Fall 12-18-2020

Bayesian Modeling for Longitudinal Count Data: Applications in Biomedical Research

Morshed Alam
University of Nebraska Medical Center

Bayesian Modeling for Longitudinal Count Data: Applications in Biomedical Research

by

Morshed Alam

A Dissertation

Presented to the Faculty of

the University of Nebraska Graduate College

in Partial Fulfillment of the requirements

for the Degree of Doctor of Philosophy

Biostatistics

Graduate Program

Under the supervision of Professors Jane Meza

and

Yeongjin Gwon

University of Nebraska Medical Center

Omaha, Nebraska

December 8, 2020

Supervisory committee:

Fang Yu, Ph.D Kendra Schmid, Ph.D

Abstract

Title: Bayesian Modeling for Longitudinal Count Data: Applications in Biomedical Research

Student: Morshed Alam

Degree: Doctor of Philosophy Degree

School: College of Public Health, UNMC

Date: December 08, 2020

Biomedical count data such as the number of seizures for epilepsy patients, number of new tumors at each visit or the number vomiting after each chemo-radiation for the cancer patients are common. Often these counts are measured longitudinally from patients or within clusters in multi-site trials. The Poisson and negative binomial models may not be adequate when data exhibit over or under-dispersion, respectively. On the contrary, a variety of dispersion conditions in count data can be captured by Conway-Maxwell Poisson (CMP) model.

This doctoral dissertation relegates to developing a statistical methodology to model longitudinal count data distributed as CMP via mixed effect modeling approach. We propose a Bayesian CMP regression model. Specifically, we develop a regression model with random intercept and slope to capture heterogeneity among subjects and dependence over time. In addition, a Bayesian generalized additive mixed effect model based on CMP is proposed by assuming a non-linear shape of the functional relationship between mean of longitudinal response and covariates. Case studies demonstrating the usefulness of the proposed methodology by using real life clinical trial data are also presented. We apply an adaptive variant of Hamiltonian MCMC to carry out Bayesian computation. The Deviance Information Criterion (DIC), along with other Bayesian model assessment criteria such as (LPML), (WAIC), (LOO) are used for model comparisons.

Both in simulation studies and real data analysis, we conclude that in terms of model fitting, CMP models outperform the competing models when data exhibit dispersion.

Acknowledgements

In deep gratitude to all those who helped make this dissertation possible. In particular,it is my pleasure to express my sincere thanks to my dissertation advisors, Dr.Jane Meza, and Dr.Yeongjin Gwon, for their invaluable effort and guidance to accomplish my work. I offer my thanks to the committee members, Dr. Kendra Schmid, and Dr.Fang Yu for their continued advice and feedback throughout the entire process.

I also take the opportunity to thank Dr. Andrew W. Swift, and Dr. Mahbubul Majumder (Department of mathematics, UNO) for their help in coding process.

I want to thank Dr. David Guwatudde (Makerere University School of Public Health), Dr. Wafaie Fawzi (Harvard School of Public Health) and Dr Danstan Bagenda (UNMC) for sharing the data from their original study, and in particular to Dr. Bagenda for arranging the entire process including the preparation of the relevant data and codebook from their original paper, and answering my questions.

I also acknowledge Zaeema Naveed, and Ibur Rahman for supporting me in various ways. I would like to thank my parents, parents-in-law for their support and encouragement during my academic career. Finally, I thank my wife, Farhana Ahmed, my sons Md Mahir Ahnaf, and Mahbir Saheim for being supportive to me always.

Dedication

This dissertation is dedicated to my parents

Mr. Mohammad Shibly

and

Mrs. Sayera Begum

whose affection, love and prays make me possible to get success and honor

also

my beloved wife **Mrs.Farhana Ahmed** for her unprecedented sacrifice in achieving my academic goal

List of Abbreviations

GLM	Generalized Linear Model
CMP	Conway-Maxwell Poisson
GAM	Generalized Additive Model
MLE	Maximum Likelihood Estimate
CDF	Cumulative Density Function
PMF	Probability Mass Function
JGQL	Joint Generalized Quasi-likelihood
GMM	Generalized Method of Moments
MPL	Maximum Pseudo-likelihood
GEE	Generalized Estimating Equation
MCMC	Markov Chain Monte Carlo
HMC	Hamiltonian MCMC
NUTS	No-U-Turn-Samplers
BGLMM	Bayesian Generalized Linear Mixed Effect Model
BGAMM	Bayesian Generalized Additive Mixed Effect Model
NB	Negative Binomial distribution
IRLS	Iterative Weighted Least Squared Method
IW	Inverse-Wishart distribution
DIC	Deviance Information Criteria
WAIC	Watanabe-Akaike Information Criterion
PSIS-LOO	Pareto Smoothed Importance Sampling Leave One Out cross validation
CPO	Conditional Predictive Ordinate
LPML	Logarithm of Pseudo-Marginal Likelihood
ESS	Effective Number of Sample Size
HPD	Highest Posterior Density

Notations

y_{ij} count response at time j from subject i

t_{ij} time j for subject i

$\boldsymbol{y}_i$ count response vector from subject i

$\boldsymbol{y}_{(-i)}$ count response vector deleting subject i

θ shape parameter of CMP distribution

Θ collection of all parameters to be estimated

ϕ dispersion parameter of CMP distribution

$\boldsymbol{x}_{ij}$ fixed effect covariate vector

$\boldsymbol{X}_i$ design matrix for fixed effect covariates

$\boldsymbol{z}_{ij}$ random effect covariate vector

$\boldsymbol{Z}_i$ design matrix for random effects

$\boldsymbol{\beta}$ vector of regression coefficients

$\boldsymbol{\zeta_i}$ vector of random effects from subject i

σ_0^2 variance of random intercepts

σ_1^2 variance of random slopes

ρ correlation between random intercepts and slopes

$\boldsymbol{D}$ variance-covariance matrix for random effects

$\mathcal{L}(\cdot)$ likelihood function

$f(\cdot)$ probability density function

$\pi(\cdot)$ prior distribution

$U(\cdot,\cdot)$ uniform distribution

$N(\cdot,\cdot)$ normal distribution

$LN(\cdot,\cdot)$ log-normal distribution

$IW(\cdot,\cdot)$ inverse-Wishart distribution

φ degrees of freedom, a parameter of inverse-Wishart distribution

$\boldsymbol{\omega}$ scale matrix, a parameter of inverse-Wishart distribution

$\boldsymbol{\Omega}$ correlation matrix for random intercept and slope

$LKJ \sim (\eta)$ LKJ (Lewandowski, Kurowicka and Joe) distribution with parameter η

$\boldsymbol{L}$ upper or lower triangular matrix in Cholesky decomposition of $\boldsymbol{\Omega}$

$|\boldsymbol{J}|$ Jacobian

B number of MCMC samples

$\hat{(\cdot)}$ posterior mean of the parameter

p_D Effective Number of Parameters

Contents

Abstract ii

Acknowledgements iv

List of Abbreviations vi

Notations vii

List of Tables xii

List of Figures xiii

1 Introduction **1**

1.1 General Background 1

1.2 Literature Review 2

1.2.1 Chronology of CMP Distribution and its Applications 2

1.2.2 CMP Distribution 4

1.2.3 CMP in GLM 4

1.2.4 GAM in Count Data 6

1.3 Gaps in Literature 6

1.4 Specific Aims 7

2 Bayesian Generalized Linear Mixed Effect model for Longitudinal Count Data Distributed as Conway-Maxwell Poisson **10**

2.1 Introduction 10

2.2 The Proposed Model 12

2.2.1 CMP Regression Model 12
2.2.2 The Likelihood Function 13
2.3 Bayesian Inference 14
2.3.1 Priors and Posteriors 14
2.3.2 Bayesian Model Assessment 17
2.3.2.1 Conditional Predictive Ordinate (CPO) and Logarithm of Pseudo-Marginal Likelihood (LPML) 18
2.3.2.2 Leave One-Out-Cross-Validation (LOO) 19
2.3.2.3 Watanabe–Akaike Information Criterion (WAIC) . . 20
2.3.2.4 Deviance Information Criterion (DIC) 20
2.4 Simulation Study 21
2.5 Sensitivity Analysis 27
2.6 Application to Randomized Controlled Clinical trials Data 29
2.6.1 Analysis of Epilepsy data 29
2.6.2 Analysis of Multivitamin Supplementation in HIV Infected Adults Data 36
2.7 Discussion 38

3 Bayesian Generalized Additive Model for Longitudinal Count Data Distributed as Conway-Maxwell Poisson 47

3.1 Introduction 47
3.2 The Proposed Model 49
3.2.1 General form of B-Spline model 49
3.2.2 Statistical Model 50
3.2.3 The Likelihood Function 51
3.3 Bayesian Inference 51
3.3.1 Priors and Posteriors 51
3.4 Illustration of GAM 52

3.4.1 Specific model 53
3.5 Discussion 57

4 Association Between Multivitamin Supplementation and Overall well-being in HIV Patients:Application of the Bayesian Generalized Mixed Effect Model 63

4.1 Introduction 63
4.2 Methods 66
4.2.1 Data Description 66
4.2.2 Statistical Analysis 67
4.3 Results 68
4.4 Discussion 70
4.5 CMP Model Implementation 72

5 Conclusion and Future Research Directions 84

A 90

B 91

C 93

D 95

E 97

List of Tables

2.1 Simulation results for under-dispersed data ($\phi = 2.50$) 23

2.2 Simulation results for Over-dispersed data ($\phi = 0.30$) 25

2.3 Simulation results for Equi-dispersed data ($\phi = 1.0$) 26

2.4 Posterior summary for over-dispersed data with different choices of hyperparameters in priors 28

2.5 Posterior summary for over-dispersed data with different choices of prior distributions 30

2.6 Posterior summary under $\mathcal{M}_1$ 33

2.7 Posterior summary under $\mathcal{M}_2$ 34

2.8 Posterior summary under $\mathcal{M}_3$ 35

2.9 Posterior summary under the mixed effect model on Multivitamin supplementation in HIV infected adults data 37

3.1 Posterior summary under Poisson, NB, and CMP model 56

4.1 Posterior summary under Poisson, Negative binomial and CMP model 69

B.1 95% Credible Interval from simulated Under-dispersed ($\phi = 2.50$) data 91

B.2 95% Credible Interval from simulated Over-dispersed $\phi = 0.30$ data . 92

B.3 95% Credible Interval from simulated Equi-dispersed $\phi = 1.00$ data . 92

List of Figures

2.1 LKJ density plot for 2×2 Correlation Matrix $\boldsymbol{\Omega}$ 41

2.2 Boxplot for subject-wise CPO for CMP model with $\phi = 2.50$ 41

2.3 Comparison of posterior means of regression parameters with changing $\boldsymbol{\beta}$ prior parameters 42

2.4 Comparison of posterior means of regression parameters with changing ϕ prior parameters 42

2.5 Comparison of posterior means of regression parameters with changing σ_ζ^2 prior parameters 43

2.6 Comparison of posterior means of regression parameters with changing $\boldsymbol{L}$ prior parameters 43

2.7 Comparison of posterior means of regression parameters with changing prior distributions 44

2.8 Profiles of seizure rates for selected subjects 44

2.9 Treatment arm-wise profile of seizure rates 45

2.10 Weekly average seizure rates 45

2.11 Trace plots for $\mathcal{M}_3$ (CMP model) 46

2.12 Pair plots for $\mathcal{M}_3$ (CMP model) 46

2.13 ACF plots for$\mathcal{M}_3$ (CMP model) 46

3.1 Distribution of the data 59

3.2 Relationship between counts and age 59

3.3 Relationship between log(counts) and age 60
3.4 Trace plots for CMP model with B-spline 60
3.5 Density plots for CMP model with B-spline 61
3.6 Metropolis acceptance rate for CMP model with B-spline 61
3.7 Histogram for Rhats for CMP model with B-spline 62
3.8 ACF plots for CMP model with B-spline 62

4.1 Posterior summary of the parameters from CMP model with 95% Credible Interval 69

C.1 Boxplot for subject-wise CPO for CMP model with $\phi = 0.30$ 93
C.2 Boxplot for subject-wise CPO for CMP model with $\phi = 1.00$ 94

Chapter 1

Introduction

1.1 General Background

Count data is a common phenomenon in biomedical and public health research. Examples include number of hospitalizations of patients, number of physician visits, number of epileptic seizures, distinct multiple sclerosis lesions, number of tender or swollen joints of rheumatoid arthritis patients, symptom counts pertinent to a particular therapeutic regime , number of medicines consumed by patients for treatment of a specific disease, number of dental caries for patients, number of road accidents, and number of injury or death per road accidents, etc. Sometimes these data are longitudinally measured or clustered among sites such as hospitals or clinics, hence induce within-patient or within-cluster correlation. The underlying correlation invalidates the crucial assumption of independence which is a base for many statistical methods. To make valid inference, one must consider subject- or cluster-specific heterogeneity by incorporating random effects in the model. Parameter estimates in random effect models provide subject-specific interpretation (G. Fitzmaurice et al., 2008). Under-or over-dispersion is another issue to consider in count data models. Inclusion of random effects in the model accounts for additional variability in the data arise from repeated

measures from same subject, and over- or under-dispersion involved in the counting process (Morris et al., 2017).

The Poisson distribution is commonly used method for count data modeling. Since Poisson distribution belongs to the exponential family, modeling in the generalized linear model (GLM) framework under the Poisson assumption is simple. However, the Poisson distribution is not the best choice to deal with over- or under-dispersed data as they violate equal mean and variance assumption. The Negative Binomial model performs well for over-dispersed, but not for under-dispersed data. Over-dispersion is highly discussed in the literature while under-dispersion in the data is also common, especially in the case of rare events. Several generalizations of Poisson models such as generalized Poisson, restricted generalized Poisson have been suggested in the literature (del Castillo & Pérez-Casany, 2005; Famoye, 1993; Famoye et al., 2004; Consul & Jain, 1973; Ridout & Besbeas, 2004). However, the downside of generalized Poisson models is the inability to capture some level of dispersion due to truncation of the dispersion parameter under certain conditions (Famoye, 1993). In this case the Conway-Maxwell Poisson (CMP) distribution is the better choice as it can capture a wide range of dispersion, and belongs to the exponential family for a non-varying dispersion parameter (Conway & Maxwell, 1962; K. F. Sellers & Shmueli, 2010).

1.2 Literature Review

1.2.1 Chronology of CMP Distribution and its Applications

The CMP distribution was developed by Conway and Maxwell (Conway & Maxwell, 1962). Minka et al. (2003) discussed approximate computational schemes for moments, and maximum likelihood estimates (MLE) of the parameters of CMP distribution. Shmueli et al. (2005) studied the discrete nature of the distribution, probabilistic

properties, special cases of CMP, and the parameter estimation techniques (weighted least square, maximum likelihood, and Bayesian). In addition, they discussed several extensions of CMP distribution such as zero inflated, zero deflated CMP distributions, and CMP binomial distribution with a hint to extension to CMP multinomial distribution.

The works, particularly by Shemueli et al.(2005) revived further research on the CMP distribution during the last decade. Theoretical developments of CMP are found in literature. For instances, the conjugate analysis of CMP (Kadane et al., 2006), the explicit and exact expression of the cumulative density function (CDF) of CMP along with its useful moments (Nadarajah, 2009), properties of CMP and its generalization of binomial distribution (referred as CMB) (Daly & Gaunt, 2015), a bi-variate CMP as a generalization of the bivariate count data model (K. F. Sellers et al., 2016), sum of the CMP and its special cases (K. F. Sellers et al., 2017), asymptotic expansion for the normalizing constant of the CMP (Gaunt et al., 2019). Sellers et al. (2012) provides a good review of CMP modeling approaches, their applications, and proposed a generalized control chart for CMP data.

Literature reveals the earlier applications of CMP as the modeling of the state dependent service rates (Conway & Maxwell, 1962), word length in linguistics (Wimmer & Altmann, 1996; Wimmer et al., 1994), count data in marketing and e-Commerce (Boatwright et al., 2003; Borle et al., 2006; Kalyanam et al., 2007; Borle et al., 2005; Singh et al., 2009), motor vehicle crashes (Lord et al., 2010, 2008), cure rate models with application to cutaneous melanoma data (Rodrigues et al., 2009), seizure counts in epileptic patients (K. F. Sellers et al., 2017). Survival data analysis with a Weibull-Conway-Maxwell-Poisson distribution was suggested by Gupta & Huang (2017). Choo-Wosoba et al. (2018) studied longitudinal fluoride exposure with dental fluorosis and dental caries .

1.2.2 CMP Distribution

Let y_i be the count response for subject i, then the probability mass function (PMF) of the CMP distribution (Conway & Maxwell, 1962) with shape parameter θ_i and dispersion parameter ϕ is given by

$$P(Y_i = y_i) = \frac{\theta_i^{y_i}}{(y_i!)^{\phi} Z(\theta_i, \phi)}, \tag{1.1}$$

where $\theta_i > 0$, $\phi \geq 0$, and $Z(\theta_i, \phi) = \sum_{k=0}^{\infty} \frac{\theta_i^k}{(k!)^{\phi}}$ is the normalizing constant for $i = 1, \ldots, I$.

The dispersion parameter, $\phi = 1, < 1, > 1$ indicate equi-dispersion, over-dispersion and under dispersion respectively. Special cases of CMP are Poisson ($\phi = 1$), Geometric ($\phi = 0, \theta_i < 1$), and Bernoulli ($\phi \to \infty, p_i = \theta_i/(1+\theta_i)$), where p_i is the probability of success. There is no simple closed form for linking θ_i and ϕ. However, a formulation is given by Ralph Snider mentioned in K. F. Sellers et al. (2012) as $\theta_i = E(Y^{\phi})$, where θ_i is the expected value of power transform counts with power ϕ. The approximated mean and variance are $E(Y) \approx \theta_i^{1/\phi} - (\phi - 1)/2\phi, V(Y) \approx (\theta_i^{1/\phi})/\phi$ respectively. This approximation might not be accurate for $\phi > 1$ or $\theta^{1/\phi} < 10$ (Shmueli et al., 2005). When ϕ is close to 1 then θ_i approximates the mean. However, in case of over-dispersion ($\phi < 1$) or under-dispersion($\phi > 1$), θ_i deviates substantially from the mean.

1.2.3 CMP in GLM

K. F. Sellers & Shmueli (2010) proposed a GLM approach by using a logarithmic link for the shape parameter of CMP distribution. The authors discussed estimation, prediction, inference, model diagnostics, interpretation, and test of over dispersion in frequentist approach. Lord et al. (2010) fitted a GLM using dual links for shape

and dispersion parameters respectively and compared gamma and Poisson models. To account for group level dispersion in the data, K. F. Sellers & Shmueli (2013) developed a generalization of CMP regression model.

Khan & Jowaheer (2013) modeled the shape parameter of CMP distribution to study longitudinal count data. They estimated the parameters using a joint generalized quasi-likelihood estimating equation (JGQL) and generalized method of moments (GMM). Choo-Wosoba et al. (2016) studied zero inflated clustered count data. They used a modified Newton-Raphson method, maximum pseudo-likelihood (MPL), and generalized estimating equation (GEE) techniques for parameter estimation. K. F. Sellers & Raim (2016) studied a zero-inflated CMP model. Morris et al. (2017) extended the CMP generalized linear regression model by incorporating a random intercept for clustered data. They used a logarithmic link for the shape parameter and indirectly linked it with the mean of the CMP model. This model is implemented both in SAS (PROC NLMIXED) and R (integrate function for numerical integration for approximation of marginal likelihood, nlminb function for optimization). The COMPoissonReg R-package (K. Sellers et al., 2019) and COUNTREG procedure in SAS (SAS Institute Inc. 2014) support CMP GLM and its zero inflated variants. Choo-Wosoba & Datta (2018) studied zero inflated count data distributed as CMP with cluster specific random effects using Gaussian-Hermite quadrature.

A number of Bayesian approaches to GLM with CMP distributional assumption of count data found in literature. For instance, Guikema & Goffelt (2008) proposed GLMs by modeling log link of shape and dispersion parameters simultaneously. They also modeled the centrality parameter (by using a transformation of the shape and dispersion parameter) with a log link in Bayesian setting by using Gibbs sampler. Chanialidis et al. (2014) proposed a method to estimate normalizing constant of CMP using retrospective sampling algorithm. Further, Chanialidis et al. (2018) proposed a rejection sampling approach combined with exchange algorithm

that does not require evaluation of normalizing constant. Wu et al. (2013) studied a Bayesian spatio-temporal Conway–Maxwell Poisson model with dynamic dispersion parameter. Choo-Wosoba et al. (2018) proposed a Bayesian approach of GLM in presence of many zero counts in the data.

1.2.4 GAM in Count Data

The Generalized additive model (GAM) was first introduced by T. J. Hastie & Tibshirani (1990). They states that in clinical trials and observational studies the GAM is useful for two reasons. It helps to prevent from model miss-specification that may lead to invalid conclusions regarding treatment efficacy. The GAMs also provide information regarding the relationship between prognostic factors and disease risk (T. J. Hastie & Tibshirani, 1990). Ruppert et al. (2003) discussed the GAM as a generalized nonparametric regression model.The generalized additive models for the length of stay in hospitals was studied by Herwartz et al. (2016) in Bayesian setting. They modeled count data by using Poisson, Negative Binomial, zero inflated variants of Poisson, and Hurdle model by incorporating group-specific random effect. T. Hastie (2008) developed a R package 'gam' to implement GAM. The available software for dealing with GAM are VGAM (Yee et al., 2020), polspline (Kooperberg, 2015), mgcv (Wood & Wood, 2015), and gamlss (Stasinopoulos et al., 2017).

The distinctive features of allowing flexible predictor effects, interpretability, and availability of ready-to use software have resulted widespread application of GAMs. However, literature on GAM for count data is limited.

1.3 Gaps in Literature

The CMP is a general distribution that can capture a wide range of dispersion in count data. Although biomedical research generates count data frequently, the application

of CMP in the biomedical field is still limited. Morris et al. (2017) used longitudinal epilepsy data and showed that CMP model fits better than the corresponding Poisson and Negative Binomial models. They included subject- specific random intercept in the model and used Akaike Information Criteria (AIC) (Akaikei, 1973) for model comparisons. To account for more subject or cluster specific heterogeneity in count data, we may need to fit models with random intercept and slope. However, literature indicates that such model has not yet been studied in CMP model setting. Besides, none of the existing packages supports CMP generalized Mixed Effect Model.

A crucial aspect of GAM is to capture the nonlinear relationship between the link function and continuous covariates. The GAM was studied in literature using count data including their zero inflated variants (Harezlak et al., 2018). The CMP distributional assumption of count data not yet been considered in GAM framework. In addition, the available packages to deal with GAM do not support Conway-Maxwell-Poisson modeling of count data generated from any of the cross sectional, longitudinal or clustered type of studies.

1.4 Specific Aims

Dealing with CMP model is complicated. The CMP probability mass function includes a normalizing constant which is an infinite series, and leads to intractable integration or differentiation. The normalizing constant in CMP is being evaluated by (a) truncation (Morris et al., 2017; Chanialidis et al., 2018) (b) asymptotic approximation (Minka et al., 2003), and (c) by using an MCMC scheme based on retrospective sampling method (Chanialidis et al., 2014). The difficulty of this evaluation restricts the wide application of the distribution for complex models, especially, when models includes high-dimensional random effects. Inclusion of random effects, makes the CMP mixed model more complicated since the random effects are to be integrated

out during estimation, especially in classical setting. In addition, dealing with flexible semi-parametric model (e.g. GAM) is also complicated in classical setting for CMP model.

To avoid these complexities mentioned above, a Bayesian approach will be helpful. In Bayesian settings, Guikema & Goffelt (2008) used Gibbs Samplers, Choo-Wosoba & Datta (2018) used Gibbs sampler and Metropolis-Hastings algorithms , and Chanialidis et al. (2018) used rejection sampling with exchange algorithm. However, these methods are computationally intensive, and they did not deal with longitudinal data by including higher order random effects.

