij} + \mathrm{Alcohol}_i + \mathrm{Income}_i + b_{i2}\right)$$

$$\Pr\left(\mathrm{Stress\ interference}_{ij} \le k\right) = \Phi\left(\alpha_k - (\beta_{30} + \beta_{31}\mathrm{Age2}_{ij} + \mathrm{Gender}_i + \mathrm{Alcohol}_i + \mathrm{Income}_i + b_{i3})\right)$$

- ❖ Age does not play a very important role in modelling for binary outcomes Depression.
- ❖ Interestingly, people with higher incomes seem to have a lower BMI value and a reduced probability of being depressed.

- Age2 variable is insignificant-separate model, this could issues of under fitting the model or bias in estimates obtained with the separate approach.
- BMI values tend to decrease slightly with age, but heart attack may become more severe as their age increases.
- There is no much difference in efficiency between the two approaches, which may be due to weak correlation as measurements are only taken at baseline, year three and year six.
- People who drink alcohol once/twice a week have less probability of being depressed compare those who drink daily/most days, once/twice a month, once/twice a year, or never drink.
- People who never drink have less probability of suffering from severe attack compare to others.
- People with higher incomes less likely to suffer from severe stress interference.
- Finally, male have a higher risk of suffering from stress interference than female.

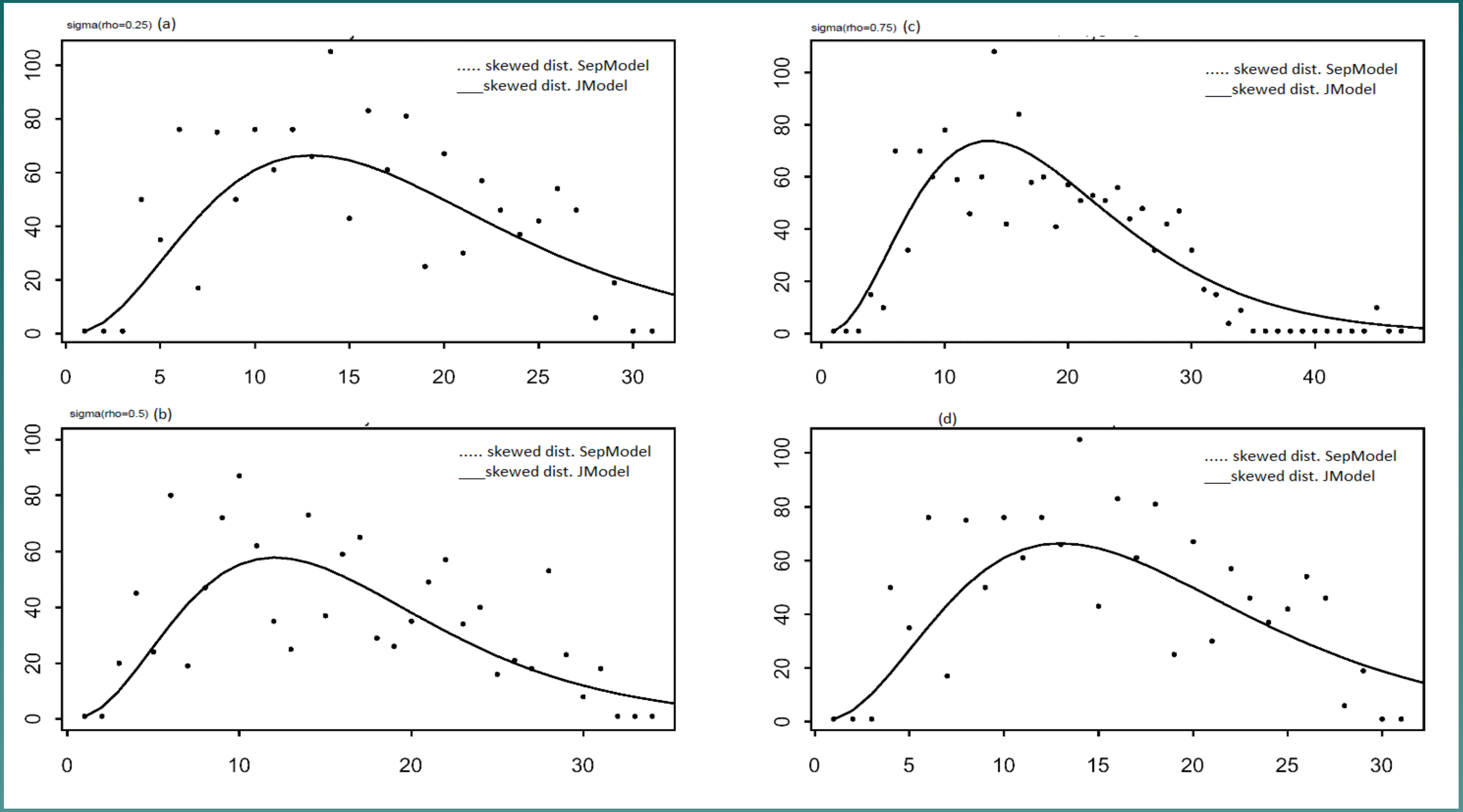


Figure 1 shows that the MCMC analysis shows that the standard deviations for the estimated regression coefficients by the joint modelling method are much more stable than these by the separate modelling approach

Conclusion

- From the results, it is suggested that the joint model was better in estimating trivariate longitudinal outcomes compare to separate model. Also, show a strong association among the three longitudinal outcomes.
- It takes about six hours to have a MCMC chain with length of 11,000 converges.
- Variable selection with respect to random effects in the joint models is also an open issue and deserves a further investigation.
- Finally, we think that marginal generalized linear models may also be useful in this context of joint models.