We attempt to use No-U-Turn-Samplers (NUTS), an adaptive variant of Hamiltonian MCMC (HMC) to draw sample from a posterior distribution. Details of NUTS and notable advantages of using HMC studied elsewhere (Hoffman & Gelman, 2014). The use of HMC facilitates quicker convergence for high-dimensional models irrespective of conjugate priors, and produces less auto correlated samples in comparison to other sampling techniques.

The specific aims are as follows:

- **Aim1:**

  To develop a Bayesian generalized linear mixed effect model (BGLMM) for longitudinal count data distributed as CMP.

  We hypothesize that inclusion of random effects in the CMP model will capture more subject or cluster specific heterogeneity, and will perform better in terms of model fit than the corresponding Poisson and negative binomial (NB) mixed effect models. In addition, the proposed Bayesian model will be more flexible to incorporate random effects than in models in the frequentist approach.

- **Aim2:**

  To develop a Bayesian generalized additive mixed effect model (BGAMM) for

longitudinal data distributed as CMP.

We hypothesize that the fitted BGAMM with the data distributed as CMP will perform better in terms of model fit than the corresponding Poisson and negative binomial (NB) additive mixed effect models. We also anticipate that the proposed Bayesian model will be more flexible to incorporate random effects than in the models in the frequentist approach.

- **Aim3:**

  CMP mixed effect implementation in STAN language, and to apply the proposed models in real biomedical data analysis.

  We will apply these models to real datasets arising from randomized controlled trials and compare them by using available model assessment tools, and provide a tutorial on CMP modeling in STAN

Accomplishing these specific aims will allow us to model biomedical longitudinal or clustered count data with a flexible distribution. We anticipate, the CMP model will outperform competing count data models in terms of model fitting.

The inclusion of random intercept and slope in the CMP mixed effect model, fitting a GAM in the context of CMP distribution, and the use of HMC to implement the models will be the additions to the existing literature to study longitudinal count data.

In Chapter 2, we develop a generalized linear mixed effect model. Development of a generalized additive mixed effect model is outlined in Chapter 3. In Chapter 4, we illustrate an application of the proposed model to real data arising from a clinical trial. Overall conclusion and Future research directions are mentioned in Chapter 5.

Chapter 2

Bayesian Generalized Linear Mixed Effect model for Longitudinal Count Data Distributed as Conway-Maxwell Poisson

2.1 Introduction

In this chapter we assume that longitudinal or cluster count data follow Conway-Maxwell Poisson (CMP) distribution. Statistical modeling on count response with CMP distributional assumption can accommodate a wide spectrum of dispersion in the data (Conway & Maxwell, 1962). Longitudinal or cluster count data analysis, by generalized linear mixed effect model, often includes random intercept (or random intercept and slope) to capture subject or cluster specific heterogeneity in the data. In such a model, in classical approach, parameter estimation requires maximization of marginal likelihood by integrating out the random effects. However, the integral with respect to the random effects becomes intractable due to the involvement

of an infinite normalizing constant in CMP probability mass function. To override the intractability, most classical approaches experience difficulties in approximating the likelihood with Laplacian or quadrature methods (Choo-Wosoba et al., 2018). Modeling with high-dimensional random effects or/and small sample size, we might experience a substantial deviation in the shape of the integrand function from that of the Gaussian density. The standard Laplace approximation may be inaccurate in such a situation (Ruli et al., 2016). They also noted that in the classical approach, model fitting via an iterative weighted least squared (IRLS) method encounters a non-convergence problem when Fisher's scoring matrix exhibits low rank for a given diagonal weight matrix, and enhanced complexity leads to over-fitting of the model.

On the contrary, a Bayesian method can handle a mixed effect model by avoiding the quadrature method or approximations to the likelihood function (Choo-Wosoba et al., 2018). A Bayesian approach avoids these approximations by applying iterative MCMC sampling schemes to draw the values of the random effects, and allows more flexibility to choose a versatile form of random effects design matrix. Inclusion of priors enables addressing the convergence issues, and avoiding over-fitting of a model (Ruli et al., 2016). With suitable priors, Bayesian methods provide inference based on posterior summary instead of maximizing the log-likelihood function.

K. F. Sellers & Shmueli (2010) developed the CMP generalized linear model (GLM). Later on, in the frequentist setting, Morris et al. (2017) extended the GLM by incorporating subject-specific random intercept for clustered count data with an arbitrary truncation of normalizing constant. As a further extension, we propose a Bayesian generalized mixed effect model (BGLMM) for longitudinal count data by including random intercept and slope which avoids the integrational intractability. Although the focus of the study is to incorporate random intercept and slope in the mixed effect model with CMP distributional assumption of the data, we also explore models with random intercept only in real data analysis.

Section 2.2 includes the proposed statistical model while Bayesian inference along with model assessment criteria are presented in Section 2.3. A detailed simulation study and sensitivity analysis are presented in Section 2.4 and Section 2.5 respectively. Application of the model with clinical trials data are illustrated in Section 2.6. Discussion of the study is noted in Section 2.7. The full conditionals noted in Appendix A.

2.2 The Proposed Model

2.2.1 CMP Regression Model

Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$ be the independent count response vector of subject i for $i = 1, \ldots, I$ and $j = 1, \ldots, n_i$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i})^T$ be a $(n_i \times (p+1))$ design matrix of fixed effect covariates, where $\boldsymbol{x}_{ij}^T = (1, x_{ij1}, \ldots, x_{ijp})$ is a $(p+1)$ dimensional covariate vector, $\boldsymbol{Z}_i = (\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{in_i})^T$ be a $(n_i \times q), (q \leq (p+1))$ known design matrix, where $\boldsymbol{z}_{ij}^T = (1, z_{ij1}, \ldots, z_{ij(q-1)})$ is a q-dimensional covariate vector, $\boldsymbol{\zeta}_i$ be a q-dimensional vector of random effects for the subject i, and $\boldsymbol{\beta}$ be a $(p+1)$-dimensional vector of regression coefficients. Then, a generalized linear mixed effect model (GLMM) is defined by

$$\mathrm{E}(\boldsymbol{y}_i|\boldsymbol{\beta}, \boldsymbol{\zeta}_i) = g^{-1}(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\zeta}_i), \tag{2.1}$$

where $g^{-1}(\cdot)$ is a link function.

Further, let θ_{ij} be the shape parameter of CMP distribution associated with a response, y_{ij} from subject i at time j. Then, the GLMM for longitudinal count response (distributed as CMP) is given by

$$\log(\theta_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\zeta}_i, \tag{2.2}$$

where $\boldsymbol{\zeta}_i \sim N_q(\mathbf{0}, \boldsymbol{D})$, $\boldsymbol{D}$ is the covariance matrix for random effects. Let $\boldsymbol{\zeta}_{qi} = (\boldsymbol{\zeta}_0, \boldsymbol{\zeta}_1)$ be the vector of random intercept and random slope, then

$$\boldsymbol{D} = \begin{bmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{bmatrix},$$

where σ_0^2 and σ_1^2 are the variances of intercept and slope, respectively, and ρ is the correlation between them.

2.2.2 The Likelihood Function

By considering the random effects as the latent variables, the complete data likelihood can be written in the following form:

$$\mathcal{L}(\boldsymbol{\beta}, \phi, \boldsymbol{D}) = \prod_{i=1}^{I} \left[\prod_{j=1}^{n_i} f(y_{ij}|\boldsymbol{\beta}, \phi, \boldsymbol{\zeta_i}) f(\boldsymbol{\zeta_i}|\boldsymbol{D}) \right]. \tag{2.3}$$

When the response y_{ij} distributed as CMP, we can write equation (2.3) as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}, \phi, \boldsymbol{D}) = & \prod_{i=1}^{I}\prod_{j=1}^{n_i} \left(\frac{1}{y_{ij}!}\right)^{\phi} \times \left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\zeta}_i\right)\right)^{y_{ij}} \quad \times \left(\frac{\sum_{k=0}^{\infty} \left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\zeta}_i\right)\right)^k}{(k!)^{\phi}} \right)^{-1} \\ & \times \prod_{i=1}^{I} (2\pi)^{-(q/2)} \frac{\exp\left(-\frac{1}{2}\left[\boldsymbol{\zeta}_i^T \boldsymbol{D}^{-1} \boldsymbol{\zeta}_i\right]\right)}{|\ \boldsymbol{D}|^{\frac{1}{2}}}. \end{aligned} \tag{2.4}$$

2.3 Bayesian Inference

The Bayesian model fitting requires specification of prior distributions, generation of the joint posterior distribution for the parameters of interest, and obtaining MCMC samples from the posterior distribution by using suitable samplers to avoid complex or intractable integration. Bayesian computation enables generating posterior summary and exploring different characteristics of the parameters.

2.3.1 Priors and Posteriors

In order to fit a Bayesian GLMM, we assume that $\boldsymbol{\beta}$, $\boldsymbol{D}$, and ϕ are independent apriori. Thus,

$$\pi(\boldsymbol{\beta}, \boldsymbol{D}, \phi) = \pi(\boldsymbol{\beta}) \times \pi(\boldsymbol{D}) \times \pi(\phi). \tag{2.5}$$

We further assume that $\boldsymbol{\beta} \sim N_p(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0), \phi \sim LN(\mu_\phi, \psi)$, and $\boldsymbol{D} \sim IW(\varphi, \boldsymbol{\omega})$. Then the joint posterior distribution under the proposed model is given by

$$\begin{aligned}
\pi(\boldsymbol{\beta}, \boldsymbol{D}, \boldsymbol{\zeta}, \phi \mid \boldsymbol{y}) &= \left[\prod_{i=1}^{I}\prod_{j=1}^{n_i} f(y_{ij} \mid \boldsymbol{\beta}, \phi, \boldsymbol{\zeta_i}) \times \pi(\boldsymbol{\beta})\right] \times \left[\prod_{i=1}^{I} f(\boldsymbol{\zeta}_i \mid \boldsymbol{D}) \times \pi(\boldsymbol{D}) \times \pi(\phi)\right] \\
&\propto \prod_{i=1}^{I}\prod_{j=1}^{n_i} \left(\frac{1}{y_{ij}!}\right)^{\phi} \times \Big(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\zeta}_i\right)\Big)^{y_{ij}} \\
&\times \left(\frac{\sum_{k=0}^{\infty}\left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\zeta}_i\right)\right)^k}{(k!)^{\phi}}\right)^{-1} \\
&\times \frac{\exp\left(-\frac{1}{2}\Big[(\boldsymbol{\beta}-\boldsymbol{\beta}_0)^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\Big]\right)}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} \times \prod_{i=1}^{I}\frac{\exp\left(-\frac{1}{2}\left[\boldsymbol{\zeta}_i^T\boldsymbol{D}^{-1}\boldsymbol{\zeta}_i\right]\right)}{|\boldsymbol{D}|^{\frac{1}{2}}} \\
&\times |\boldsymbol{D}|^{-\frac{(\varphi+q+1)}{2}} \exp\Big(-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\omega}\boldsymbol{D}^{-1}\right)\Big) \times \frac{\exp\Big(-\frac{1}{2}\big(\frac{\log\phi-\mu_\phi}{\psi}\big)^2\Big)}{\psi\phi}.
\end{aligned} \tag{2.6}$$

In (2.6), we use inverse-Wishart (IW) distribution for $\boldsymbol{D}$ and log-normal distribution for ϕ. However, a wide variety of priors can be used with different parameter specifications such as uniform prior for $\boldsymbol{\beta}$, multivariate-t prior for random effects, and

any distribution with positive support such as half-Cauchy, gamma or uniform for ϕ, inverse gamma for variance parameters of the random effects. As discussed in Barnard et al. (2000), the inverse-Wishart distribution is a natural choice of $\boldsymbol{D}$ due to conjugacy. The alternative choices of priors are scaled inverse-Wishart (O'Malley & Zaslavsky, 2008), hierarchical half-t prior (A. Huang et al., 2013), restricted Wishart distribution (Wang et al., 2018), and separation of covariance strategy (Barnard et al., 2000). However, Alvarez et al. (2014) noted that the IW prior suffers from the limitations, such as (a) a single degree of freedom parameter φ controls uncertainty in all variance parameters and does not allow flexibility to incorporate various amount of prior knowledge to various variance components (Gelman et al., 2013), (b) if $\varphi > 1$, the marginal distribution of variance parameter retains lower density close to the zero region and causes bias in the posterior estimate of variance (Gelman et al., 2006), and (c) it imposes dependency between variance and correlation, meaning that larger variances are associated with extreme correlations (near to $+1$ and -1) while smaller variances are associated with correlations near to zero (Tokuda et al., 2011). However, we can also consider distribution of correlation or its variants as a prior. Alternatively, there is a separation strategy proposed in Barnard et al. (2000), which treats variance and correlations independently. For example, using the separation strategy, the covariance matrix $\boldsymbol{D}$ is decomposed as

$$\boldsymbol{D} = \begin{bmatrix} \sigma_0 & 0 \\ 0 & \sigma_1 \end{bmatrix} \boldsymbol{\Omega} \begin{bmatrix} \sigma_0 & 0 \\ 0 & \sigma_1 \end{bmatrix},$$

where $\boldsymbol{\Omega}$ is a correlation matrix, and defined as

$$\boldsymbol{\Omega} = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}.$$

Lewandowski et al. (2009) suggests the LKJ (Lewandowski, Kurowicka and Joe) distribution to sample correlation matrix $\mathbf{\Omega}$ uniformly from a space of positive definite correlation matrices. Currently, the LKJ distribution is used as a default option in Stan language (Carpenter et al., 2017).

If $\mathbf{\Omega} \sim LKJ(\eta)$, then the density is given in Lewandowski et al. (2009) as

$$f(\mathbf{\Omega}) \propto \det(\mathbf{\Omega})^{(\eta-1)}.$$

The parameter η controls the shape of the distribution. The value $\eta = 1$ indicates the prior is uniform over all valid correlation matrices, $\eta = 2$ indicates ρ retains high concentration at closer to zero region. The extreme correlation become less plausible as the value of η increases (McElreath, 2020). The identity matrix is the modal correlation matrix when $\eta > 1$, and the density has a trough at the identity matrix when $0 < \eta < 1$ (Stan user's guide, version 2.18).

The density of $\mathbf{\Omega}$ under LKJ distribution with different values of the shape parameter is shown in Figure 2.1.

Instead of directly modeling correlation matrix $\mathbf{\Omega}$ with LKJ density, Stan language provides an implicit parameterization in terms of Cholesky decomposition. For $\eta > 0$, the Cholesky decomposition of $\mathbf{\Omega}$ is given by

$$\mathbf{\Omega} = \boldsymbol{L}\boldsymbol{L}^T = \begin{bmatrix} l_{11} & 0 \\ l_{12} & l_{22} \end{bmatrix} \begin{bmatrix} l_{11} & l_{12} \\ 0 & l_{22} \end{bmatrix},$$

where $\boldsymbol{L}$ is a lower triangular matrix with $l_{mm} > 0$, for $m = 1, 2$ and each row $\boldsymbol{L}_m$ has unit Euclidian length.

The general form of LKJ density of a $M \times M$ lower triangular matrix $\boldsymbol{L}$ is given as

$$h(\boldsymbol{L}|\eta) \propto |\boldsymbol{J}| \det(\boldsymbol{L}\boldsymbol{L}^T)^{(\eta-1)} \quad \propto \prod_{m=2}^{M} l_{mm}^{M-m+2\eta-2}, \tag{2.7}$$

where $h(\cdot)$ is the density of LKJ distribution, $|\boldsymbol{J}|$ is the Jacobian for the transformation from $\boldsymbol{\Omega}$ to $\boldsymbol{L}$, and η has the similar interpretation as mentioned above. However, $\eta = 1$ does not imply that distribution of $\boldsymbol{L}$ is constant while distribution of $\boldsymbol{L}\boldsymbol{L}^T$ is constant for the same (Stan user's guide, version 2.18). In our analysis we use distribution of $\boldsymbol{L}$ as a prior.

2.3.2 Bayesian Model Assessment

The commonly available Bayesian model comparison tools are Deviance Information Criteria (DIC) (Spiegelhalter et al., 2002), Watanabe-Akaike Information Criterion (WAIC), popularly known as widely applicable Bayesian information criterion (Watanabe, 2013), Pareto smoothed importance sampling leave-one-out cross validation (PSIS-LOO) and K-fold-Cross Validation (K-fold CV) (Vehtari et al., 2017). DIC suffers from problems as it uses point estimation rather than being fully Bayesian (Plummer, 2008; Van Der Linde, 2005), DIC is not defined for singular models (Vehtari et al., 2017). The WAIC can overcome some pitfalls of DIC, although there is no theoretical basis why WAIC is unreliable in some situations (Vehtari et al., 2017). The k-fold CV provides more reliable results when importance sampling LOO (IS-LOO) fails for a large number of data points (Vehtari et al., 2017). No method can dominate all others. Therefore, as the Bayesian model assessment criteria, we attempt to calculate Monte-Carlo version of conditional predictive ordinate (CPO), and corresponding logarithm of pseudo-marginal likelihood (LPML) as discussed in (Geisser & Eddy, 1979; Zhang et al., 2017; Gelfand et al., 1992; Chen et al., 2012).

In addition, we calculate WAIC, Bayesian Leave One Out cross validation (LOO), and Deviance Information Criteria (DIC) for Bayesian model assessment. Among the competing models the best model is the one having the largest LPML, while the lowest values of LOO, WAIC, and DIC indicate the best fit of the model.

2.3.2.1 Conditional Predictive Ordinate (CPO) and Logarithm of Pseudo-Marginal Likelihood (LPML)

The conditional predictive ordinate (CPO) is calculated based on leave-one-out cross validation. To define CPO, let us consider, $\boldsymbol{y} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_I)^T$ be the longitudinal response data. Then, the CPO provides the estimate of probability of observing a future $\boldsymbol{y_i}$ given the observed $\boldsymbol{y}_{(-i)}$. The vector $\boldsymbol{y}_{(-i)} = (\boldsymbol{y}_1, \ldots, \boldsymbol{y}_{i-1}, \boldsymbol{y}_{i+1}, \ldots, \boldsymbol{y}_I)^T$ denotes all observations deleting the data points for ith subject. It is the posterior probability of observing the value of $\boldsymbol{y_i}$ when the model is fitted to all data except $\boldsymbol{y_i}$. Higher value of CPO indicates a better fit of the model to $\boldsymbol{y_i}$ while a lower value reveals that $\boldsymbol{y_i}$ is an outlier and influential observation. Following Geisser & Eddy (1979); Chen et al. (2012), with the given notation the CPO for subject i is defined as

$$\text{CPO}_i = \int f(\boldsymbol{y}_i \mid \Theta)\pi\big(\Theta \mid \boldsymbol{y}_{(-i)}\big)d\boldsymbol{\Theta}. \tag{2.8}$$

On simplification, as noted in (Zhang et al., 2017), we can write

$$\begin{aligned} \text{CPO}_i &= \left[\int \frac{1}{f(\boldsymbol{y_i} \mid \boldsymbol{\Theta})}\pi(\boldsymbol{\Theta} \mid \boldsymbol{y})d\boldsymbol{\Theta}\right]^{-1} \\ &= \left[\, E_{\boldsymbol{\Theta}|\boldsymbol{y}}\left(\frac{1}{f(\boldsymbol{y_i}|\boldsymbol{\Theta})}\right)\right]^{-1}, \end{aligned} \tag{2.9}$$

where $\boldsymbol{\Theta}$ includes ($\boldsymbol{\beta}$, ϕ, and $\mathbf{D}$). Following Chen et al. (2012); Zhang et al. (2017), the Monte Carlo estimate of the CPO (B is the number of MCMC samples) is given by

$$\widehat{\text{CPO}}_i = \left[\frac{1}{B}\sum_{b=1}^{B}\frac{1}{f(\boldsymbol{y_i}|\boldsymbol{\Theta}^{(b)})}\right]^{-1}. \tag{2.10}$$

Then, the LPML is defined as

$$\text{LPML} = \sum_{i=1}^{I}\log(\widehat{\text{CPO}}_i). \tag{2.11}$$

2.3.2.2 Leave One-Out-Cross-Validation (LOO)

Following the similar notations in section 2.3.2.1, the leave-one-out predictive density for a given dataset deleting the ith data point is given as

$$\text{elpd}_{\text{loo}} = \sum_{i=1}^{I}\log f(\boldsymbol{y_i}|\boldsymbol{y}_{(-i)}), \tag{2.12}$$

where

$$f(\boldsymbol{y_i}|\boldsymbol{y}_{(-i)}) = \int f(\boldsymbol{y_i}|\boldsymbol{\Theta})\pi(\boldsymbol{\Theta}|\boldsymbol{y}_{(-i)})d\boldsymbol{\Theta}. \tag{2.13}$$

On simplification, as noted in Vehtari et al. (2017), from MCMC samples, (2.13) can be approximated as

$$f(\boldsymbol{y_i}|\boldsymbol{y}_{(-i)}) \approx \Big[\frac{1}{B}\sum_{b=1}^{B}\frac{1}{f(\boldsymbol{y_i}|\boldsymbol{\Theta}^{(b)})}\Big]^{-1}. \tag{2.14}$$

Then, to provide the output on the conventional scale of deviance, we can write

$$\text{LOO} = -2\sum_{i=1}^{I}\log\Big[\frac{1}{B}\sum_{b=1}^{B}\frac{1}{f(\boldsymbol{y_i}|\boldsymbol{\Theta}^{(b)})}\Big]^{-1}. \tag{2.15}$$

2.3.2.3 Watanabe–Akaike Information Criterion (WAIC)

The WAIC is an alternative measure to estimating the expected log pointwise predictive density and is defined as

$$\widehat{\text{elpd}}_{\text{waic}} = \widehat{\text{lpd}} - \widehat{\text{p}}_{\text{waic}}, \tag{2.16}$$

where the computed log point-wise predictive density ($\widehat{\text{lpd}}$) is

$$\widehat{\text{lpd}} = \sum_{i=1}^{I} \log \left[\frac{1}{B} \sum_{b=1}^{B} \frac{1}{f(\boldsymbol{y_i}|\boldsymbol{\Theta}^{(b)})} \right]^{-1}, \tag{2.17}$$

and the estimated effective number of parameters ($\widehat{\text{p}}_{\text{waic}}$) is defined as

$$\widehat{\text{p}}_{\text{waic}} = \sum_{i=1}^{I} \text{Var}\Big(\log f(\boldsymbol{y_i}|\boldsymbol{\Theta}^{(b)}) \Big). \tag{2.18}$$

As mentioned in Vehtari et al. (2017), on the conventional scale of deviance or AIC we can write

$$\text{WAIC} = -2\Big[\frac{1}{I} \sum_{i=1}^{I} \log \Big\{ \frac{1}{B} \sum_{b=1}^{B} f(\boldsymbol{y_i}|\boldsymbol{\Theta}^{(b)}) \Big\} \Big] - \sum_{i=1}^{I} \text{Var}\Big(\log f(\boldsymbol{y_i}|\boldsymbol{\Theta}^{(b)}) \Big). \tag{2.19}$$

2.3.2.4 Deviance Information Criterion (DIC)

The Deviance Information Criteria (DIC) proposed by Spiegelhalter et al. (2002) is the most widely used method for Bayesian model comparison which includes goodness-of-fit of the model as well as complexity of the model in terms of effective number of parameters (p_D). Let $\boldsymbol{y}$ be generated from the probability model $f(\boldsymbol{y}|\boldsymbol{\Theta})$, then the marginal distribution of $\boldsymbol{y}$, $f(\boldsymbol{y}) = \int_{\boldsymbol{\Theta}} f(\boldsymbol{y}|\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})d\boldsymbol{\Theta}$, where $\pi(\boldsymbol{\Theta})$ is the prior distribution of $\boldsymbol{\Theta}$. The posterior distribution, $\pi(\boldsymbol{\Theta}|\boldsymbol{y}) \propto f(\boldsymbol{y}|\boldsymbol{\Theta})\pi(\boldsymbol{\Theta})$. Then, the DIC is defined as

$$\text{DIC} = \overline{\Delta(\boldsymbol{\Theta})} + p_D, \tag{2.20}$$

where $\Delta(\boldsymbol{\Theta}) = -2\log f(\boldsymbol{y} \mid \boldsymbol{\Theta})$, is the deviance function, $\overline{\Delta(\boldsymbol{\Theta})}$ is the posterior mean of deviance, and $p_D = \overline{\Delta(\boldsymbol{\Theta})} - \Delta(\bar{\boldsymbol{\Theta}})$ is the effective number of parameters. On simplification,

$$\text{DIC} = 2\overline{\Delta(\boldsymbol{\Theta})} - \Delta(\bar{\boldsymbol{\Theta}}). \tag{2.21}$$

In addition, we monitor convergence of the MCMC chains by observing scale reduction statistics $\hat{R}$ that measures the ratio of the average variance of samples within each chain to the variance of the pooled samples across chains. A value of $\hat{R}$ closer to 1 indicates that each set of B simulated values is close to the target distribution (Gelman et al., 1992). To monitor the performance of the MCMC samples, we also observe effective number of sample size (ESS) where the higher ESS is an indication of the higher number of independent MCMC samples, trace plots visualizing the convergence status of MCMC samples, auto-correlation plots (ACF plots) where exponential shape is an indication of producing non-autocorrelated samples, pair plots and density plots illustrating the non-disruption in MCMC samples for the posterior means of the parameters (some of them are noted in Section 2.6).

2.4 Simulation Study

We perform a simulation study to demonstrate the flexibility and performance of the CMP mixed effect model with correlated subject-specific random intercept and slope. The datasets resemble clinical trial type of data where treatment arm consists of 40% of the subjects. We generate 100 longitudinal datasets each of having $n = 100$ subjects with 5 measurements in 5 distinct time points ($t = 0, 1, \ldots 4$) in under-dispersed ($\phi = 2.50$), over-dispersed ($\phi = 0.30$) and equi-dispersed ($\phi = 1.00$) conditions respectively. The time t_{ij} is then converted as binary variable (0 =baseline or 1=post-baseline). We assume that patient condition improve by 1% during post-baseline period, as it happens in some clinical trials due to counselling effect irrespective of

intervention. The only explanatory variable, $x \sim$ Bernoulli(0.40) mimics a treatment assignment to 40% of the patients, and we assume a moderate beneficial (5%) effect of treatment on the subjects. Random effects are generated with the specifications of $\zeta_0 \sim N(0, 0.05^2)$ and $\zeta_1 \sim N(0, 0.02^2)$. Correlation, ρ between random intercept and slope is considered as -0.50. The true values of the regression coefficients β_1=-0.05, and β_2=-0.01 represent the beneficial effect of treatment and time respectively. We consider β_0=-0.8. Then the response counts, y_{ij}s are generated with the specification, $y_{ij} \sim CMP(\theta_{ij}, \phi)$, where $\theta_{ij} = \exp(\beta_0 + \beta_1 x_{ij} + (\beta_2 + \zeta_{1i})t_{ij} + \zeta_{0i})$.

In each dispersion condition we applied CMP, negative binomial (NB), and Poisson model respectively with $5,000$ iterations (50% warm-ups) in each of the 4 chains. We compute Monte-Carlo versions of CPO and LPML, LOO, WAIC, and DIC along with their respective inter-quartile range (IQR) for the purpose of model assessments. Posterior mean, standard error (SE), mean squared error (MSE), and coverage probability (CP%) (in 95% credible interval) are reported in Tables 2.1, 2.2, and 2.3 for three dispersion conditions respectively. We also perform graphical illustration of subject-wise CPOs from all samples for all dispersion conditions to observe existence of extreme observations. However, we report here only the graph for CMP model with under-dispersed data in Figure 2.2, and the rest are reported in Appendix C.

Table 2.1 presents the results obtained from CMP, NB, and Poisson model assuming under-dispersed ($\phi = 2.50$) simulated data. The CMP model produces the posterior mean of dispersion parameter as $\phi = 2.82$ with a MSE=0.48. The result supports that data are under-dispersed since $\phi > 1$. On the other hand, NB model produces a negligible value for ϕ (0.04). However, dispersion parameters in CMP and NB carry different interpretations. When value of the dispersion parameter in NB approaches to zero, it indicates there is no over dispersion in the data, and data distribution approaches to the Poisson distribution. On the other hand, when the

Table 2.1: Simulation results for under-dispersed data ($\phi = 2.50$)

Model	Param	True	Mean	SE	MSE	CP (%)	LPML (IQR)	LOO (IQR)	WAIC (IQR)	DIC (IQR)
CMP	ϕ	2.50	2.82	0.61	0.48	91%	−99.93	−199.87	714.39	714.39
	β_0	−0.80	−0.87	0.18	0.01	94%	(3.45)	(6.90)	(26.45)	(26.67)
	β_1	−0.05	0.03	0.18	0.04	92%				
	β_2	−0.01	0.01	0.24	0.06	95%				
	σ_0	0.05	0.13	0.05	0.01	100%				
	σ_1	0.02	0.14	0.06	0.02	100%				
	ρ	−0.50	−0.02	0.02	0.22	100%				
NB	ϕ	-	0.04	0.00	-	-	−101.35	−200.71	713.01	732.27
	β_0	−0.80	−1.12	0.18	0.13	69%	(3.51)	(7.03)	(26.13)	(26.29)
	β_1	−0.05	0.02	0.14	0.02	95%				
	β_2	−0.01	0.02	0.19	0.04	97%				
	σ_0	0.05	0.07	0.01	0.00	100%				
	σ_1	0.02	0.08	0.01	0.00	100%				
	ρ	−0.50	−0.04	0.01	0.43	100%				
Poisson	β_0	−0.80	−1.11	0.18	0.13	69%	−100.33	−200.66	729.72	730.95
	β_1	−0.05	0.02	0.13	0.02	94%	(3.51)	(7.03)	(26.06)	(26.35)
	β_2	−0.01	0.02	0.19	0.04	96%				
	σ_0	0.05	0.07	0.01	0.00	100%				
	σ_1	0.02	0.07	0.01	0.00	100%				
	ρ	−0.50	−0.04	0.01	0.21	100%				

dispersion parameter in CMP equals to 1, it indicates there is no over dispersion in the data, and data distribution approaches to the Poisson distribution.

In the case of NB the regression coefficients are β_0=−1.12, β_1= 0.02, β_2= 0.02, while in the Poisson model the corresponding values are β_0=−1.11, β_1= 0.02, β_2= 0.02. In terms of parameter estimates we experience subtle differences between the two models. Sellers and Shmueli,(2010) reports that both Poisson and NB produce almost same regression parameter estimates for under-dispersed count response.

The posterior means of regression coefficient for intercept, treatment, and time in CMP model are β_0=−0.87, β_1=0.03,and β_2= 0.01. Apparently, it is observed that there are substantial differences in the parameter estimates between CMP and other two models. An approximate conversion (β/ϕ) of the CMP regression parameters enable a direct comparison among the models, and are almost accurate for larger count response. The converted values for CMP model $\beta_0/2.5$=−0.30, $\beta_1/2.5$=0.01,

and $\beta_2/2.5$=0.002 illustrate smaller effect sizes of the respective covariates on the response than that of produced by other models. However, this comparison may not be accurate as the under-dispersed simulation set up produces smaller counts in this case. We observe that MSE for intercept parameter in the CMP model is the lowest while for other parameters a bit higher than that of other models.

In comparison to the NB and Poisson model, the higher values of the estimates for standard deviations $\sigma_0 = 0.13$ and $\sigma_1 = 0.14$ for random intercept and slope in CMP model explain higher spread around the population level intercept and slope. The negative sign in the correlation coefficients ρ between random intercept and slope across models indicate that subjects having higher initial responses have slower rate of improvement.

Based on 95% credible interval, the coverage probabilities for all parameters are higher than 92% in CMP model while the same for both Poisson and NB are higher than 69%. Both credible interval and highest posterior density (HPD) (reported in Appendix B) interval for all parameters from all models are almost coincided which illustrates the symmetry of posterior means for respective parameters.

Although there exists no remarkable deviations in the values of LPML and LOO across three models, we observe that CMP model results the highest value of LPML (-99.93), and NB model results the lowest value of LOO (-200.71) . In contrast, CMP model results the lowest values of WAIC (714.394), and DIC (714.393) followed by Poisson model with WAIC (729.72), and DIC (730.95) respectively. It is observed that CMP model fits the best on the basis of LPML, WAIC, and DIC. On the other hand, NB model fits the best based on LOO with a negligible difference.

We applied CMP, NB, and Poisson model to simulated over-dispersed ($\phi = 0.30$) data and report the results in Table 2.2. We observe that the posterior mean of dispersion parameter in CMP model is ϕ=0.39 with a MSE=0.04, and reveals the existence of over-dispersion. The negative signs in treatment effect β_1=-0.01

Table 2.2: Simulation results for Over-dispersed data ($\phi = 0.30$)

Model	Param	True	Mean	SE	MSE	CP (%)	LPML (IQR)	LOO (IQR)	WAIC (IQR)	DIC (IQR)
CMP	ϕ	0.30	0.39	0.19	0.04	89%	−77.00	−154.69	1039.48	1038.40
	β_0	−0.80	−0.83	0.19	0.04	88%	(6.28)	(12.55)	(49.44)	(49.85)
	β_1	−0.05	−0.01	0.10	0.01	97%				
	β_2	−0.01	−0.02	0.16	0.03	91%				
	σ_0	0.05	0.16	0.05	0.01	100%				
	σ_1	0.02	0.17	0.05	0.03	98%				
	ρ	−0.50	−0.17	0.05	0.11	100%				
NB	ϕ	-	0.45	0.19	-	-	−77.00	−153.95	1040.76	1040.25
	β_0	−0.80	−0.58	0.18	0.08	72%	(6.11)	(12.38)	(47.78)	(47.92)
	β_1	−0.05	−0.01	0.13	0.02	96%				
	β_2	−0.01	−0.02	0.20	0.04	90%				
	σ_0	0.05	0.22	0.03	0.03	99%				
	σ_1	0.02	0.24	0.07	0.05	95%				
	ρ	−0.50	−0.18	0.06	0.11	100%				
Poisson	β_0	−0.80	−0.64	0.20	0.06	85%	−75.00	−150.29	1055.56	1045.37
	β_1	−0.05	−0.01	0.13	0.02	97%	(5.49)	(10.98)	(50.56)	(47.02)
	β_2	−0.01	0.00	0.21	0.04	89%				
	σ_0	0.05	0.38	0.16	0.13	76%				
	σ_1	0.02	0.41	0.17	0.18	44%				
	ρ	−0.50	−0.33	0.17	0.06	100%				

and time effect β_2=−0.02 and the corresponding transformed values ($\beta_1/0.39$=−0.03) and ($\beta_2/0.39$=−0.03) produced by CMP model indicate positive impacts of both treatment and time on count response, and they are markedly different than that of Poisson and NB model. The MSEs for β_1 and β_2 in CMP models are 0.01 and 0.03 respectively, also smaller than that of other models. The MSEs for σ_0 and σ_1 in CMP model are much smaller than NB and Poisson models respectively. Coverage probabilities in CMP model for all parameters are above 88% while the same for NB model are more than 72% and for Poisson more than 44%. We observe that Poisson model produces the highest value of LPML (−75.00) among the three models. On the contrary, the lowest values of LOO (−154.69), WAIC (1039.48), and DIC (1038.40) illustrate the best fit of CMP model among others. In addition, NB model fits better than the Poisson model in case of over-dispersed data which is expected.

Table 2.3: Simulation results for Equi-dispersed data ($\phi = 1.0$)

Model	Param	True	Mean	SE	MSE	CP (%)	LPML (IQR)	LOO (IQR)	WAIC (IQR)	DIC (IQR)
CMP	ϕ	1.00	1.23	0.30	0.14	84%	90.10	−180.25	859.80	859.48
	β_0	−0.80	−0.81	0.16	0.03	96%	(4.92)	(9.82)	(42.60)	(43.12)
	β_1	−0.05	0.00	0.15	0.02	94%				
	β_2	−0.01	−0.04	0.18	0.03	98%				
	σ_0	0.05	0.23	0.07	0.04	100%				
	σ_1	0.02	0.25	0.08	0.05	88%				
	ρ	−0.50	−0.18	0.05	0.11	100%				
NB	ϕ	-	0.07	0.03	-	-	90.50	−180.93	859.12	859.88
	β_0	−0.80	−0.85	0.15	0.02	96%	(4.91)	(9.81)	(42.51)	(43.30)
	β_1	−0.05	−0.00	0.14	0.02	94%				
	β_2	−0.01	−0.05	0.16	0.03	98%				
	σ_0	0.06	0.18	0.04	0.02	100%				
	σ_1	0.02	0.20	0.04	0.04	100%				
	ρ	−0.50	−0.16	0.03	0.12	100%				
Poisson	β_0	−0.80	−0.85	0.15	0.02	96%	90.30	−180.66	858.73	859.10
	β_1	−0.05	−0.00	0.14	0.02	94%	(4.78)	(9.45)	(41.65)	(42.97)
	β_2	−0.01	−0.05	0.16	0.03	98%				
	σ_0	0.05	0.19	0.05	0.02	100%				
	σ_1	0.02	0.21	0.05	0.03	98%				
	ρ	−0.50	−0.16	0.03	0.11	100%				

We report simulation results from equi-dispersed data in Table 2.3. We notice that Poisson model retains the lowest WAIC (858.73) and DIC (859.10) values than that of other models indicating the best fit while the second alternative is CMP model based on DIC. On the contrary, both LPML and LOO values illustrate that NB model fits better than others. However, differences in LPML and LOO across models are minuscule. MSE for parameters in Poisson model are smaller or at least equivalent to other models.

Both WAIC and DIC provide consistent conclusions regarding model fit for all models while both LPML and LOO provides contradictory conclusions.

Figure 2.2 presents boxplots for subject-wise CPO from 100 under-dispersed simulated samples. We produce similar graph for over- and equi-dispersed data cases too (reported in Appendix C). Extreme low values of CPO indicate that the respective data points are outliers. We observe that out of 100 subjects very few retain extreme

low CPO values for some samples, which indicate that simulation data are more or less homogeneous.

2.5 Sensitivity Analysis

We perform a sensitivity analysis for the over-dispersed ($\phi = 0.30$) simulated data by changing hyper parameters in priors as well as prior distributions to investigate their influences on posteriors means. The results are presented in Table 2.4 and Table 2.5 respectively. We apply CMP model with correlated random intercept and slope with the following prior specifications $\beta_k \sim N(0, 10000), \phi \sim LN(0, 15), \zeta \sim N(0, \sigma_\zeta^2), \sigma_\zeta^2 \sim U(0, \infty), \boldsymbol{L} \sim LKJ(2)$. Then we change parameter values in five scenarios for each of the cases from O_1 to O_4 (see Table 2.4).

In Table 2.4 we observe that posterior means of the regression parameters are almost identical for all scenarios (S_1 to S_5) for the cases O_1 to O_4. In O_4, we notice that for different choices of parameter η in LKJ distribution the posterior means of the parameters σ_0, σ_1, and ρ vary remarkably. In particular, we experience notable deviations in these parameter values when η moves downward from 1.5 to 0.5, meaning that when density moves towards uniformity.

Table 2.5 illustrates the effect of prior changes on posterior means. In scenarios S_1 to S_5, we consider defuse normal prior for regression co-efficients, gamma or half-Cauchy prior for ϕ in place of log normal, and inverse gamma or half-Cauchy for σ_0^2 and σ_1^2 instead of uniform, and experience no remarkable changes in regression co-efficients. However, we notice substantial changes in standard deviations and correlation parameters across scenarios. It is evident that regression estimates are not much sensitive to the priors while estimates for σ_0, σ_1, ϕ, and ρ are remarkably sensitive.

Table 2.4: Posterior summary for over-dispersed data with different choices of hyper-parameters in priors

Scenarios	Posterior means						
	$\hat{\beta_0}$ (95%HPD)	$\hat{\beta_1}$ (95%HPD)	$\hat{\beta_2}$ (95%HPD)	$\hat{\sigma_0}$ (95%HPD)	$\hat{\sigma_1}$ (95%HPD)	$\hat{\rho}$ (95%HPD)	$\hat{\phi}$ (95%HPD)
O_1 :**Parameters for β_k vary**							
$S_1 : \beta_k \sim N(0, 10000)$	−0.71	0.01	−0.16	0.29	0.33	−0.22	0.40
	(−1.01, −0.42)	(−0.22, 0.22)	(−0.43, 0.10)	(0.00, 0.79)	(0.00, 0.92)	(−0.98, 0.62)	(0.01, 0.70)
$S_2 : \beta_k \sim N(0, 100)$	−0.71	0.01	−0.16	0.26	0.31	−0.19	0.39
	(−1.01, −0.42)	(−0.22, 0.22)	(−0.41, 0.08)	(0.00, 0.65)	(0.00, 0.79)	(−0.94, 0.64)	(0.01, 0.68)
$S_3 : \beta_k \sim N(0, 50)$	−0.70	0.01	−0.16	0.28	0.33	−0.20	0.40
	(−1.01, −0.42)	(−0.22, 0.22)	(−0.41, 0.09)	(0.00, 1.72)	(0.00, 0.86)	(−0.94, 0.67)	(0.01, 0.70)
$S_4 : \beta_k \sim N(0, 3)$	−0.71	0.01	−0.16	0.28	0.34	−0.22	0.40
	(−0.99, −0.42)	(−0.22, 0.22)	(−0.42, 0.08)	(0.00, 0.75)	(0.00, 0.89)	(−0.94, 0.67)	(0.01, 0.70)
$S_5 : \beta_k \sim N(0, 2)$	−0.70	0.00	−0.17	0.29	0.35	−0.21	0.41
	(−1.00, −0.41)	(−0.21, 0.22)	(−0.41, 0.10)	(0.00, 0.78)	(0.00, 0.97)	(−0.97, 0.65)	(0.06, 0.75)
O_2 :**Parameters for ϕ vary**							
$S_1 : \phi \sim LN(0, 100)$	−0.70	0.01	−0.16	0.27	0.33	−0.20	0.40
	(−1.00, −0.42)	(−0.21, 0.23)	(−0.41, 0.10)	(0.00, 0.70)	(0.00, 0.84)	(−0.94, 0.66)	(0.02, 0.71)
$S_2 : \phi \sim LN(0, 30)$	−0.71	0.01	−0.16	0.29	0.34	−0.21	0.40
	(−0.99, −0.40)	(−0.21, 0.23)	(−0.41, 0.11)	(0.00, 0.77)	(0.00, 0.91)	(−0.95, 0.66)	(0.01, 0.70)
$S_3 : \phi \sim LN(0, 10)$	−0.70	0.01	−0.17	0.26	0.32	−0.20	0.40
	(−0.99, −0.40)	(−0.21, 0.23)	(−0.40, 0.10)	(0.00, 0.69)	(0.00, 0.81)	(−0.97, 0.64)	(0.02, 0.71)
$S_4 : \phi \sim LN(0, 3)$	−0.70	0.01	−0.16	0.27	0.33	−0.21	0.41
	(−1.00, −0.42)	(−0.21, 0.22)	(−0.42, 0.10)	(0.01, 0.96)	(0.01, 1.14)	(−0.94, 0.73)	(0.07, 0.75)
$S_5 : \phi \sim LN(0, 0.5)$	−0.63	−0.01	−0.18	0.34	0.40	−0.24	0.56
	(−0.92, −0.34)	(−0.22, 0.23)	(−0.43, 0.09)	(0.00, 0.85)	(0.00, 1.00)	(−0.98, 0.61)	(0.31, 0.83)
O_3 :**Parameters for σ_ζ^2 vary**							
$S_1 : \sigma_\zeta^2 \sim U(0, 1000)$	−0.70	0.01	−0.16	0.30	0.36	-0.23	0.41
	(−1.00, −0.40)	(−0.21, 0.23)	(−0.42, 0.09)	(0.00, 0.82)	(0.00, 0.98)	(−0.97, 0.61)	(0.02, 0.74)
$S_2 : \sigma_\zeta^2 \sim U(0, 30)$	−0.70	0.01	−0.16	0.31	0.36	-0.22	0.41
	(−1.00, −0.41)	(−0.22, 0.22)	(−0.41, 0.09)	(0.00, 0.98)	(0.00, 1.10)	(−0.97, 0.63)	(0.01, 0.75)
$S_3 : \sigma_\zeta^2 \sim U(0, 10)$	−0.71	0.01	−0.16	0.26	0.32	-0.20	0.39
	(−1.01, −0.41)	(−0.21, 0.23)	(−0.40, 0.10)	(0.00, 0.68)	(0.00, 0.82)	(−0.99, 0.64)	(0.01, 0.68)
$S_4 : \sigma_\zeta^2 \sim U(0, 3)$	−0.71	−0.01	−0.16	0.29	0.31	-0.22	0.41
	(−1.00, −0.42)	(−0.21, 0.23)	(−0.42, 0.11)	(0.00, 0.81)	(0.00, 0.98)	(−0.99, 0.63)	(0.01, 0.72)
$S_5 : \sigma_\zeta^2 \sim U(0, 1)$	−0.71	0.01	−0.16	0.25	0.30	-0.19	0.39
	(−1.01, −0.43)	(−0.22, 0.23)	(−0.42, 0.08)	(0.00, 0.63)	(0.00, 0.73)	(−0.96, 0.64)	(0.01, 0.68
O_4 :**Parameters for $\boldsymbol{L}$ vary**							
$S_1 : \boldsymbol{L} \sim LKJ(5)$	−0.70	0.01	−0.17	0.23	0.28	−0.08	0.40
	(−0.99, −0.41)	(−0.22, 0.22)	(−0.42, 0.07)	(0.00, 0.55)	(0.00, 0.67)	(−0.70, 0.50)	(0.01, 0.68)
$S_2 : \boldsymbol{L} \sim LKJ(3)$	−0.70	0.01	−0.17	0.25	0.30	−0.13	0.39
	(−1.01, −0.41)	(−0.21, 0.22)	(−0.42, 0.09)	(0.00, 0.62)	(0.00, 0.62)	(−0.84, 0.62)	(0.05, 0.72)
$S_3 : \boldsymbol{L} \sim LKJ(1.5)$	−0.71	0.01	−0.16	0.31	0.38	−0.27	0.41
	(−1.01, −0.42)	(−0.21, 0.23)	(−0.42, 0.09)	(0.00, 0.85)	(0.00, 1.02)	(−1.00, 0.68)	(0.01, 0.73)
$S_4 : \boldsymbol{L} \sim LKJ(1)$	−0.72	0.01	−0.14	0.36	0.43	−0.38	0.41
	(−1.01, −0.42)	(−0.21, 0.21)	(−0.39, 0.13)	(0.00, 0.98)	(0.00, 1.19)	(−1.00, 0.75)	(0.04, 0.73)
$S_5 : \boldsymbol{L} \sim LKJ(0.5)$	−0.73	0.01	−0.11	0.55	0.67	−0.61	0.45
	(−1.06, −0.44)	(−0.21, 0.24)	(−0.41, 0.20)	(0.00, 1.53)	(0.00, 1.80)	(−1.00, 0.83)	(0.01, 1.02)

To compare the regression parameter estimates across scenarios, we construct forest plots, illustrated in Figures 2.3 to 2.7, for 95% HPD intervals and their respective mean values (solid circles) obtained from Table 2.4 and Table 2.5. All the figures evidence that posterior means for regression parameters are quite robust across scenarios.

2.6 Application to Randomized Controlled Clinical trials Data

To demonstrate the applicability of our proposed model we use data from two distinct clinical trials. The first example involves data from placebo-controlled clinical trial of epilepsy patients, and the second example relates to multi-vitamin supplement in HIV patients. We report a short description for each dataset along with corresponding analysis results.

2.6.1 Analysis of Epilepsy data

We fit models with random effects on Epilepsy data discussed in Thall & Vail (1990). The dataset consists of the number of seizures for 59 patients suffering from epilepsy, 31 of them are assigned to the progabide group, the treatment arm, and the rest in the placebo group. Seizure rates are longitudinally measured in an initial eight weeks before baseline and then in every two weeks in four consecutive treatment periods. The mean and variance of the seizure rate is 12.85 and 349.17 respectively, potentially an indication of over-dispersion in the data. Figure 2.8 presents the individual profile of seizure rates for selected subjects from placebo and progabide group. The first measurement at zero is the number of seizures in eight weeks intervals before randomization and the rests are in every two weeks. We illustrate the treatment arm-wise profiles in Figure 2.9, and notice that one subject in progabide group retains

Table 2.5: Posterior summary for over-dispersed data with different choices of prior distributions

Scenarios	Posterior means						
	$\hat{\beta_0}$ (95%HPD)	$\hat{\beta_1}$ (95%HPD)	$\hat{\beta_2}$ (95%HPD)	$\hat{\sigma_0}$ (95%HPD)	$\hat{\sigma_1}$ (95%HPD)	$\hat{\rho}$ (95%HPD)	$\hat{\phi}$ (95%HPD)
O_5 :**Priors vary**							
$S_1 : \beta_k \sim N(0, 10000), \phi \sim \Gamma(2, 2),$							
$\zeta \sim N(0, \sigma_\zeta^2), \sigma_\zeta^2 \sim IG(0.1, 0.1), \boldsymbol{L} \sim LKJ(2)$	-0.67	0.01	-0.17	0.22	0.26	-0.13	0.46
	$(-0.94, -0.39)$	$(-0.22, 0.22)$	$(-0.41, 0.09)$	(0.02, 0.53)	(0.02, 0.64)	$(-0.94, 0.66)$	(0.17, 0.74)
$S_2 : \beta_k \sim N(0, 10000), \phi \sim \Gamma(3, 3),$							
$\zeta \sim N(0, \sigma_\zeta^2), \sigma_\zeta^2 \sim IG(0.1, 0.1), \boldsymbol{L} \sim LKJ(2)$	-0.65	0.01	-0.17	0.49	0.56	-0.45	0.55
	$(-0.94, -0.35)$	$(-0.24, 0.24)$	$(-0.43, 0.12)$	(0.15, 0.95)	(0.15, 1.11)	$(-0.99, 0.34)$	(0.25, 0.84
$S_3 : \beta_k \sim N(0, 10000), \phi \sim \Gamma(2, 2),$							
$\zeta \sim N(0, \sigma_\zeta^2), \sigma_\zeta^2 \sim Cauchy(0, 10), \boldsymbol{L} \sim LKJ(2)$	-0.66	0.01	-0.17	0.32	0.39	-0.24	0.51
	$(-0.93, -0.36)$	$(-0.22, 0.24)$	$(-0.45, 0.09)$	(0.00, 0.86)	(0.00, 1.06)	$(-0.97, 0.63)$	(0.23, 0.81)
$S_4 : \beta_k \sim N(0, 10000), \phi \sim Cauchy(0, 10),$							
$\zeta \sim N(0, \sigma_\zeta^2), \sigma_\zeta^2 \sim IG(2, 2), \boldsymbol{L} \sim LKJ(2)$	-0.67	0.01	-0.15	0.60	0.69	-0.58	0.55
	$(-0.97, -0.36)$	$(-0.23, 0.24)$	$(-0.42, 0.14)$	(0.24, 1.07)	(0.23, 1.24)	(-0.99, 0.09)	(0.23, 0.87)
$S_5 : \beta_k \sim N(0, 10000), \phi \sim Cauchy(0, 10),$							
$\zeta \sim N(0, \sigma_\zeta^2), \sigma_\zeta^2 \sim Cauchy(0, 10), \boldsymbol{L} \sim LKJ(2)$	-0.67	0.01	-0.17	0.31	0.37	-0.23	0.48
	$(-0.95, -0.38)$	$(-0.23, 0.22)$	$(-0.43, 0.08)$	(0.00,0.78)	(0.00, 0.92)	$(-0.96, 0.64)$	(0.16, 0.80)

extreme number of seizures. Trends in seizure rates are not obvious from both the graphs. In Figure 2.10, we present treatment arm-wise average weekly seizure rates, and observe that at the end of eight weeks subjects in progabide group experience lesser seizure rates.

Time is an indicator variable of a period after baseline (0 if baseline, 1 if after baseline), trt is defined as 1 if a patient receives an anti-epileptic drug (progabide) and 0 if placebo, T_{ij} is the offset, length of time period in weeks (8 if baseline, 2 if after baseline). By observing the profiles, it is reasonable to assume that there is a natural heterogeneity among subjects both in their baseline level and in the changes in expected counts over time, rationalizes the inclusion of subject-wise random effects in the model.

We consider three models and mention them one by one:

$\mathcal{M}_1$: A mixed effect model with subject-specific random intercept, and we specify the model as

$$\log(\theta_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\zeta}_i, \tag{2.22}$$

where $\boldsymbol{x}_{ij}^T = (1, trt_{ij}, time_{ij}, trt_{ij} \times time_{ij}), \boldsymbol{z}_{ij}^T = 1, \boldsymbol{\zeta}_i = (\zeta_{0i})$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$. More specifically,

$$\log(\theta_{ij}) = (\beta_0 + \zeta_{0i}) + \beta_1 \times trt_{ij} + \beta_2 \times time_{ij} + \beta_3 \times trt_{ij} \times time_{ij} + log(T_{ij}) \tag{2.23}$$

$\mathcal{M}_2$: A mixed effect model with uncorrelated random intercept and slope. With the general model noted in (2.22) the specifications are:
$\boldsymbol{x}_{ij}^T = (1, trt_{ij}, time_{ij}, trt_{ij} \times time_{ij}), \boldsymbol{z}_{ij}^T = (1, time_{ij}), \boldsymbol{\zeta}_i = (\zeta_{0i}, \zeta_{1i})^T$, $\boldsymbol{\beta} =$

$(\beta_0, \beta_1, \beta_2, \beta_3)^T$, and the variance-covariance matrix for random effects

$$\boldsymbol{D} = \begin{bmatrix} \sigma_0^2 & 0 \\ 0 & \sigma_1^2 \end{bmatrix}.$$

In other words,

$$\log(\theta_{ij}) = (\beta_0 + \zeta_{0i}) + \beta_1 \times trt_{ij} + (\beta_2 + \zeta_{1i}) \times time_{ij} + \beta_3 \times trt_{ij} \times time_{ij} + log(T_{ij}) \quad (2.24)$$

$\mathcal{M}_3$: A mixed effect model with correlated random intercept and slope. Following the similar notations in $\mathcal{M}_1$ and $\mathcal{M}_2$ the specifications are:
$\boldsymbol{x}_{ij}^T = (1, trt_{ij}, time_{ij}, trt_{ij} \times time_{ij}), \boldsymbol{z}_{ij}^T = (1, time_{ij}), \boldsymbol{\zeta}_i = (\zeta_{0i}, \zeta_{1i})^T$,
$\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$, and the variance-covariance matrix for random effects

$$\boldsymbol{D} = \begin{bmatrix} \sigma_0^2 & \rho\sigma_0\sigma_1 \\ \rho\sigma_0\sigma_1 & \sigma_1^2 \end{bmatrix}.$$

In particular,

$$\log(\theta_{ij}) = (\beta_0 + \zeta_{0i}) + \beta_1 \times trt_{ij} + (\beta_2 + \zeta_{1i}) \times time_{ij} + \beta_3 \times trt_{ij} \times time_{ij} + log(T_{ij}), \quad (2.25)$$

In our analysis, we consider the following prior distributions:

$$\phi \sim LN(0, 15)$$
$$\boldsymbol{L} \sim LKJ(2)$$
$$\zeta_q \sim N(0, \sigma_{\zeta_q}^2), \sigma_{\zeta_q}^2 \sim IG(0.1, 0.1) \text{ for } q = 0, 1$$
$$\beta_k \sim N(0, 10000) \text{ for } k = 0, 1, 2, 3$$

To fit the models $\mathcal{M}_1$ and $\mathcal{M}_2$ we do not require prior for $\boldsymbol{L}$ as the models do not include correlation between random effects. We assume that β_k can take any value with normal mean 0 and standard deviation 10000. The dispersion parameter $\phi > 0$

and we assign a log normal prior with mean zero and moderate standard deviation for ϕ. Choo-Wosoba et al. (2018) use log normal distribution as a prior for ϕ. Since $\sigma^2_{\zeta_i}$ is a scale parameter with a lower bound of zero, we assign a inverse-gamma prior for $\sigma^2_{\zeta_i}$, as suggested in Gelman et al. (2006). For the correlated random effects we assume that the random effects are weakly correlated, the value $\eta = 2$, indicates that the correlation ρ is close to zero. We perform Bayesian analysis by using Stan language (named after Stanislaw Ulam, a mathematician) and rstan (R package) in four chains with 5000 iterations having 2500 warm-up each. Since, from the simulation study we obtain intuitive and consistent results across models for DIC, we report DIC for $\mathcal{M}_1$ and $\mathcal{M}_2$. For $\mathcal{M}_3$ we report LPML, LOO, WAIC, and DIC for the model assessments among the competing models CMP, NB and Poisson.

Table 2.6: Posterior summary under $\mathcal{M}_1$

Parameter	Poisson			NB			CMP			
	Mean	Std	95% HPD	Mean	Std	95% HPD	Mean	Std	95% HPD	$\hat{R}$
β_0	1.03	0.16	(0.71,1.34)	1.09	0.18	(0.74,1.45)	−1.19	0.13	(−1.44,−0.93)	1
β_1	0.12	0.05	(0.02,0.21)	0.02	0.10	(−0.18,0.22)	−0.01	0.08	(−0.17,0.14)	1
β_2	−0.03	0.22	(−0.47,0.42)	0.08	0.25	(−0.41,0.58)	0.98	0.05	(0.87,1.0)	1
β_3	−0.10	0.06	(−0.23,0.02)	−0.32	0.14	(−0.60,−0.04)	−0.04	0.04	(−0.11,0.03)	1
σ	0.81	0.08	(0.66,0.98)	0.85	0.09	(0.68,1.02)	0.27	0.04	(0.20,0.36)	1
ϕ	-	-	-	0.15	0.03	(0.11,0.21)	0.30	0.04	(0.23,0.37)	1
DIC	6591.67			3291.92			2604.62			
$\overline{\Delta(\boldsymbol{\Theta})}$	6439.85			3179.1			2561.15			
p_D(approx.)	152			113			44			

The output of the models $\mathcal{M}_1$, $\mathcal{M}_2$, and $\mathcal{M}_3$ are reported in Tables 2.6, 2.7, and 2.8 respectively. Table 2.6 presents posterior mean of the parameters obtained from $\mathcal{M}_1$ for Poisson, NB , and CMP model, where σ indicates the standard deviation of subject specific random intercept. The results show that $\sigma = 0.27$ is the least for CMP model. Posterior mean for dispersion parameter ϕ both in NB ($\phi = 0.15$) and in CMP ($\phi = 0.30$) postulate that the study data are over-dispersed. Although not significant, the value β_1=−0.01, 95% HPD: (−0.17,0.14) in CMP model shows that progabide has positive impact on number of seizures. The negative sign for the interaction co-efficients in all three models reveal the beneficial effect of study drug in

reducing number of seizures for epilepsy patients. The values $\hat{R} = 1$ for all parameters in CMP model express that the posterior is close to the target distribution. The least value of DIC (2604.62) for CMP illustrates the best fit in comparison to NB and Poisson models.

Table 2.7: Posterior summary under $\mathcal{M}_2$

Parameter	Poisson			NB			CMP			
	Mean	Std	95% HPD	Mean	Std	95% HPD	Mean	Std	95% HPD	$\hat{R}$
β_0	1.05	0.15	(0.77,1.35)	1.09	0.16	(0.77,1.41)	-0.70	0.19	(−1.07,−0.31)	1
β_1	0.02	0.12	(−0.21,24)	0.00	0.12	(−0.24,0.25)	0.02	0.10	(−0.17,0.21)	1
β_2	0.06	0.20	(−0.34,0.47)	0.06	0.23	(−0.37,0.50)	0.73	0.09	(0.54,0.91)	1
β_3	−0.31	0.16	(−0.62,0.00)	−0.32	0.17	(−0.66, 0.02)	−0.14	0.09	(−0.32,0.03)	1
σ_0	0.75	0.08	(0.61,0.91)	0.75	0.09	(0.58,0.93)	0.34	0.06	(0.24,0.46)	1
σ_1	0.52	0.07	(0.39,0.65)	0.42	0.11	(0.22,0.62)	0.24	0.05	(0.16,0.34)	1
ϕ	-	-	-	0.11	0.03	(0.07,0.17)	0.44	0.06	(0.34,0.56)	1
DIC	7612.10			3767.87			2899.88			
$\overline{\Delta(\boldsymbol{\Theta})}$	7312.48			3617.93			2861.42			
p_D(approx.)	240			150			39			

We report posterior summary of the parameters for $\mathcal{M}_2$ from Poisson, NB , and CMP models in Table 2.7. Here we consider intercept and slope are independent. We observe that $\sigma_0 = 0.34$ and $\sigma_1 = 0.24$ are the least(s) for CMP model. The values $\phi = 0.11$ for NB and $\phi = 0.44$ for CMP evidence the existence of over-dispersion in the data. The negative sign for the interaction between progabide and time coefficients in all three models reveal the beneficial effect of study drug in reducing number of seizures for epilepsy patients, although not significant in case of CMP. The values $\hat{R} = 1$ for all parameters in CMP model express good performance of MCMC samples. Likewise $\mathcal{M}_1$, the least value of DIC (2899.88) for CMP illustrates the best fit of the model.

Posterior summary of the parameters for $\mathcal{M}_3$ from Poisson, NB, and CMP models is presented in Table 2.8. The model considers correlated random intercept and slope. We observe that $\sigma_0 = 0.30$ and $\sigma_1 = 0.22$ are the least(s) for CMP model. The standard deviation of posterior means for all parameters in CMP model are the least. The CMP model shows a significant correlation between random intercept and

Table 2.8: Posterior summary under $\mathcal{M}_3$

Parameter	Poisson			NB			CMP			
	Mean	Std	95% HPD	Mean	Std	95% HPD	Mean	Std	95% HPD	$\hat{R}$
β_0	1.07	0.15	(0.78,1.34)	1.12	0.15	(0.83,1.1.42)	-0.79	0.19	$(-1.14,-0.42)$	1
β_1	0.05	0.20	$(-0.33,-0.45)$	0.05	0.21	$(-0.36,0.46)$	0.01	0.09	$(-0.16,0.19)$	1
β_2	0.00	0.12	$(-0.23,0.23)$	-0.03	0.12	$(-0.26,0.20)$	0.74	0.09	(0.56,0.91)	1
β_3	-0.31	0.16	$(-0.63,0.02)$	-0.32	0.17	$(-0.64,0.01)$	-0.13	0.08	$(-0.29,0.03)$	1
σ_0	0.75	0.08	(0.60,0.90)	0.68	0.09	(0.51,0.86)	0.30	0.05	(0.20,0.41)	1
σ_1	0.52	0.07	(0.39,0.65)	0.37	0.09	(0.19,0.54)	0.22	0.04	(0.14,0.31)	1
ρ	0.14	0.16	$(-0.17,0.43)$	0.54	0.22	(0.06,0.91)	0.45	0.18	(0.08,0.78)	1
ϕ	-	-	-	0.12	0.03	(0.08,0.17)	0.41	0.06	(0.30,0.52)	1
LPML	-77.29			-63.29			-25.90			
LOO	912.100			746.84			305.60			
WAIC	13,694.43			15,802.57			5,315.11			
DIC	6,861.52			3,581.57			2,900.48			
$\overline{\Delta(\boldsymbol{\Theta})}$	6727.94			3463.82			2878.90			
p_D	134			118			22			

slope resulting by $\rho = 0.45$ with a 95% HPD:(0.08, 0.78). The values $\hat{R} = 1$ for all parameters in CMP model indicate good mixing of the MCMC samples. The highest value of LPML (-25.90) and the least values of LOO (305.59), WAIC (5315.11), and DIC (2900.46) for CMP model illustrate the best fit of the model among others.

The values $\phi = 0.12$ for NB and $\phi = 0.41$ for CMP illustrate that data are over-dispersed. Although not significant, the negative sign for the regression coefficients of the interaction between progabide and time in three models illustrate that progabide has a positive impact in reducing number of seizures for epilepsy patients over time. The regression parameter estimates in three models are not directly comparable since θ_i does not represent mean of CMP as it does for NB and Poisson distributions. A transformation, β_k/ϕ provides a crude comparison with β_ks from Poisson and NB model (K. F. Sellers & Shmueli, 2010), which is almost accurate for larger counts having mean greater than 10. Therefore, the converted coefficients for CMP are $\beta_0/0.41=-1.93$, $\beta_1/0.41=0.02$, $\beta_2/0.41=1.80$, and $\beta_3/0.41=-0.317$. The coefficients across models are quite different except for the interaction term. The interaction coefficients in Poisson, NB, and CMP reveal 26.65%, 27.38% , and 27.17% reduction

respectively in the seizure counts in progabide group from baseline to post baseline in comparison to placebo group.

Although we produce trace plots, pair plots, and ACF plots for all parameters form the three models $\mathcal{M}_1$ to $\mathcal{M}_3$ in CMP, NB, and Poisson model setting, we report here only for CMP model under $\mathcal{M}_3$. The trace plots in Figure 2.8 illustrates the good mixing and convergence of MCMC samples for all parameters. The pair plots in Figure 2.9 reveals density of the posterior means, and evidences no issues in MCMC samples. The exponential shapes of ACF plots in Figure 2.10 for the regression coefficients divulge a sign of non-auto correlated samples generation in MCMC.

2.6.2 Analysis of Multivitamin Supplementation in HIV Infected Adults Data

We illustrate another application of the proposed model to the data from a longitudinal randomized double-blinded placebo controlled clinical trial. The trial explores the beneficial effect of multivitamin among HIV-infected adults receiving highly active antiretroviral therapy (HAART) in Uganda reported in Guwatudde et al. (2012). The adults received either a multivitamin (MV) supplement (including vitamin B-complex, C, and E) or placebo. We consider 354 subjects each of having measurements at visits 3, 6, 12, and 18 months for the analysis data set. The number of missing pills (mean =14.03, variance=43.55) during last month of each visit as an indirect measure of non-adherence to the study medication is considered as a response variable, and trt (MV or Placebo) and weight measures of the subject at each visit are taken as the covariates. We run the CMP model with uncorrelated random intercept and slope with the similar notation mentioned in Section 2.6.1.

The form of the subject specific random effects models is

$$\log(\theta_{ij}) = (\beta_0+\zeta_{0i})+\beta_1\times trt_{ij}+(\beta_2+\zeta_{1i})\times time_{ij}+\beta_3\times\log(weight_{ij})+log(T_{ij}), \quad (2.26)$$

where T_{ij} is the offset term. As an offset, we consider $\log 3$ for the visits at 3, and 6 months, and $\log 6$ for the remaining two visits.

We conduct the Bayesian analysis with the following prior specifications

$$\phi \sim LN(0, 15)$$
$$\zeta_q \sim N(0, \sigma^2_{\zeta_q}), \sigma^2_{\zeta_q} \sim IG(0.1, 0.1) \text{ for } q = 0, 1$$
$$\beta_k \sim N(0, 10000) \text{ for } k = 0, 1, 2, 3$$

Table 2.9: Posterior summary under the mixed effect model on Multivitamin supplementation in HIV infected adults data

Parameter	CMP			
	Mean	Std	95% HPD	$\hat{R}$
β_0	0.46	0.17	(0.14, 0.80)	1
β_1	−0.02	0.02	(−0.05, 0.01)	1
β_2	−0.33	0.01	(−0.34, −0.31)	1
β_3	−0.01	0.04	(−0.09, 0.07)	1
σ_0	0.07	0.01	(0.04, 0.09)	1
σ_1	0.06	0.01	(0.05, 0.08)	1
ϕ	0.54	0.02	(0.50, 0.57)	1

From Table 2.9 we see, although not statistically significant, those who are receiving multivitamin (β_1=−0.02, 95% HPD: (−0.05, 0.01)), and gaining weight (β_3=−0.01, 95% HPD: (−0.09, 0.07)) are less likely to miss the intervention medication, indicating better adherence to the study medication. As the time of intervention goes up adherence to the study medication (β_2=−0.33, 95% HPD: (−0.34, −0.31)) significantly increases. The values of $\hat{R}s = 1$ for all parameters reveal good mixing of the MCMC samples, and convergence of the model.

2.7 Discussion

K. F. Sellers & Shmueli (2010) mentions, in terms of model fit and predictive power, CMP model outperforms NB and Poisson by having its ability to account for a wide variety of dispersion conditions in a parsimonious way. Due to the longer tail, CMP model can capture extreme observations. As special cases, a number of data distributions such as Poisson, geometric, Bernoulli distributions can be generated from CMP distribution. Instead of fitting separate models, the CMP model enables us to fit a single model for various dispersion conditions. This distinctive feature introduces CMP model as a flexible regression model for count data. In this study we propose a novel Bayesian approach to fit a CMP generalized mixed effect model by using No-U-Turn Sampling (NUTS), a variant of Hamiltonian MCMC. In particular, we incorporate random intercept and slope in CMP mixed effect model to capture subject specific heterogeneity and inherent dispersion prevailing in the longitudinal count data. We examine the model performance based on both simulated and real data by using four Bayesian model assessment criteria namely, LPML, LOO, WAIC, and DIC. Especially, we simulate data in under-, over-, and equi- dispersed conditions and apply Poisson, negative binomial, and CMP model in each of the dispersion conditions.

We experience from both simulated and real data analysis that each of the model assessment criteria does not equally perform in all situations, similar evidence mentioned in Vehtari et al. (2017). From the simulated data we see that WAIC and DIC perform consistently in all situations. On the contrary, both LOO and LPML performances are conflicting in case of under-and equi-dispersed data with subtle deviations. We experience, in case of epilepsy data all four measurements are giving similar conclusion. A close look to CPO calculation for epilepsy data reveals that for some data points it provides unusual value for CPO which impacts LPML. Calculation of Monte-Carlo version of LOO is also similar to CPO calculation, and

probably that is why they behave similarly. In addition, WAIC involves calculation of variance of log posterior likelihood across simulations is not reliable when it exceeds 0.04 (Vehtari et al., 2017). Therefore, it is reasonable to rely on DIC, at least for this case as it is widely used method although retains some limitations (Vehtari et al., 2017; Plummer, 2008; Van Der Linde, 2005).

In Our study, based on DIC, both the simulation and real data analysis reveal that CMP model fits better than NB and Poisson for over- or under-dispersed data. In case of under-dispersed simulated data, surprisingly, we observe that parameter estimates in Poisson and NB model almost equivalent, while a bit different than that of CMP model. K. F. Sellers & Shmueli (2010) found similar results in their analysis. Therefore, it is reasonable to say that in case of under-dispersed data both Poisson and NB produce similar results. The CMP model retains the lowest WAIC and DIC but MSEs for the parameters are higher or equal except for the intercept in comparison to other models. Poisson and NB are not optimal model for under-dispersed data. Although predictive performance is not satisfactory, based on DIC and WAIC, the CMP model may be viewed as the best model followed by Poisson. However, an extensive simulation study by considering a wide range of under-dispersion levels and true effect size may result different conclusion. In case of over-dispersed simulated data we experience that CMP is the best model followed by NB. Poisson seems to be the best model for equi-dispersed data while CMP is the second alternative.

The regression parameter estimates in three models are not directly comparable since θ_i does not represent mean of CMP distribution like of NB and Poisson distributions. A conversion, β_k/ϕ provides an approximate comparison with β_ks from Poisson and NB model (K. F. Sellers & Shmueli, 2010) for high counts. After transformation, head to head comparisons reveal that parameter estimates are different at least for some cases. By choosing an incorrect model there remains potential chance of loosing information on the effect sizes of the covariates of interest.

The sensitivity analysis with over-dispersed data illustrates that when LKJ density approaches to uniform type of shape the posterior means of standard deviations and correlation become more sensitive while a little sensitivity is noticed in case of regression coefficients.

Epilepsy data are well studied in literature, and results are known regarding dispersion and efficacy of progabide, a study drug. We examine whether our proposed model could produce the similar results. The data are over-dispersed, and most of the studies resulted non-efficacy of the progabide (Leppik et al., 1987). Our model sufficiently produced the similar results to the previous studies. The model fitting on multivitamin data to assess the adherence to medication in terms of missing pills also depicts good convergence.

In epilepsy data analysis we encounter some unusual CPO values, and high values of variance of log posterior likelihood across simulations that makes the reliability of LPML and WAIC questionable, need further exploration. In this study we did not deal with zero inflated longitudinal counts, missing mechanism of longitudinal data, and modeling duel links (modeling log link of dispersion parameter) in the context of CMP distribution which remains for further extension. In all cases the model checking by Rhat, trace plots, pair plots,ACF plots were quite satisfactory.

With the advent of computation performed in this study, dealing with count response, the proposed model is easily extendable to study subject and cluster specific variability in multi-site clinical trials by adding cluster or site specific random effects in the model. The model would be potentially useful and superior in biomedical, public health, and business research while dealing with dispersed periodic count responses.

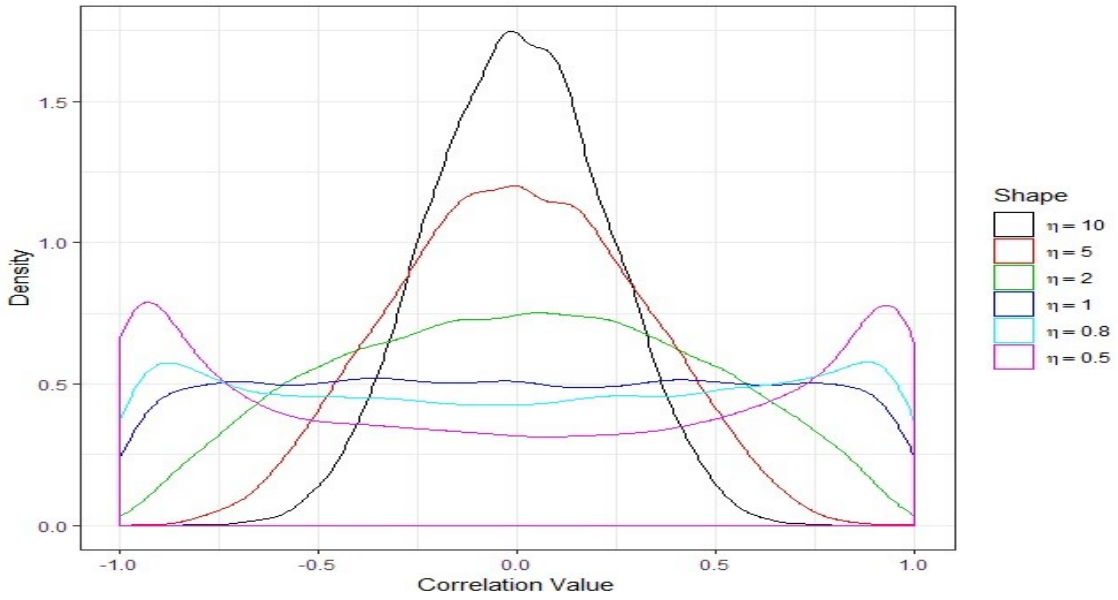

Figure 2.1: LKJ density plot for 2×2 Correlation Matrix $\mathbf{\Omega}$

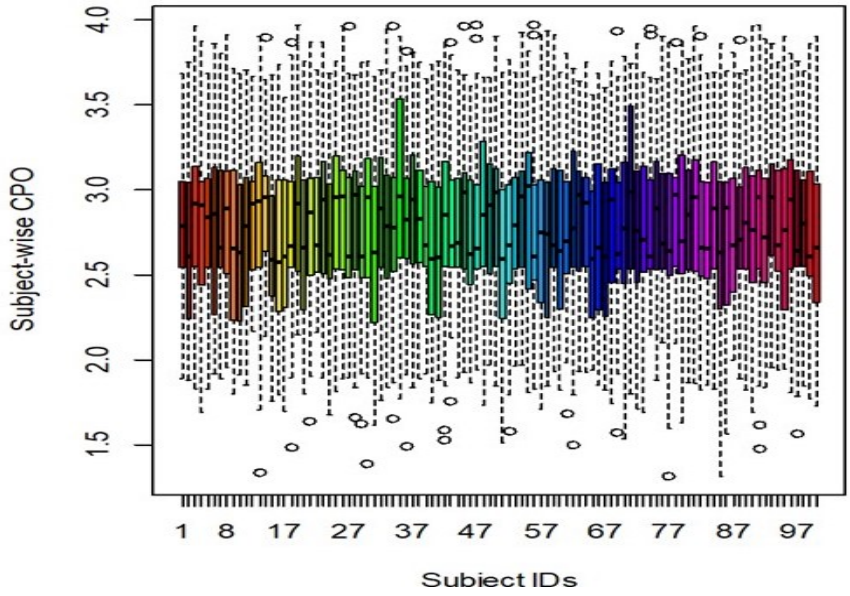

Figure 2.2: Boxplot for subject-wise CPO for CMP model with $\phi = 2.50$

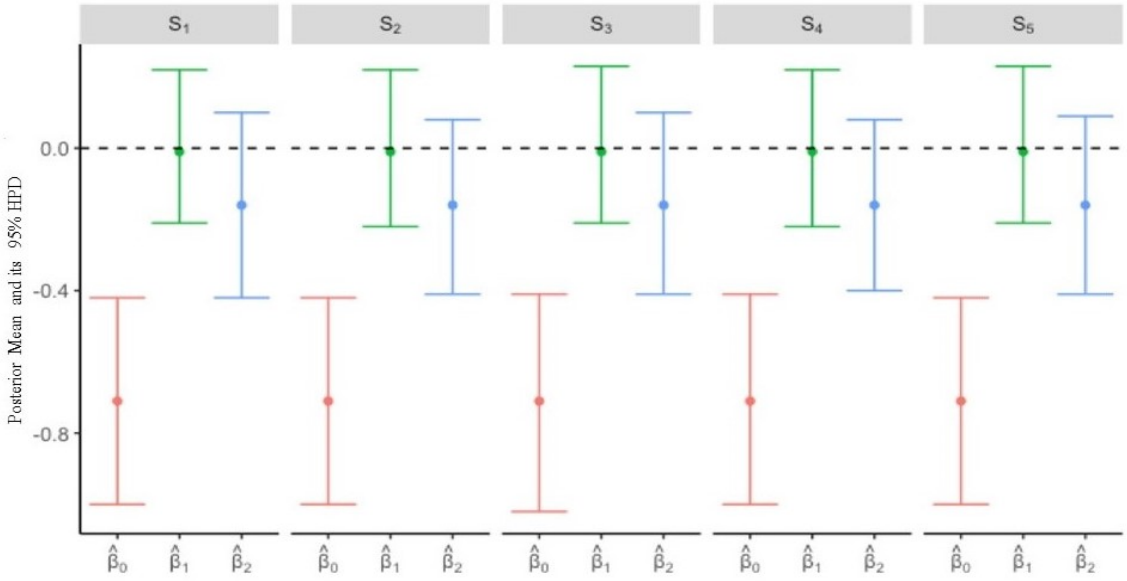

Figure 2.3: Comparison of posterior means of regression parameters with changing $\boldsymbol{\beta}$ prior parameters

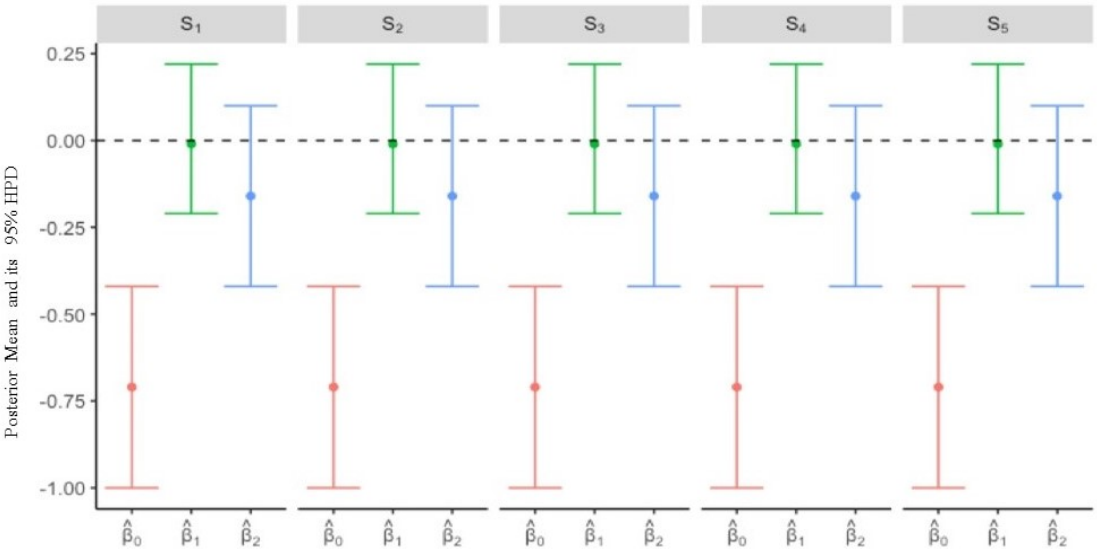

Figure 2.4: Comparison of posterior means of regression parameters with changing ϕ prior parameters

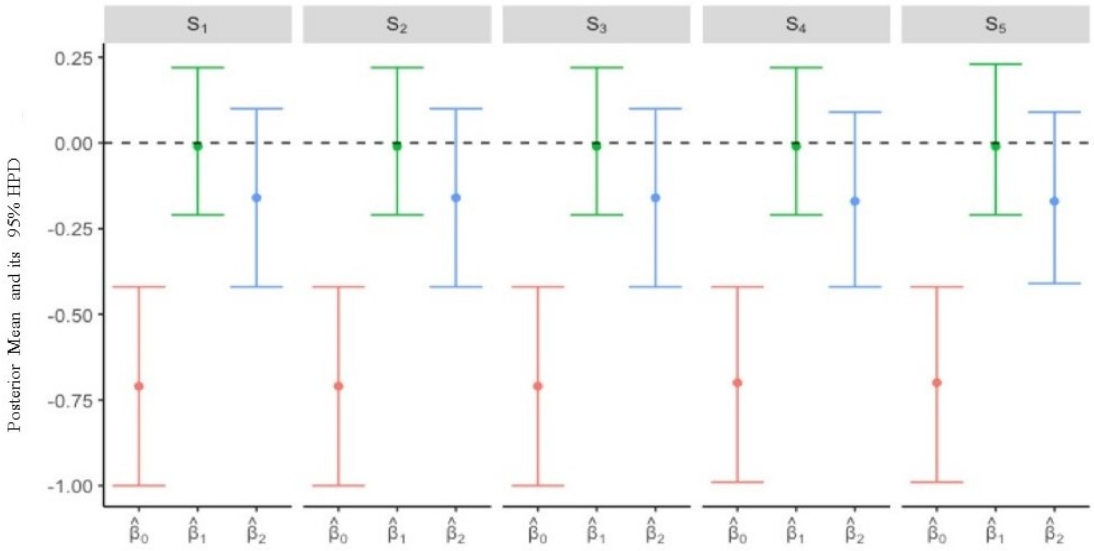

Figure 2.5: Comparison of posterior means of regression parameters with changing σ_ζ^2 prior parameters

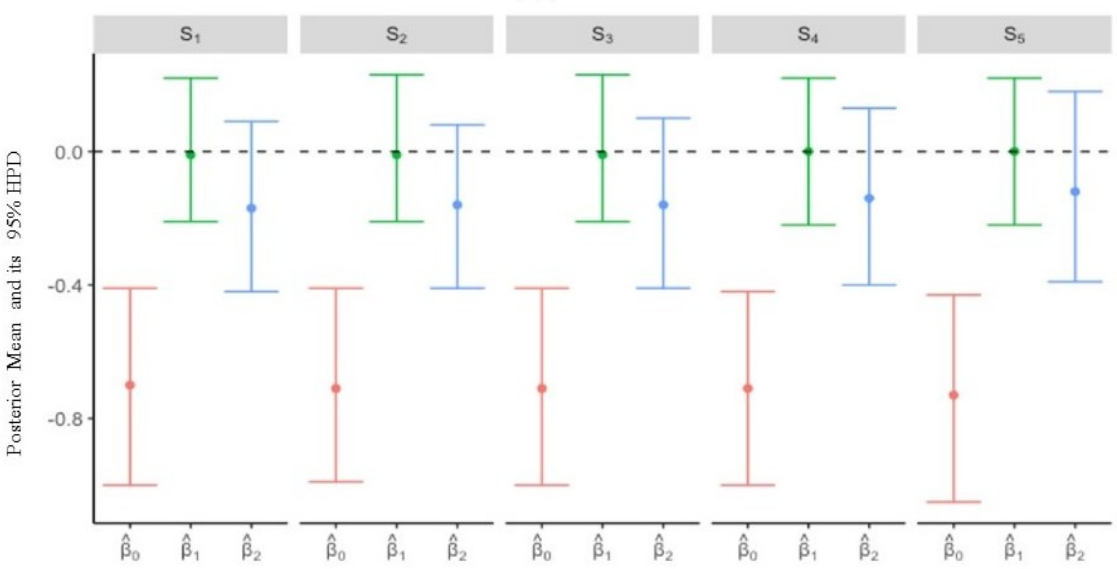

Figure 2.6: Comparison of posterior means of regression parameters with changing $\boldsymbol{L}$ prior parameters

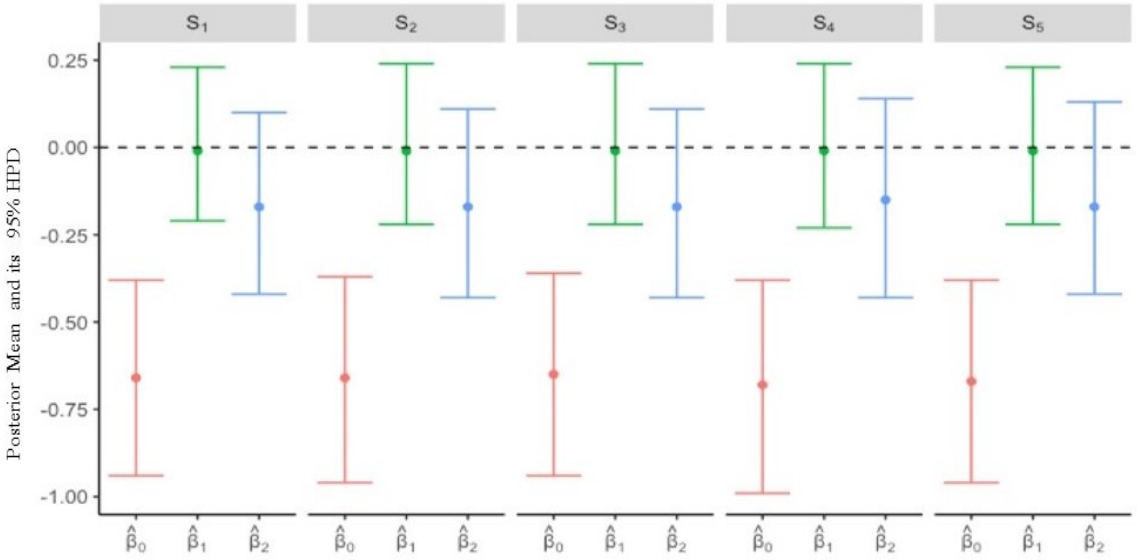

Figure 2.7: Comparison of posterior means of regression parameters with changing prior distributions

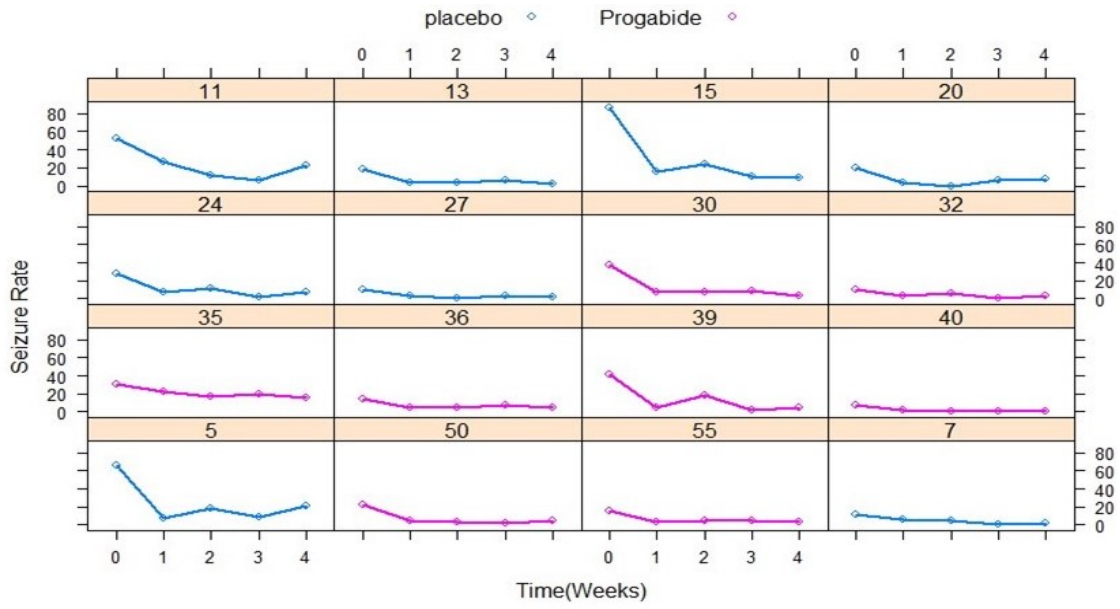

Figure 2.8: Profiles of seizure rates for selected subjects

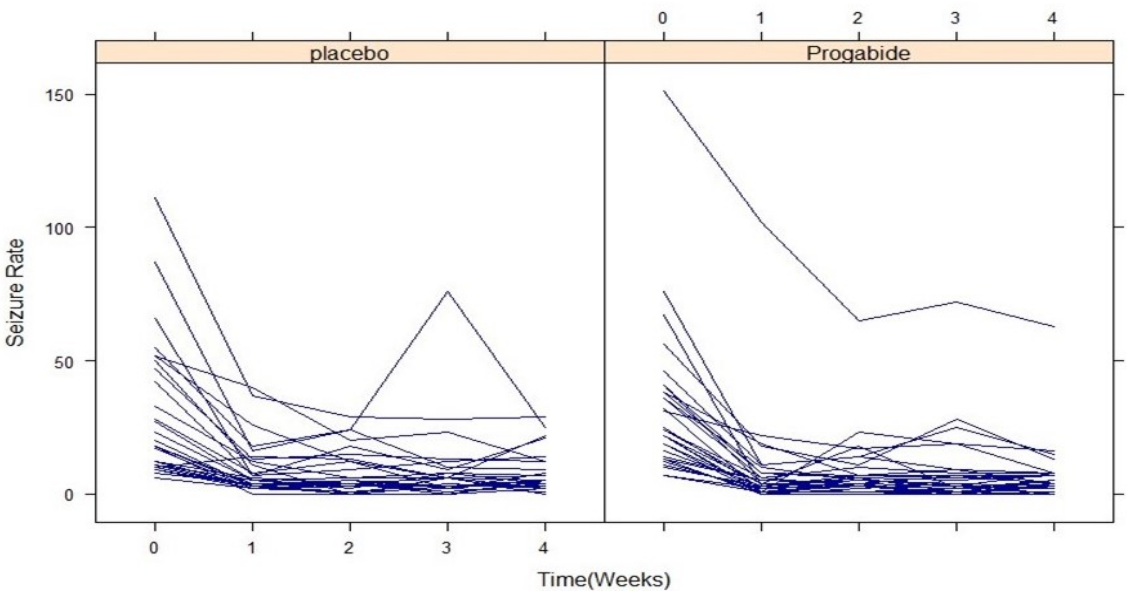

Figure 2.9: Treatment arm-wise profile of seizure rates

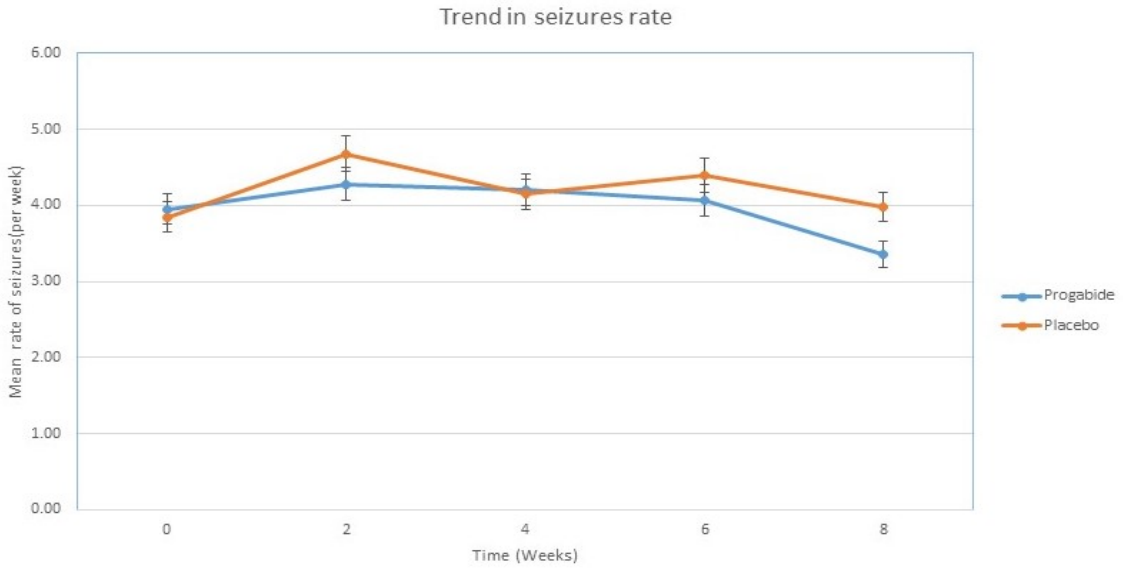

Figure 2.10: Weekly average seizure rates

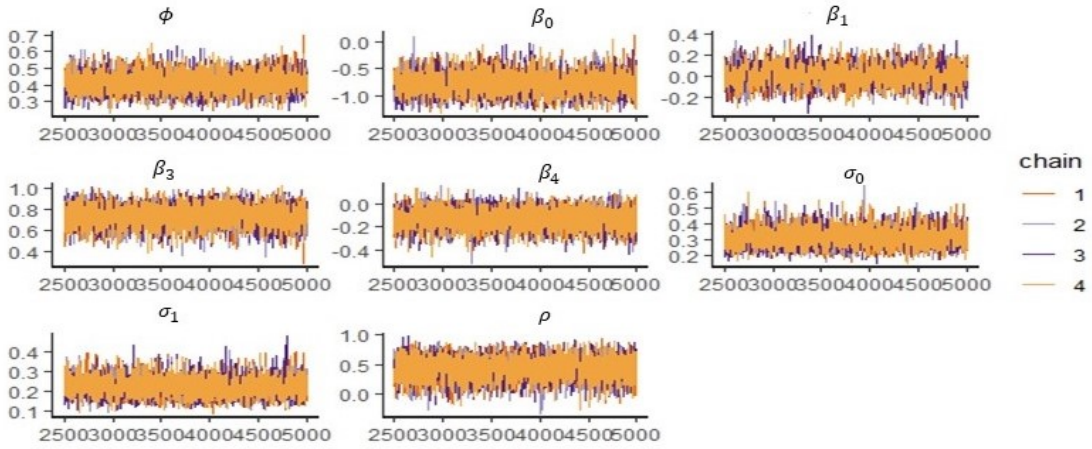

Figure 2.11: Trace plots for $\mathcal{M}_3$ (CMP model)

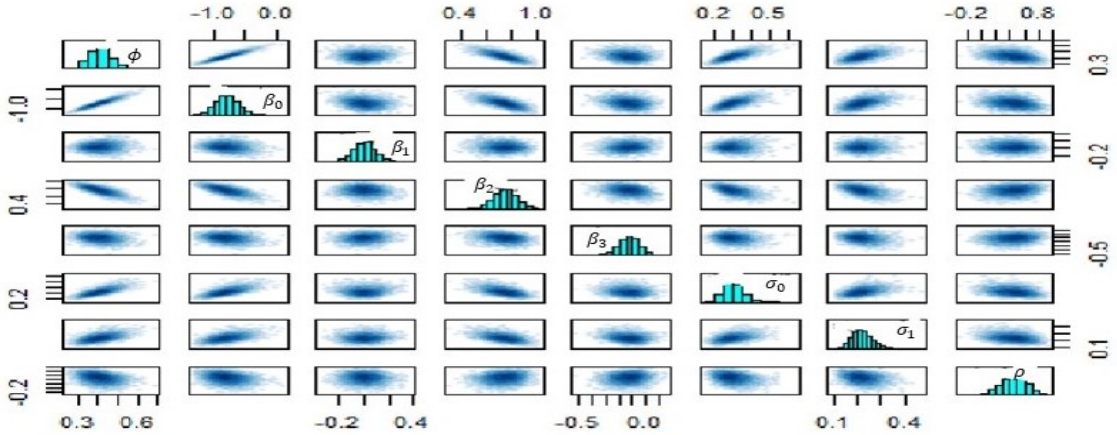

Figure 2.12: Pair plots for $\mathcal{M}_3$ (CMP model)

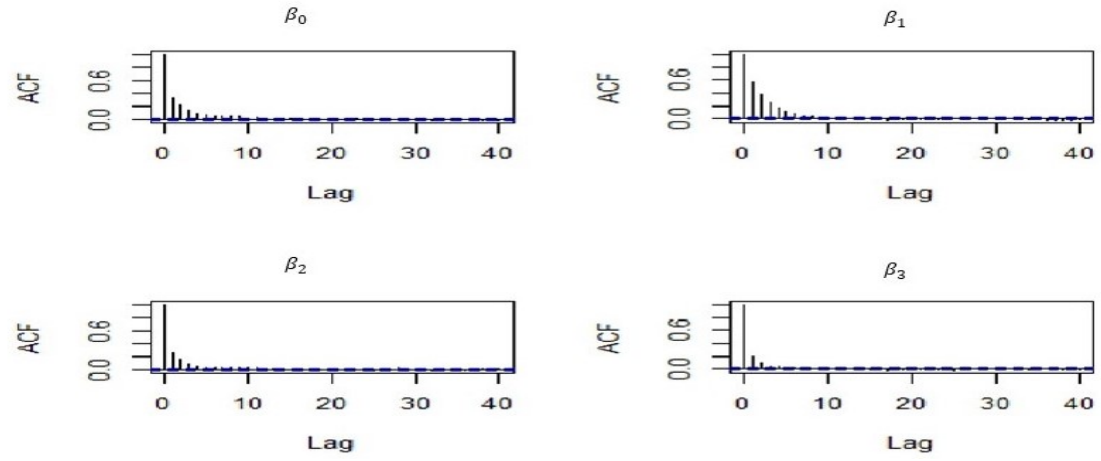

Figure 2.13: ACF plots for$\mathcal{M}_3$ (CMP model)

Chapter 3

Bayesian Generalized Additive Model for Longitudinal Count Data Distributed as Conway-Maxwell Poisson

3.1 Introduction

The parametric model, such as linear mixed effect model assumes that the shape of the functional relationship between mean of the longitudinal response and covariates is known and linear. The parametric model explains the relationship with a relatively smaller number of regression coefficients, maintains model parsimony, and keeps parameter interpretation simple. However, in clinical trials and observational studies, there are situations where a non-linear relationship exists between response and prognostic factors (T. J. Hastie & Tibshirani, 1990). In such a situation, using fewer parameters may be too restrictive in capturing the non-linear functional relationship. Longitudinal model has to allow grater flexibility to account for the

relationship via non/semi-parametric modeling approach. In non-parametric regression model, the shape of the functional relationship is not settled down in advance, instead, it is largely determined by the data itself (G. M. Fitzmaurice et al., 2012).

In practice, we encounter count data analysis arising from a variety of studies (e.g, cross-sectional, longitudinal or clustered). Likewise, linear mixed effect model, the generalized linear parametric model may not be flexible to capture non-linearity between the link function and covariates while modeling count data. To address the situation, we can add more flexibility in the model by replacing the linear predictor with splines. The resulting model is termed as generalized additive model (GAM). According to T. J. Hastie & Tibshirani (1990), the GAM prevents model misspecification, hence provides reasonable inference for the parameters of interest. The GAM was studied in the literature by using count data with Poisson and Negative Binomial distributional assumptions including their zero-inflated variants in Harezlak et al. (2018). The available packages dealing with GAM do not support Conway-Maxwell Poisson (CMP) distributional assumption of the count data. To the best of our knowledge, no literature is available regarding GAM with longitudinal data distributed as CMP both in frequentist and Bayesian settings. In order to capture non-linearity between link function and covariates, to account for subject-specific heterogeneity, and to avoid integrational complexity (discussed in Chapter 2) we propose a Bayesian generalized additive mixed model (BGAMM) for longitudinal count data distributed as CMP.

This chapter is organized as follows. Section 2 includes the proposed statistical model following Bayesian inference in Section 3. An illustration of the proposed model with an application to a hypothetical data is presented in Section 4, and a short discussion is noted in Section 5, and full conditionals in Appendix B.

3.2 The Proposed Model

3.2.1 General form of B-Spline model

A spline of degree F is a function constructed by connecting polynomial segments of degree F so that the function is continuous, has $(F-1)$ continuous derivatives, and Fth derivative between knots is constant. The linear mixed effect model of penalized splines for longitudinal response demonstrates that the mean response is a function of fixed effects, and two sets of random effects. The first set of random effects ζ_i allows each individual to have her/his own piece-wise linear curve that is offset from the smooth population averaged curve by ζ_i for $i = 1, \ldots, I$, and the additional random effects, γ_c for $c = 1, \ldots, C$, are the coefficients for the truncated line functions, $(t_{ij} - S_c)_+$ for $j = 1, \ldots, n_i$, that produce a smooth regression function, $\Psi(t_{ij})$. The amount of smoothing depends on the relative value of variance of γ (σ^2_γ). The γ_c takes care of the non-linear trend in the mean response, and ζ_i, varying across subjects, accounts for correlation among the repeated measures. To fix this idea, we consider the following expression

$$\mathrm{E}(y_{ij}|\boldsymbol{X}) = \boldsymbol{X}_i\boldsymbol{\beta} + \Psi(t_{ij}) + \zeta_i, \tag{3.1}$$

where, $\boldsymbol{X}_i\boldsymbol{\beta}$ is the parametric part, a linear function of covariate $\boldsymbol{X}$, and $\Psi(t_{ij}) = \sum_{c=1}^{C} \gamma_c (t_{ij} - S_c)_+$ is the nonparametric part, $(t_{ij} - S_c)_+ = (t_{ij} - S_c)$ if $(t_{ij} - S_c) > 0$, and equal to zero otherwise. The S_c are the knot locations in the piece-wise linear function of time t_{ij}, y_{ij} denotes the jth response on the ith individual at time t_{ij}, $\zeta_i \sim N(0, \sigma^2_\zeta)$, and $\gamma_c \sim N(0, \sigma^2_\gamma)$.

3.2.2 Statistical Model

Let $\boldsymbol{y}_i = (y_{i1}, \ldots, y_{in_i})^T$ be the independent count response vector of subject i for $i = 1, \ldots, I$ and $j = 1, \ldots, n_i$, $\boldsymbol{X}_i = (\boldsymbol{x}_{i1}, \ldots, \boldsymbol{x}_{in_i})^T$ be a $(n_i \times (p+1))$ design matrix of fixed effect covariates, where $\boldsymbol{x}_{ij}^T = (1, x_{ij1}, \ldots, x_{ijp})$ is a $(p+1)$ dimensional covariate vector, $\boldsymbol{Z}_i = (\boldsymbol{z}_{i1}, \ldots, \boldsymbol{z}_{in_i})^T$ be a $(n_i \times C), (C \leq (p+1))$ known design matrix, where $\boldsymbol{z}_{ij}^T = \Big((x_{ij1} - S_1)_+, \ldots, (x_{ij1} - S_c)_+, \ldots, (x_{ij1} - S_C)_+\Big)$ is a C-dimensional basis vector, $\boldsymbol{\zeta} = (\zeta_1, \ldots, \zeta_I)^T$ is the subject specific random intercept vector, $\tilde{\boldsymbol{\zeta}}_i = (\zeta_i, \ldots, \zeta_i)^T$, a n_i dimensional vector, and $\boldsymbol{\beta} = (\beta_0, \beta_1, \ldots, \beta p)^T$ is a $(p+1)$ dimensional fixed effect co-efficient vector, $\boldsymbol{\gamma} = (\gamma_1, \gamma_2, \ldots, \gamma_C)^T$ is a C dimensional spline coefficient vector. Then, a generalized additive mixed effect model (GAMM) is given by

$$\mathrm{E}(\boldsymbol{y}_i | \boldsymbol{\beta}, \boldsymbol{\gamma}, \tilde{\boldsymbol{\zeta}}_i) = g^{-1}\Big(\boldsymbol{X}_i\boldsymbol{\beta} + \boldsymbol{Z}_i\boldsymbol{\gamma} + \tilde{\boldsymbol{\zeta}}_i\Big), \tag{3.2}$$

where $g^{-1}(\cdot)$ is link function.

Further, let θ_{ij} be the shape parameter of CMP distribution associated with jth component of $\boldsymbol{y}_i$. Then, the GAMM for longitudinal count response (distributed as CMP) is given by

$$\log(\theta_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\gamma} + \zeta_i, \tag{3.3}$$

where $\zeta \sim N(0, \sigma_\zeta^2)$, and $\gamma_c \sim N(0, \sigma_\gamma^2)$.

3.2.3 The Likelihood Function

By considering the random effects as latent variables, the complete data likelihood can be written in the following form

$$\mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \phi, \zeta_i) = \prod_{i=1}^{I} \left[\prod_{j=1}^{n_i} f(y_{ij} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \phi, \zeta_i) f(\zeta_i \mid \sigma_\zeta^2) \right]. \tag{3.4}$$

When response y_{ij} distributed as CMP we can write equation (3.4) as

$$\begin{aligned} \mathcal{L}(\boldsymbol{\beta}, \boldsymbol{\gamma}, \phi, \sigma_\zeta^2) = & \prod_{i=1}^{I} \prod_{j=1}^{n_i} \left(\frac{1}{y_{ij}!} \right)^{\phi} \times \left(\exp\left(\boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{z}_{ij}^T \boldsymbol{\gamma} + \zeta_i \right) \right)^{y_{ij}} \\ & \times \left(\frac{\sum_{k=0}^{\infty} \left(\exp\left(\boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{z}_{ij}^T \boldsymbol{\gamma} + \zeta_i \right) \right)^k}{(k!)^{\phi}} \right)^{-1} \\ & \times \prod_{i=1}^{I} (2\pi\sigma_\zeta^2)^{-(1/2)} \exp\Big(- \frac{\zeta_i^2}{2\sigma_\zeta^2} \Big). \end{aligned} \tag{3.5}$$

3.3 Bayesian Inference

The Bayesian modeling needs specification of prior distributions for the parameters under consideration, and generation of the corresponding posterior distribution. Then, obtaining MCMC samples from the posterior distribution by using suitable samplers, and generating posterior summary (means) to avoid complex or intractable integration. Bayesian analysis enables us to explore different characteristics of the parameters. Inclusion of zeros in the 95% HPD intervals of the regression parameters reveal the non-significance of the respective traits associated with the parameter.

3.3.1 Priors and Posteriors

In order to fit a BGAMM, we assume that $\boldsymbol{\beta}$, $\boldsymbol{\gamma}$, σ_ζ^2, and ϕ are independent apriori. Then

$$\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \phi, \sigma_\zeta^2) = \pi(\boldsymbol{\beta}) \times \pi(\boldsymbol{\gamma}) \times \pi(\phi) \times \pi(\sigma_\zeta^2). \tag{3.6}$$

We further assume that, $\boldsymbol{\beta} \sim N_p(\boldsymbol{\beta}_0, \boldsymbol{\Sigma}_0), \boldsymbol{\gamma} \sim N_q(\boldsymbol{\gamma}_0, \boldsymbol{\Sigma}_\gamma), \phi \sim LN(\mu_\phi, \psi)$,and $\sigma_\zeta^2 \sim IG(\alpha, \delta)$. Then the posterior distribution under the proposed model is given by

$$
\begin{aligned}
\pi(\boldsymbol{\beta}, \boldsymbol{\gamma}, \boldsymbol{\zeta}, \phi | \boldsymbol{y}) &= \left[\prod_{i=1}^{I} \prod_{j=1}^{n_i} f(y_{ij} | \boldsymbol{\beta}, \boldsymbol{\gamma}, \phi, \zeta_i) \times \pi(\boldsymbol{\beta}) \times \pi(\boldsymbol{\gamma}) \right] \times \left[\prod_{1=1}^{I} f(\zeta_i | \sigma_\zeta^2) \times \pi(\sigma_\zeta^2) \times \pi(\phi) \right] \\
&\propto \prod_{i=1}^{I} \prod_{j=1}^{n_i} \left(\frac{1}{y_{ij}!} \right)^{\phi} \times \left(\exp\left(\boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{z}_{ij}^T \boldsymbol{\gamma} + \zeta_i \right) \right)^{y_{ij}} \\
&\times \left(\frac{\sum_{k=0}^{\infty} \left(\exp\left(\boldsymbol{x}_{ij}^T \boldsymbol{\beta} + \boldsymbol{z}_{ij}^T \boldsymbol{\gamma} + \zeta_i \right) \right)^k}{(k!)^\phi} \right)^{-1} \times \frac{\exp\left(-\frac{1}{2} \left[(\boldsymbol{\beta} - \boldsymbol{\beta_0})^T \boldsymbol{\Sigma}_0^{-1} (\boldsymbol{\beta} - \boldsymbol{\beta_0}) \right] \right)}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} \\
&\times \frac{\exp\left(-\frac{1}{2} \left[(\boldsymbol{\gamma} - \boldsymbol{\gamma}_0)^T \boldsymbol{\Sigma}_\gamma^{-1} (\boldsymbol{\gamma} - \boldsymbol{\gamma}_0) \right] \right)}{|\boldsymbol{\Sigma}_\gamma|^{\frac{1}{2}}} \times \prod_{i=1}^{I} (\sigma_\zeta^2)^{-(1/2)} \exp\left(-\frac{\zeta_i^2}{2\sigma_\zeta^2} \right) \\
&\times \frac{\delta^\alpha}{\Gamma(\alpha)} \sigma_\zeta^{-2(\alpha+1)} \exp\left(-\frac{\delta}{\alpha} \right) \times \frac{\exp\left(-\frac{1}{2} \left(\frac{\log_{\phi - \mu_\phi}}{\psi} \right)^2 \right)}{\psi \phi}.
\end{aligned}
\tag{3.7}
$$

3.4 Illustration of GAM

We illustrate the proposed model with a hypothetical data example. Data consist of count measurements from 50 subjects in five occasions (at different ages of the subjects). The continuous age variable is generated from normal distribution with a standards deviation 2 and a randomly selected mean from uniform distribution having a support (15, 50). The age variable is considered as varying time in the mixed effect GAM (no fixed time points of measurements). The range of age in analysis data is $(51.10 - 13.12)$ years. Treatment variable is generated from Bernoulli distribution with probability, $p_r = 0.55$. We generate under-dispersed count data as $y_{ij} \sim CMP(50 \times 5, \theta_{ij}, 1.8)$, where the positive shape parameter θ_{ij} is generated by taking an absolute value of a term simulated from normal distribution with standard deviation 1 and a randomly selected mean from uniform distribution with a support

(1, 2). The mean and variance of the count response are 0.812 and 0.578 respectively. The Figure 3.1 illustrates the shape of count response distribution. The relationship between age vs count response, and age vs log(count response) along with their respective trends and 95% confidence intervals are presented in Figures 3.2, and 3.3 (a) and (b) respectively. In both cases we notice that the relationships are non-linear even after taking logarithmic transformation of the counts. This situation leads us to the use of generalized additive model.

3.4.1 Specific model

In a two group setting (study drug vs control, exposed vs non-exposed) the time trend can be incorporated in semi-parametric fashion by allowing the mean response change in a highly non-linear and non-predetermined way. This set up could be very inviting in clinical trials where a pre-determined analysis plan is required but the actual form of mean time trend is not known in advance. The group effect is added in a parametric way that allows a relatively simple and powerful test of it on the mean change over time. Therefore, we include treatment (trt) as a covariate in the model to examine the treatment effect on the function of mean count change. We fit Poisson, Negative binomial (NB), and CMP model with B-spline and perform model assessments by using DIC, LPML, LOO, and WAIC.

We specify a mixed effect GAM with subject-specific random intercept as

$$\log(\theta_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\gamma}_i + \zeta_i, \tag{3.8}$$

where $\boldsymbol{x}_{ij}^T = (1, trt_{ij}, age_{ij}, trt_{ij} \times age_{ij}), \zeta_{qi} = (\zeta_{0i})$, $\boldsymbol{\beta} = (\beta_0, \beta_1, \beta_2, \beta_3)^T$, and $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_C)^T$, and the basis vector for B-spline is $\boldsymbol{z}_{ij}^T = [(age_{ij} - S_1)_+, \ldots, (age_{ij} -$

$S_C)_+]$.

More specifically,

$$\begin{aligned}\log(\theta_{ij}) =&\beta_0 + \beta_1 \times trt_{ij} + \beta_2 \times age_{ij} + \beta_3 \times trt_{ij} \times age_{ij} \\ &+ \gamma_1 \times (age_{ij} - S_1)_+ + , \ldots , + \gamma_C \times (age_{ij} - S_C)_+ + \zeta_i,\end{aligned} \tag{3.9}$$

In fitting a spline regression, knots are usually specified in advance. The exact locations and number of knots (C) usually not too sensitive to the predictive value of the regression (Harrell Jr, 2013). A good number of approaches have been discussed in the literature to select the number of knots, and places in the generalized additive model (GAM) setting. Some of them are heuristic (Harrell Jr, 2013), and some of them are statistical method based (Wood & Wood, 2015). Harrell Jr (2013) suggests if the sample size, $n \geq 100$ then $C = 5$, and if $n \leq 30$ then $C = 3$. The function $gam()$ in R package mgcv (Wood & Wood, 2015) uses generalized cross-validation (GCV) approach that automatically chooses the number of knots for the model. However, *mgcv* package does not support CMP distribution. Akaike information criterion (AIC) can also be used to choose the number of knots in a GAM fitting (Van Houwelingen & Le Cessie, 1990). This approach chooses C to maximize model likelihood ratio $\chi^2 - 2C$ (Harrell Jr, 2013).

Once, the number of knots is chosen we can use them in equally spaced locations. However, other options can also be applied. Some of the alternative approaches can be mentioned here. In most cases, researchers use 3 to 5 of knots and they are placed at fixed percentiles of the data. Harrell Jr (2015) suggested heuristic percentiles which is popular to the biostatisticians. If the sample size is less than 100, Stone & Koo (1985) suggested replacing outer quantiles with 5th smallest and 5th largest and the inner three quantiles at equally spaced positions of the variable. How-

ever, in our analysis we use $C = 30$ arbitrarily to increase the model complexity, and to observe the model implementation time via our code.

The model assumes a single realization of $(\gamma_1, \ldots, \gamma_C)$, and these C random coefficients are shared by all individuals. These random coefficients must be constrained to have the same variance σ_γ^2 to avoid non-convergence in model fitting (G. M. Fitzmaurice et al., 2012). Bayesian analysis takes care the uncertainty in smoothing parameters. Therefore, the assumption $\gamma_c \sim N(0, \sigma_\gamma^2)$ is appropriate if we use O'Sullivan Spline (Harezlak et al., 2018). O'Sullivan penalized splines imitates the natural boundary behavior of smoothing splines (https://arxiv.org/abs/0707.0143). We are using such splines in our analysis. Harezlak et al. (2018) implemented a GAM with normally distributed response in rstan in Bayesian setting. We implement here a similar setting with different distributional assumption of count data.

In our analysis, we consider the following prior distributions

$$\phi \sim LN(0, 15)$$
$$\zeta \sim N(0, \sigma_\zeta^2), \sigma_\zeta^2 \sim IG(0.2, 0.2)$$
$$\beta_k \sim N(0, 10000) \text{ for } k = 0, 1, 2, 3$$
$$\gamma_c \sim N(0, \sigma_\gamma^2) \text{ for } c = 1, \ldots, 30, \text{ and}$$
$$\sigma_\gamma^2 \sim IG(0.2, 0.2)$$

We assume that β_k can take any value with normal mean 0 and standard deviation 10000. The dispersion parameter ϕ is positive, $0 < \phi < 1$ indicates over-dispersion, and $\phi > 1$ indicates under-dispersion. We assign a log normal prior with mean zero and moderate standard deviation for ϕ. Wosoba et al. (2018) use log normal distribution as a prior for ϕ. Since σ_ζ^2 is a scale parameter with a lower bound of zero, we assign a inverse-gamma prior for σ_ζ^2, as suggests in Gelman et al. (2006). We perform Bayesian analysis by using Stan language and rstan R package in four chains with 10000 iterations having 2500 warm-up each. We report LPML, LOO,

WAIC, and DIC for the model assessments among the competing models such as CMP, NB and Poisson. The results are presented in Table 4.1.

Table 3.1: Posterior summary under Poisson, NB, and CMP model

	Poisson				NB				CMP			
Parameter	Mean	Std	95% HPD	$\hat{R}$	Mean	Std	95% HPD	$\hat{R}$	Mean	Std	95% HPD	$\hat{R}$
β_0	−1.11	0.44	(−1.98,−0.24)	1	−1.12	0.46	(−2.01,−0.22)	1	−1.07	0.57	(−2.17,0.05)	1
β_1	0.66	0.56	(−39,1.78)	1	0.65	0.58	(−0.46,1.82)	1	1.00	0.72	(−0.42,2.40)	1
β_2	0.03	0.01	(0.00,0.05)	1	0.02	0.01	(0.00,0.05)	1	0.04	0.02	(0.01,0.07)	1
β_3	−0.02	0.02	(−0.05,0.01)	1	−0.03	0.02	(−0.06,0.01)	1	−0.03	0.02	(−0.08,0.01)	1
σ_ζ	0.15	0.07	(0.04,0.28)	1	0.15	0.07	(0.04,0.28)	1	0.25	0.13	(0.05,0.49)	1
σ_γ	0.16	0.08	(0.04,0.33)	1	0.17	0.08	(0.04,0.33)	1	0.21	0.12	(0.04,0.43)	1
ϕ	-	-	-	-	0.03	0.02	(0.01,0.07)	1	2.16	0.30	(1.57,2.73)	1
DIC	1011.44				977.62				721.34			
$\overline{\Delta(\boldsymbol{\Theta})}$	768.47				571.77				632.44			
p_D(approx.)	226				209.15				89			
LPML	−27.15				−13.92				0.37			
LOO	2715.09				1392.65				−36.43			
WAIC	2034.88				1493.28				719.64			

Posterior summary of the parameter $\phi = 2.16$ from CMP model reveals that data are under-dispersed as we assumed while simulation. The values for $\hat{R}$ for all parameters in all models illustrate good performances of the model. The DIC values for Poisson, NB, and CMP models are 1011.44, 977.62, and 721.34 respectively. The smallest values of DIC (721.34), LOO (−36.43), WAIC (719.64), and highest value of LPML (0.37) for CMP model, illustrate the best fit of the CMP model in comparison to Poisson and NB model. Likewise in chapter 2, we notice that both Poisson and NB model produce similar parameter estimates as the data are under-dispersed, but different from CMP model outputs. The converted regression coefficients for CMP model are $\beta_0/2.16 = -0.50$, $\beta_1/2.16 = 0.46$, $\beta_2/2.16 = 0.02$, $\beta_3/2.16 = -0.01$ are also a bit different from other models.

We produce trace-plots, density plots, plot for Metropolis acceptance rate, histogram for $\hat{R}$ and ACF plots for all parameters from CMP, and reported them in Figures 3.2 to 3.6 respectively. The caterpillar like shape of the trace plots depict the good mixing of the posterior samples across chains. The bell shapes of density plots for posterior means for all parameters depict non-disruption in MCMC sam-

pling process. In Figure 3.4 we observe that log posterior is bell shaped which is an indication of better convergence of the model. The mean metropolis acceptance rate also high. The average $\hat{R}$ is closer to 1 reveals that posterior distribution is close to the target distribution. Exponential shape of ACF plots in Figure 3.6 stipulates relatively lesser auto-correlated samples generation which is an advantage of using Hamiltonian samplers (No-U-Turn samplers) over other MCMC sampling techniques.

3.5 Discussion

Dealing with generalized linear model (GLM), K. F. Sellers & Shmueli (2010) opines that CMP model is a better alternative to NB and Poisson while count data exhibit dispersion. However, the application of CMP generalized additive mixed model dealing with longitudinal count data was unexplored in literature. In this study we attempt to examine alternatives of GAM by considering usual count data distributions such as CMP, NB, and Poisson by using a hypothetical dataset. The response data here is under-dispersed. From the analysis dataset we notice that CMP model with B-spline fits better than NB and Poisson model on the basis of all Bayesian model assessment criteria we did consider. A crude comparison of model regression parameters across models is possible with a suitable conversion (β_k/ϕ) of CMP model coefficients, we perform this point-wise comparison across models. We experience that both Poisson and NB model result almost same values for respective parameters that conforms similar findings by Sellers et al.(2010) for under-dispersed data. However, the values are different in case of CMP model.

We do not accommodate automated knot selection in GAM in CMP distributional setting. In fact, no package is available for selecting number of knots and implementing GAM when count data distribution is assumed as CMP. The heuristic percentiles for knot location suggestion by Harrell Jr (2015) may be followed. The

other alternatives might be followed by fitting GAM using mgcv R package with the logarithmic link by considering either Poisson or Negative Binomial or both to determine C first and then use that C in CMP setting. We can also check model fitting by calculating DIC for different choices of C and determine the value of C based on the lowest DIC value. However, use of excessive number of knots might minimize the roughness of the non-linear curve with a high chance of producing over-fitting model (G. M. Fitzmaurice et al., 2012). The GAM is extremely useful to study the efficacy of the intervention drug or devices in clinical trials where functional relationship between mean count response and time is found to be non-linear. We were limited due do unavailability of a good real data set to explore a scientific research question of interest in CMP GAM setting which remains for further exploration. In addition, we do not consider zero inflated longitudinal counts, missing data issues, and duel links (modeling log link of dispersion parameter) in CMP GAM fitting which also remain for further extension.

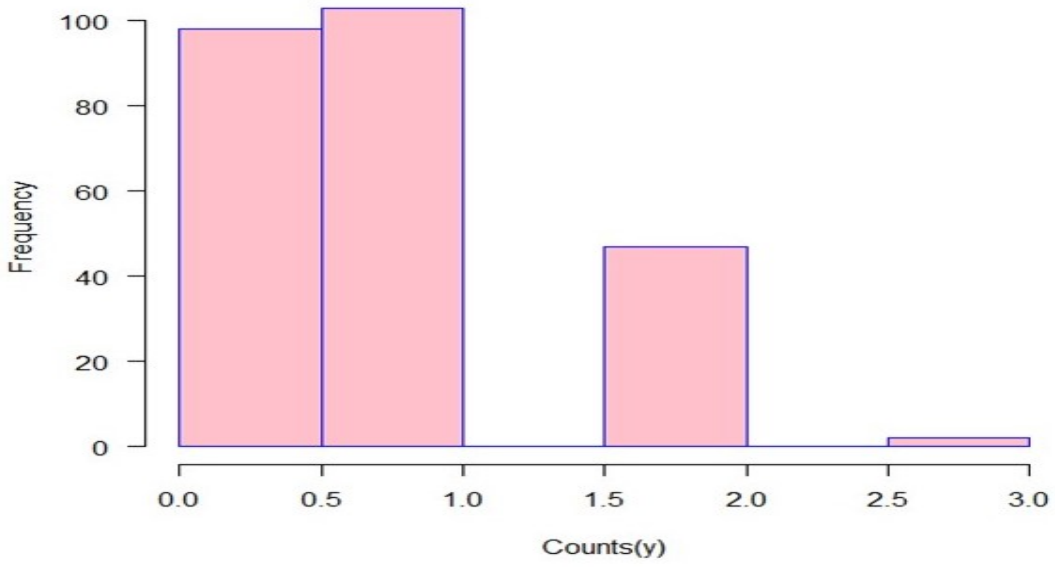

Figure 3.1: Distribution of the data

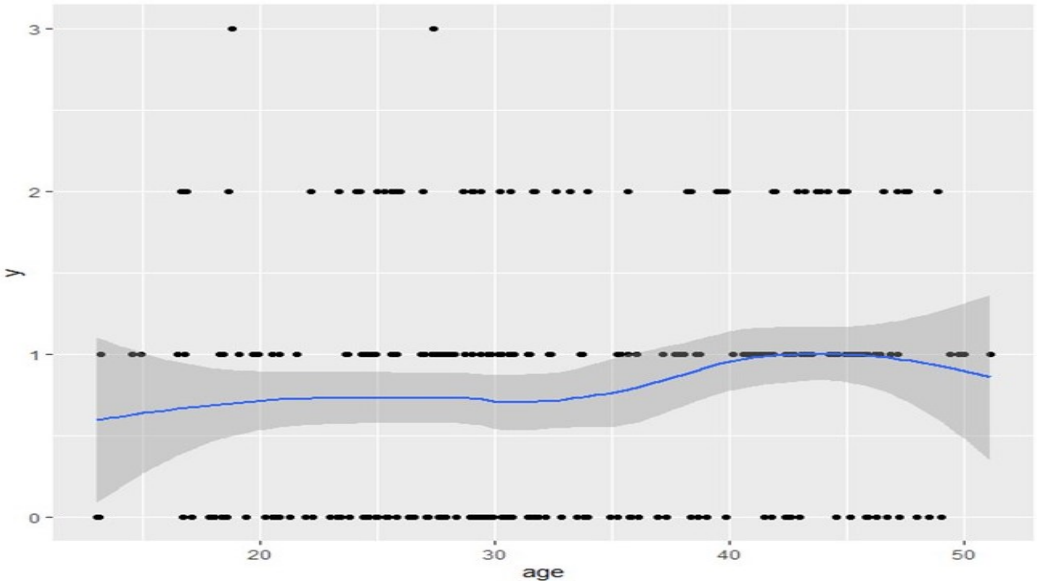

Figure 3.2: Relationship between counts and age

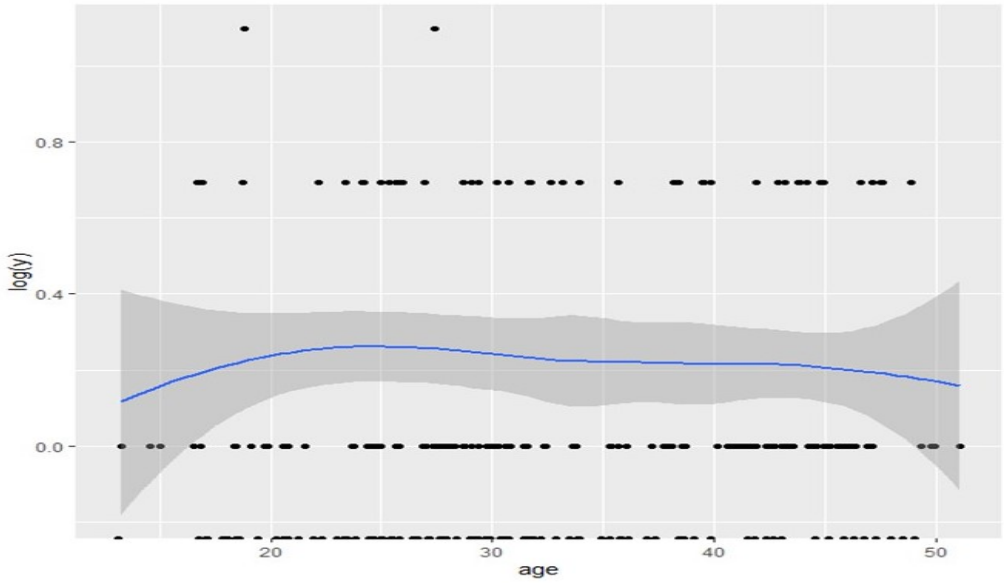

Figure 3.3: Relationship between log(counts) and age

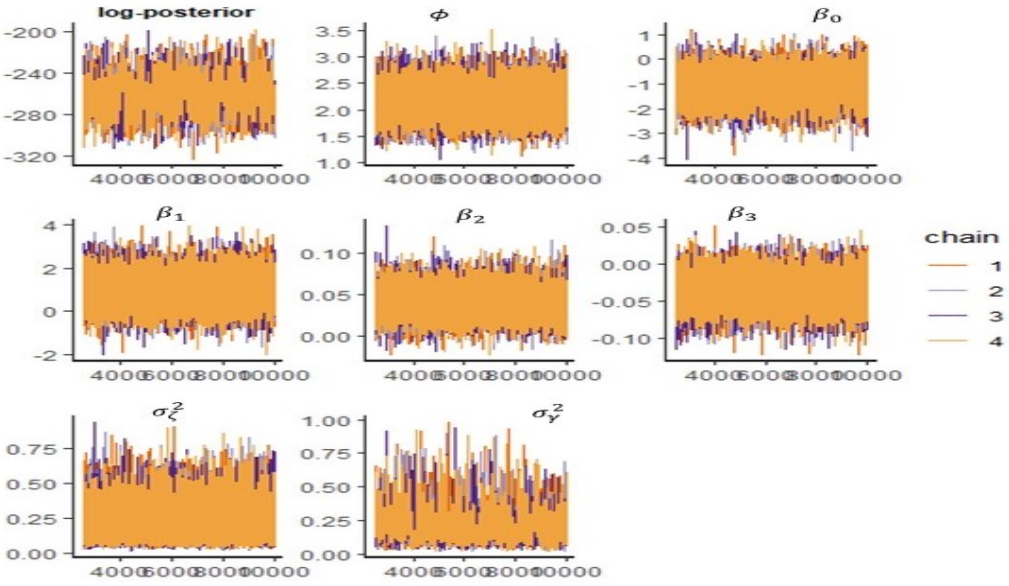

Figure 3.4: Trace plots for CMP model with B-spline

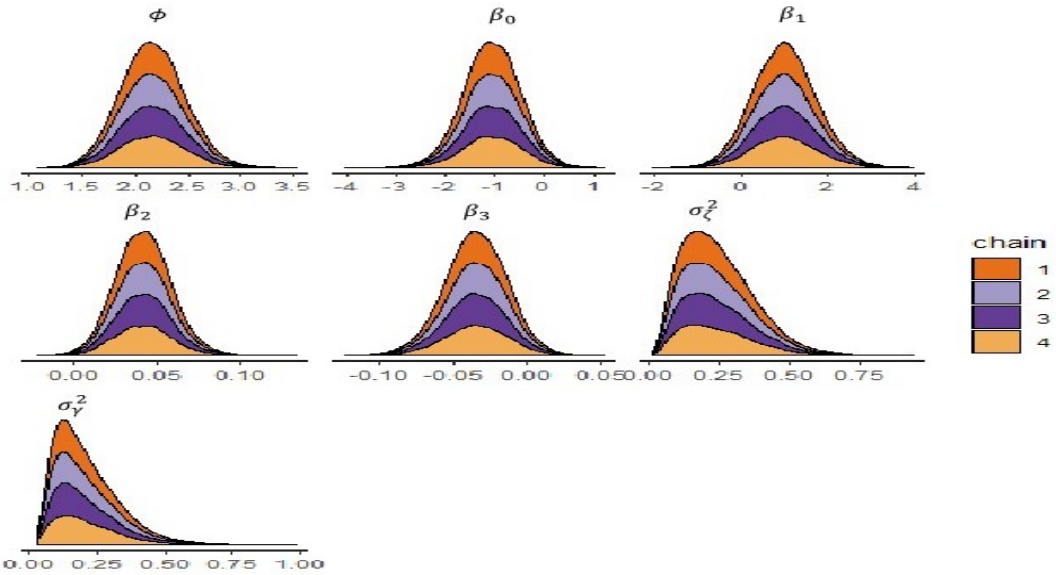

Figure 3.5: Density plots for CMP model with B-spline

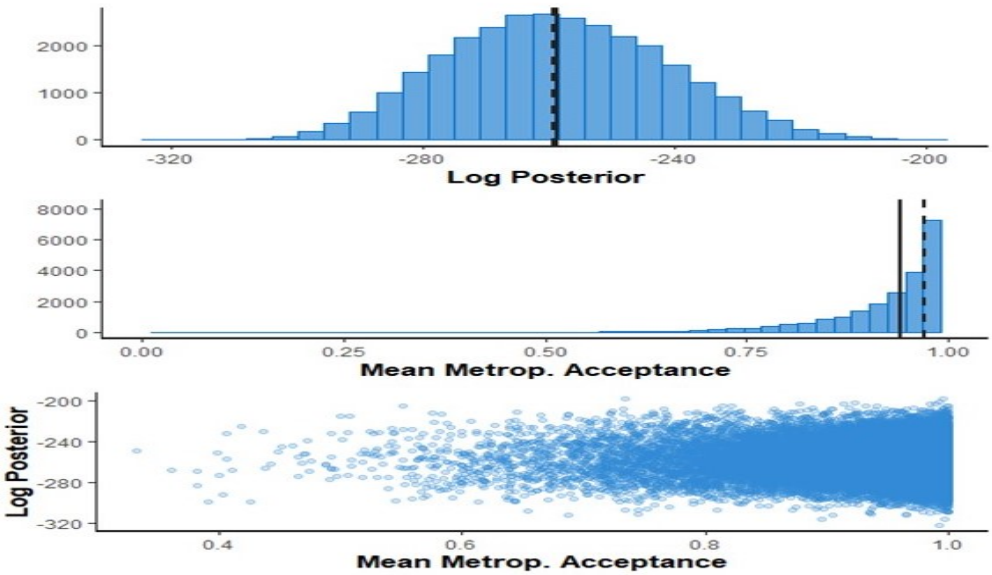

Figure 3.6: Metropolis acceptance rate for CMP model with B-spline

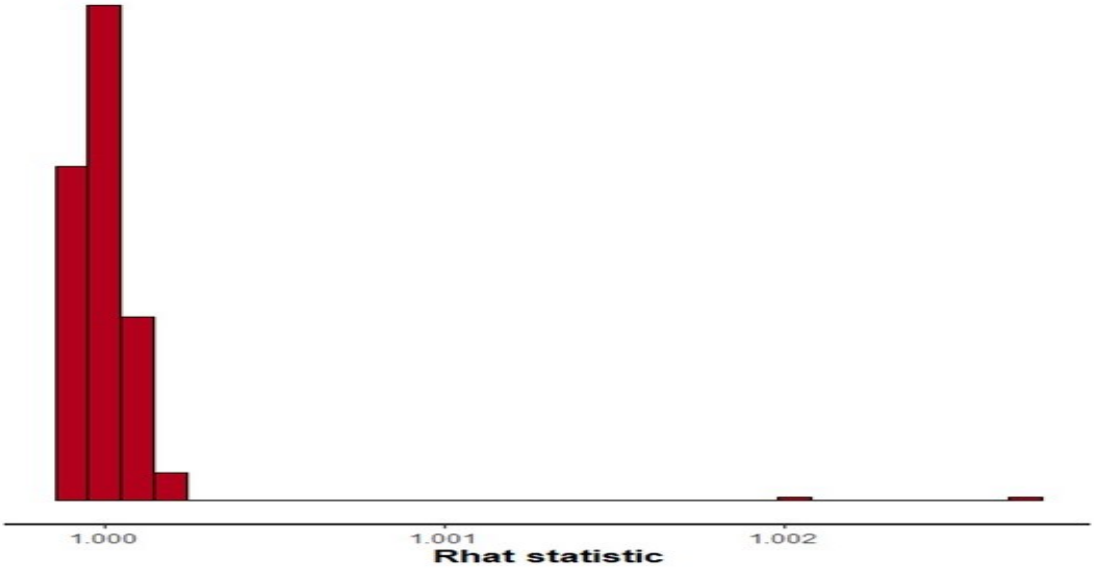

Figure 3.7: Histogram for Rhats for CMP model with B-spline

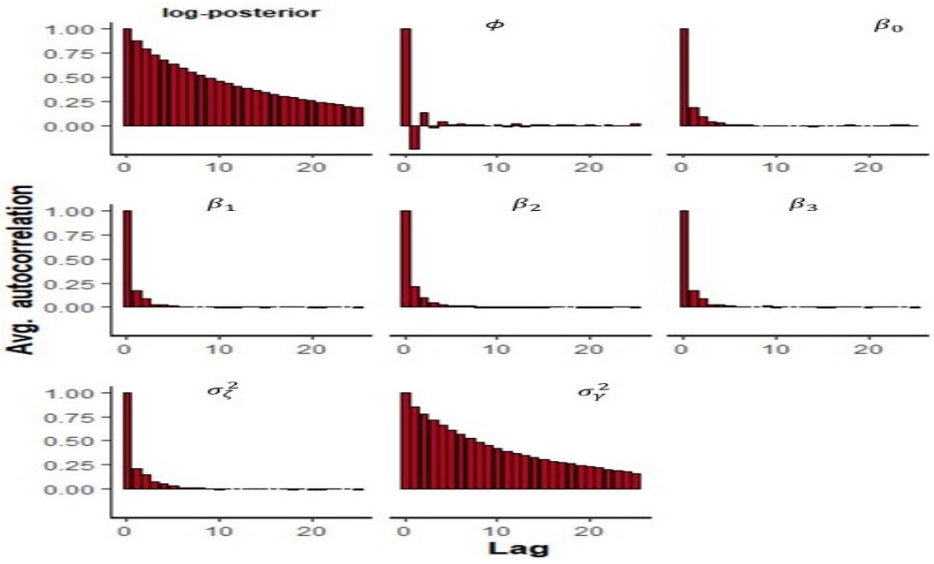

Figure 3.8: ACF plots for CMP model with B-spline

Chapter 4

Association Between Multivitamin Supplementation and Overall well-being in HIV Patients:Application of the Bayesian Generalized Mixed Effect Model

4.1 Introduction

There has been a substantial decrease in the number of new HIV infections during the past decade. However, recent estimates from United States indicate that there were 36,400 new cases in 2018 and an estimated 1.2 million people had HIV (prevalence) (for Disease Control et al., 2020). The advent of highly active combined antiretroviral therapy (cART) has tremendously altered the natural history of HIV infection

transforming the disease from a fatal infectious one to a manageable chronic condition (Deeks et al., 2013). With an enhanced life expectancy, the patterns of comorbidities have changed among the HIV-infected population. Along with the continued management of the HIV infection, this population needs care of age-related comorbidities which may be complicated by issues such as medication associated toxicities (Chu & Selwyn, 2011). Given the situation, there is continued need to exploit all potential interventions to improve the health and health related quality of life of the HIV-infected population, and nutritional interventions are one of them.

Chronic diseases in general, have certain underlying pathologic mechanisms that may be modified by nutrients such as reduction of oxidative damage by antioxidants, DNA methylation regulated by folate and B vitamins, bone metabolism regulated by vitamin D and calcium, and cell differentiation, proliferation, and growth regulated by retinol, calcium, and vitamin D (H.-Y. Huang et al., 2006). Also, it is well established that chronic diseases lead to micronutrient deficiencies that in turn further increase the risk of disease progression and symptomatology (Erickson et al., 2000). Thus, micronutrient supplementation may be beneficial in disease mitigation. Regarding HIV-infected population, vitamin D supplementation has been studied extensively. Recent studies have found that vitamin D supplementation attenuates the effect of immune activation and decreases bone turnover markers in HIV-infected population (Eckard et al., 2018; Nanayakkara et al., 2019; Sudjaritruk & Puthanakit, 2017).

The potential beneficial role of multi-micronutrient supplementation has been evaluated to a relatively lesser extent in the HIV-infected population. Interventional studies conducted on HIV-infected children have shown micronutrient supplementation beneficial in terms of improvement in CD4 counts, delay in the progression of disease, improved appetite, reduced duration of pneumonia or diarrhea and improved wasting (Gautam et al., 2014; Mda et al., 2010, 2013).

However, the interventional studies conducted on adult HIV-infected population exhibit mixed result probably dependent on the characteristics of the study population or statistical analysis used. Majority of the studies conducted on adult participants, evaluated the micronutrient supplementation through improvement in biomarkers and mortality as the outcomes, and not through the disease symptomatology. Where some studies found improved CD4 counts, reduced viral loads, decreased mortality and reduced risk of AIDS defining condition with micronutrient supplementation (Baum et al., 2013; Hemsworth et al., 2012; Zhao et al., 2009; Kaiser et al., 2006; Jiamton et al., 2003), others did not find any beneficial result (Makinde et al., 2017; Motswagole et al., 2013; PrayGod et al., 2011; Semba et al., 2007).

Given the contradictory findings of the limited multi-micronutrient interventional research conducted in the adult HIV-infected population, the subject needs further evaluation, specially in terms of an effect on the overall symptomatology of the disease. To fill the research gap, our study is using data from a randomized double-blind controlled trial to determine the effect of multivitamin supplementation (containing 1.4 mg B1, 1.4 mg B2, 1.9 mg B6, 2.6 mcg B12, 18 mg niacin, 70 mg C, 10 mg E, and 0.4 mg folic acid) in adult HIV-infected participants on overall disease symptomatology, where overall well-being of the patients accounts for the reduction in counts of sign and symptoms due to study drug.

A variety of regression models based on the Poisson distribution namely standard Poisson, negative binomial, restricted generalized Poisson regression model have been used to model such kind of count data (Winkelmann & Zimmermann, 1995). However, these models are based on certain assumptions. For example, Poisson model assumes equality of mean of variance. In practice this ideal situation happens rarely, instead, variance could be higher (over-dispersion) or lower (under-dispersion) than mean. Failure to address these properties may lead to the inefficient estimation of the model parameters. The Conway-Maxwell Poisson (CMP) as a count data distri-

bution is able to accommodate a wide range of dispersion in the data and has been proven superior to the alternative models (K. F. Sellers & Shmueli, 2010). To our knowledge no previous interventional study has used the Bayesian generalized linear mixed model (BGLMM) which enables us to deal with over/under dispersed longitudinal counts of sign and symptoms, especially when the count data follow CMP distribution. The first objective is to demonstrate the applicability of this model as an alternative to study disease symptomatology, and the second objective is to find the association between multivitamin supplements and overall well-being of the HIV patients. The second section of this study includes methods along with data source, response variable, predictor variables, model and statistical analysis. The results are included in Section 3 followed by discussion and conclusion in Section 4. We demonstrate CMP model implementation coding in the final Section.

4.2 Methods

4.2.1 Data Description

The longitudinal data for this study are collected from a randomized double-blind placebo controlled clinical trial conducted to examine the beneficial effect of multivitamin among HIV-infected adults receiving highly active antiretroviral therapy (HAART) in Uganda reported by Guwatudde et al. (2012). In this trial 400 adults are randomly assigned to either a multivitamin (MV) supplement (including vitamin B-complex, C, and E) or placebo arm with equal proportion who were continuing to receive standard medical care according to Uganda's Ministry of Health guidelines. The current study utilizes the data from 354 subjects for whom complete measurements at 3, 6, 12, and 18 months visits were available. In every visit, the subjects were asked whether they experience any sign and symptoms during last three months in forty five directions such as fatigue, general body weakness, fever, oral thrush etc.

By combining individual answers we create a response variable, namely, the number of symptoms and sign as an indication of overall well-being of the patients. Decrease in the number of sign and symptoms may be viewed as improved health condition of the subjects. The trial collected background and demographic information of the subjects, among them baseline age and gender information were available. Whether the subject receives multivitamin or placebo, baseline age , and gender of the subjects are adjusted in the model as the the covariates.

4.2.2 Statistical Analysis

Let θ_{ij} be the shape parameter of CMP distribution associated with jth component of longitudinal count response vector $\boldsymbol{y}_i$ for subject i. Then, under the notations defined in Chapter 2, the specific model to study number of sign and symptoms is given by

$$\log(\theta_{ij}) = (\beta_0 + \zeta_{0i}) + \beta_1 Trt_{ij} + (\beta_2 + \zeta_{1i}) Time_{ij} + \beta_3 Age + \beta_4 Gender, \qquad (4.1)$$

where ζ_{0i} and ζ_{1i} are the subject specific random intercept and slope respectively, and we assume they are correlated. The βs are the regression coefficients.

We perform Bayesian analysis in Stan language and rstan in four chains with 5000 iterations having 2500 warm-up each. By default Stan uses No-U-Turn Sampling (NUTS), an adaptive version Hamiltonian MCMC. In our analysis we consider the following prior distributions: $\beta_k \sim N(0, 10000)$ for $k = 0, 1, 2, 3, 4$, a diffuse normal prior. The dispersion parameter $\phi \sim LN(0, 15)$, providing positive support for ϕ. We assign $\zeta_q \sim N(0, \sigma^2_{\zeta_q}), \sigma^2_{\zeta_q} \sim IG(0.1, 0.1)$ for $q = 0, 1$, as suggested in Gelman et al. (2006). For the correlated random effects we decompose correlation matrix $\boldsymbol{\Omega}$ with Choleskey decomposition $\boldsymbol{\Omega} = \boldsymbol{L}^T \boldsymbol{L}$, where $\boldsymbol{L}$ is a lower triangular matrix,

and its distribution is assumed as $LKJ(\eta)$. We assume that the random effects are weekly correlated, the value $\eta = 2$, indicates that the correlation is close to zero. We fit Poisson, negative binomial (NB) and CMP model, and compared fitness of the models by using DIC, WAIC, LOO, and LPML.

A subgroup analysis has been conducted by segregating the data by sex with CMP model. In subgroup analysis intervention (MV or Placebo) and age are considered as the covariates. The regression parameter estimates from CMP model are not directly comparable with Poisson and NB as CMP model does not model link function of the mean directly. A transformation (β_k/ϕ) is used to compare coefficients across three (Possion, NB, and CMP) models as suggested in K. F. Sellers & Shmueli (2010). Inclusion of zeros in 95% HPD intervals for coefficients are considered as non-significance of the respective covariates. We report incidence rate ratio (IRR) for each of the covariates. An IRR$<$ 1 indicates positive impact of the trait on reducing number of signs and symptoms.

4.3 Results

We analyzed data from 354 HIV infected subjects of ages 18-67 years with a median age 36 years (IQR=11), among them 173 (48.87%) received multivitamin and the rest received placebo. Male participants are 108(30.5%). The response variable ranges from 0 to 31 with a median sign and symptom count 7 (IQR=6).

In Table 4.1, we report posterior means, standard deviation, and 95% credible intervals for the parameters obtained from Poisson, negative binomial, and CMP model outputs. In Figure 4.1, we illustrate posterior means of the CMP model parameters along with their 95% HPD intervals. The value of dispersion parameter ($\phi = 0.84$) in CMP model indicates that data are over-dispersed. The CMP model retains the lowest DIC (6918.71), LOO (693.48), WAIC (7048.49), and the highest

Table 4.1: Posterior summary under Poisson, Negative binomial and CMP model

Parameter	Poisson			NB			CMP			
	Mean	Std.	95% HPD	Mean	Std.	95% HPD	Mean	Std.	95% HPD	$\hat{R}$
β_0	2.25	0.11	(2.02,2.46)	0.81	0.06	(0.71,0.92)	1.89	0.14	(1.61,2.16)	1
β_1	−0.10	0.05	(−0.20,0.00)	−0.05	0.03	(−0.10,0.00)	−0.09	0.05	(−0.17,0.01)	1
β_2	−0.03	0.00	(−0.03,−0.02)	−0.01	0.00	(−0.02,−0.01)	−0.02	0.00	(−0.03,−0.02)	1
β_3	0.00	0.00	(−0.01,0.01)	0.00	0.00	(0.00,0.00)	0.00	0.00	(−0.01,0.00)	1
β_4	−0.26	0.06	(−0.37,−0.14)	−0.13	0.03	(−0.19,−0.07)	−0.22	0.05	(−0.32,−0.12)	1
σ_0	0.47	0.03	(0.41,0.53)	0.20	0.0.02	(0.17,0.24)	0.38	0.04	(0.31,0.46)	1
σ_1	0.03	0.00	(0.02,0.03)	0.01	0.00	(0.01,0.02)	0.02	0.00	(0.01,0.02)	1
ρ	−0.32	0.10	(−0.50,−0.12)	0.03	0.16	(−0.27,0.36)	−0.24	0.13	(−0.49,0.02)	1
ϕ	-	-	-	0.03	0.01	(0.01,0.04)	0.84	0.05	(0.75,0.93)	1
DIC	7149.73			18650.71			6918.71			
p_D(approx.)	416			52			305			
LPML	−361.89			−1044.88			−346.74			
LOO	723.79			2089.75			693.48			
WAIC	7211.97			19244.44			7048.49			

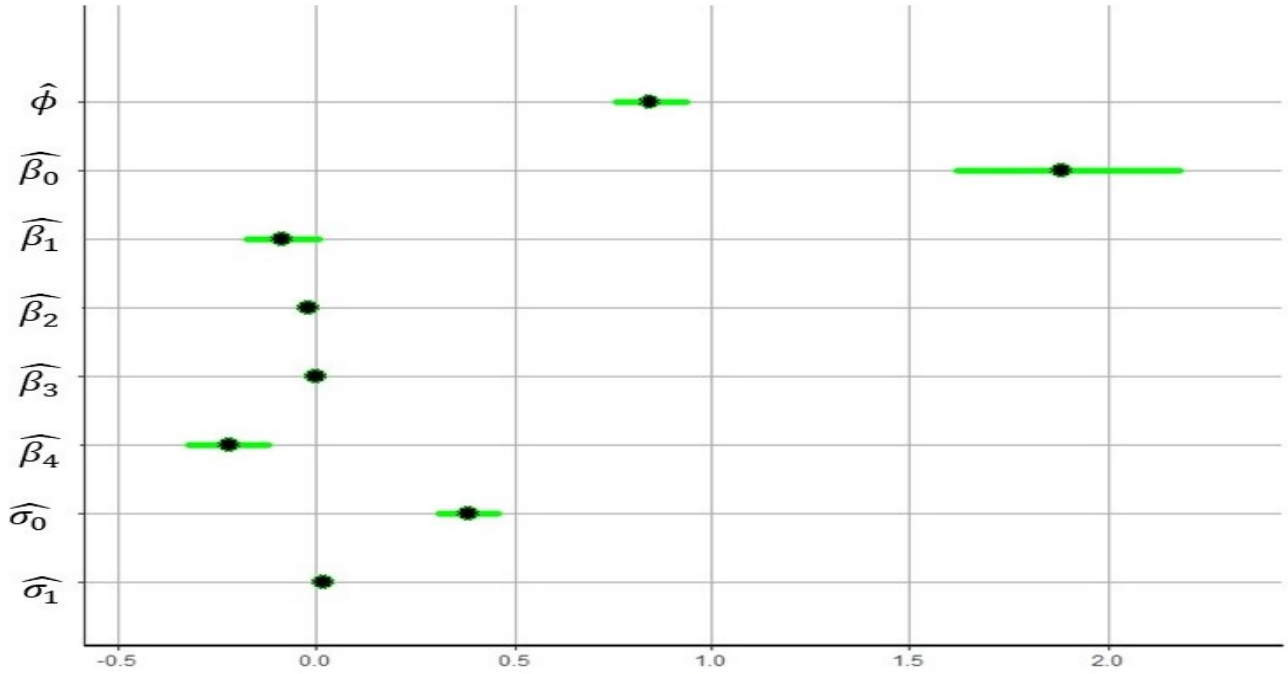

Figure 4.1: Posterior summary of the parameters from CMP model with 95% Credible Interval

(−346.74) among the three models which indicate that the CMP model fits better than Poisson, and negative binomial models. Since, the CMP model is appeared to be the best, we explain results from CMP model. From CMP model, although not significant but $\beta_1 < 0 = 97.20\%$, the incidence rate of suffering from different sign and symptoms of HIV patients who consumes multivitamin is lower [β_1=−0.09, 95%

HPD: (− 0.17, 0.01), and the corresponding IRR= exp(−0.09/0.84)=0.91] than who consumes placebo. The incidence rate of suffering from different sign and symptom of HIV patients who consume multivitamin is lowed by 9% than those who consume placebo while age and gender are adjusted.

For one month increase in the follow up time the incidence rate for sign and symptoms decreases by 2 percent[β_2=−0.02, 95% HPD: (−0.03, −0.02), and the corresponding IRR= exp(−0.02/0.84)=0.98] among the HIV patients, and statistically significant. The incidence rate ratio for signs and symptoms among male is significantly lower than that of women [β_4=−0.22, 95% HPD: (−0.32, −0.12), and corresponding IRR= exp(−0.22/0.84) = 0.77] while adjusting age, time and intervention.

The subgroup analysis illustrates that, male receiving multivitamin experience lesser number of signs and symptoms (IRR=0.91), the experience is also in the similar direction for female (IRR=0.90). As time goes up the male patients experience higher reduction of symptoms (IRR=0.96) than that of the female patients (IRR=0.98). Both in male and female groups, baseline age does not have significant role in reducing sign and symptoms. We present posterior mean along with their corresponding 95% HPD interval in Figure 4.1. It is observed that gender effect is significantly lower than zero and treatment effect is marginally lower than zero, which illustrate that male, and individuals in multivitamin group experience lesser number of sign and symptoms respectively.

4.4 Discussion

Based on all model assessment criteria, our study reveals that the CMP model is an ideal alternative to study, and identify the factors affecting the symptomology status in HIV patients receiving multivitamin as a supplementary intervention. The CMP

model has been proven as a better alternative to dispersed count data (K. F. Sellers & Shmueli, 2010; Morris et al., 2017).

It is evident from all three models that multivitamin, time, and male gender have positive impacts in reducing the number of signs and symptoms. Limited research has been conducted on the effect of multivitamin supplementation in HIV patients. However, the roles of individual vitamins such as vitamin D and A have been sufficiently studied in specific gender and comorbidity based sub-groups of HIV infected population and such supplementation has been found to be beneficial. Multivitamin supplementation has been found to be associated with lower risk of death, a higher CD4 count, lower viral loads, delayed disease progression, improved weight gain and significant improvement in hematological status in pregnant females in Tanzania (Fawzi et al., 2004, 2007; Villamor et al., 2002). Similarly, among children, probiotic and micronutrient supplementation has shown significant improvement in CD4 count and delay in progression to advanced disease (Gautam et al., 2014). In adult HIV infected population, multivitamin supplementation has shown a reduction in oxidative stress (Allard et al., 1998). Individually, vitamin D deficiency has been found to be associated with a higher all-cause mortality and AIDs event (Viard et al., 2011), while its supplementation, and increase in serum concentration over time is associated with a decrease in markers of T-cell activation, monocyte activation and Interleukin-6 (an inflammatory biomarker) among adults (Benguella et al., 2018; Eckard et al., 2018), and improved neuromuscular motor skills among children and young adults (Brown et al., 2015). Majority of vitamin A supplementation studies have been conducted in pregnant females and provide mixed results.

A systematic review of randomized control trials found no overall evidence of a positive effect of intervention in mother to child transmission of disease, but they did find significant positive effect on birth weight (Kongnyuy et al., 2009). Few studies did not find any beneficial effects of vitamin supplementation (Guwatudde

et al., 2015). We found that among males, the beneficial effects of multivitamin supplementation were more evident as compared to females (as seen by the strength of association). This is a novel finding to our knowledge as no previous interventional study has assessed the role of gender. In this study, baseline age does not have any significant association with sign and symptom.

This study has several strengths. The use of large double blind randomized control trial data in HIV patients to study the effect of multivitamin is one of them. Our results could be reliable because of large sample. Application of the Bayesian CMP mixed effect model with higher order random effects to study longitudinal count data as an alternative to other methods is another strength. However, this study is not free from limitations. In our calculation, we observe that DIC, and WAIC for NB model is substantially high in comparison to other models, seems to be counter intuitive for over-dispersed data. Further investigation may be required. The model could not adjust for other predictors especially certain comorbidities that could have affected metabolism and mechanisms of actions of the vitamins as they were not available from the original investigators. From this study, the sign of positive impact of multivitamin supplementation in HIV patient's health status would enhance the application of multivitamin to the HIV patient population as a low cost therapy. However, larger study may be conducted to produce generalizeable results.

4.5 CMP Model Implementation

No package supports mixed effect model when longitudinal count data distribution is assumed as Conway Maxwell Poisson (CMP). We implement our analysis in R environment with a R package ***rstan***. It requires both Stan and R coding. We use the matrix notation of mixed effect model of the form with the given notation in Chapter 2

$$\log(\theta_{ij}) = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{z}_{ij}^T\boldsymbol{\zeta}, \tag{4.2}$$

where $\zeta \sim N(0, \sigma_\zeta^2)$, and θ_{ij} is the shape parameter of the CMP distribution associated with longitudinal count observation y_{ij}. CMP is not a default distribution in Stan. Therefore, we had to adopt CMP distribution in Stan. The CMP model implementation process are mentioned below step-wise (please follow the codes below to match with steps):

1. Install the R package ***rstan*** with all its dependencies (R packages), and call all required libraries.

```
# Library
library(rstan)
library(ggplot2)
```

2. Load data in R and generate a fixed effect design matrix. We can make a design matrix for random effects too. However, we made here only design matrix for fixed effects.

```
 # Loading data assuming that data has trt,time,age,gender variables
data=read.csv("~data.csv")
#Design matrix for Fixed effects
X=(model.matrix(~1+trt+time+age+gender, data))
```

3. Express data as a list with the data to be used in the model, we call here it ***standata***.

```
# creating a list of data for  modeling in Rstan
```

```
standata=list(Nobs=nrow(data),
              Npreds = ncol(X),
              Ngroups=length(unique(data$ID)),
              y = data$countresponse,
              X = X,
              trt=data$trtment,
              time=(data$visit_times),
              group = as.integer(factor(data$ID)),
              age=data$base_line_age)
```

4. Stan code can be written in two ways but we have to use one of them. Firstly: write the Stan code in R environment with a name **cmpcrs** (say), all the Stan code will be within inverted comma ('...'). Secondly: we can create a separate '$\sim$ *.stan*' file in notepad **cmp.stan** (say). Finally, call the code/file as **modelcode** = **cmpcrs** or **cmp.stan**. However, here we apply first case as

```
cmpcrs = '.....'
```

5. Within Stan code user defined distribution's loglikelihood can be developed. In CMP distribution we have a normalizing constant which is an infinite sum, we have to evaluate normalizing constant in 2 steps first with a main function. Within the main function the function ' real Z (real theta, real phi) {...}' evaluates the normalizing constant $Z(\theta_i, \phi) = \sum_{k=0}^{\infty} \frac{\theta_i^k}{(k!)^\phi}$ for subject i for θ_i and fixed ϕ. The successive terms in the sum are diminishing and we included the term in the sum until difference between two term is attained to < 0.0001

```
real Z(real theta, real phi){
real sm;
```

```
real sm_prev;
int i;
real diff;
sm=0;
i=0;
diff=1;
while (diff > 0.0001){
sm_prev=sm;
sm=sm+exp(i*log(theta)-phi*lgamma(i+1));
diff=sm-sm_prev;
i=i+1;
}
return(sm);
return(i);
}
```

In the second step, we have to evaluate the vector $\boldsymbol{Zv}$ for all observations.

```
 vector Zv(vector theta, real phi){
int N = rows(theta);
vector[N] zs;
for (i in 1:N){
zs[i] = Z(theta[i], phi);
}
return(zs);
```

Finally, the log-likelihood.

```
real compoisson_lpmf(int y, vector theta, real phi){
```

```
int N = rows(y);
return-sum(log(Zv(theta,phi)))-phi*sum(lgamma(y+1)) + sum(y.*log(theta));
}
```

Note: For default distributions in Stan we do not require this step.

6. In data block define data including design matrix.

```
data {
int<lower=0> Nobs;            // number of observations
int<lower=0> Npreds;          // number of fixed effects
int<lower=0> Ngroups;        // number of subjects
 int y[Nobs];             // response variable
matrix[Nobs,Npreds] X;    // fixed effect design matrix
vector[Nobs] time;      // measurement times or visit times
int<lower=1,upper=Ngroups> group[Nobs];  // group-wise observation
}
```

7. In parameter block define parameters of interest including regression coefficients, dispersion parameter, standard deviations of random effects that we want to estimate. In this case to accommodate correlated random effects we need to define Choleskey factor (L).

```
 parameters {
real<lower=0.01> phi;   // dispersion parameter
vector[Npreds] beta;    // fixed effect regression coefficients
vector<lower=0>[2] sigma_zeta; //  variance of random effects
cholesky_factor_corr[2] L_zeta;
// L,Choleskey factor of correlation matrix
matrix[2,Ngroups] z_zeta;
```

```
// matrix of randomly distributed random variables
}
```

8. In transform parameter block estimate theta from the proposed model, and create correlated random effects by using

$$\begin{bmatrix} \zeta_{01} & \zeta_{11} \\ \zeta_{02} & \zeta_{12} \\ \dots & \dots \\ \zeta_{0I} & \zeta_{1I} \end{bmatrix} = \Big(\text{diag}(\sigma_{\zeta 0}, \sigma_{\zeta 1})\boldsymbol{L}_{\zeta}\boldsymbol{z}_{\zeta}\Big)^{T}$$

```
// transform parameter block
transformed parameters {
vector[Nobs] thetahat ; // vector of estimated theta
matrix[2,Ngroups] zeta; // matrix of random intercepts and slopes
zeta<-diag_pre_multiply(sigma_zeta,L_zeta)*z_zeta;
// creating correlated random-intercepts and slopes
// estimating thetas
for (i in 1:Nobs)
thetahat[i]<- exp(X[i]*beta+zeta[1,group[i]]+ zeta[2,group[i]]*time[i]);
}
```

9. In model block define priors of the parameters and generate posterior, here is the end of Stan coding. we recommend uniform or normal priors for regression parameters, uniform or inverse-gamma for variance parameter, log normal or Halpf-Cauchy for dispersion parameter, LKJ with eta=2 or 1.5 for L
// model block

```
model {
```

```
beta~ normal(0,10000);          // fixed effect regression coefficients
L_zeta~lkj_corr_cholesky(2.0);  // priors Choleskey factor
to_vector(z_zeta)~normal(0,1);  // prior for to-vector
phi~lognormal(0,15);            // prior for dispersion parameter
sigma_zeta~inv_gamma(0.1,0.1);  // prior for dispersion parameter
y ~ compoisson(thetahat, phi);
```

10. Now run the model with modelcode=***cmpcrs*** and data= ***standata***. Number of chains and cores may be 1 or 2 or 3 or 4. Any number of thinning can be used to avoid dependent samples for posterior properties calculations. One can monitor progress of MCMC by using any number of refresh. The step size, maximum number of tree depth, and metropolis acceptance ratio can be controlled for better performance of MCMC samples. Default warm-up is 1/2 of the iterations. However, one can select any number of warm-ups.

```
# Rstan code for model running
fit.cmpcrs_mf=stan(model_code=cmpcrs, data= standata, iter = 5000,
chains = 4, cores=4, thin=1, refresh = 1000, init_r=0.01)
```

11. Print the output, one can specify parameters and 95% credible interval to print. To monitor performance of MCMC trace plots, ACF plots, Pair plots can also be produced.

//printing parameter estimates

```
  print(fit.cmpcrs_mf, pars=c("phi","beta[1]",
  "beta[2]", "beta[3]", "beta[4]","beta[5]",
'sigmaint'), probs=c(0.025,0.5,0.975))
```

```
 Library
library(rstan)
library(ggplot2)
# Loading data assuming that data has trt,time,age,gender variables
data=read.csv("~data.csv")
//
#Design matrix for Fixed effects
X=(model.matrix(~1+trt+time+age+gender, data))
//
# Creating a list of data for Rstan modeling
standata=list(Nobs=nrow(data),
              Npreds = ncol(X),
              Ngroups=length(unique(data$ID)),
              y = data$countresponse,
              X = X,
              trt=data$trtment,
              time=(data$visit_times),
              group = as.integer(factor(data$ID)),
              age=data$base_line_age)
//
#start of STAN coding block for the correlated random intercept
#and slope with the name "cmpcrs"
cmpcrs = '
#evaluate the normalizing constant (infinite sum)

functions{
```

```
real Z(real theta, real phi){
real sm;
real sm_prev;
int i;
real diff;
sm=0;
i=0;
diff=1;
while (diff > 0.0001){
sm_prev=sm;
sm=sm+exp(i*log(theta)-phi*lgamma(i+1));
diff=sm-sm_prev;
i=i+1;
}
return(sm);
return(i);
}
vector Zv(vector theta, real phi){
int N = rows(theta);
vector[N] zs;
for (i in 1:N){
zs[i] = Z(theta[i], phi);
}
return(zs);
}
#define log likelihood
real compoisson_lpdf(vector y, vector theta, real phi){
```

```
int N = rows(y);
return-sum(log(Zv(theta,phi)))-phi*sum(lgamma(y+1)) + sum(y.*log(theta));
}
}
#data block
data {
int<lower=0> Nobs;                // number of observations
int<lower=0> Npreds;              // number of fixed effects
int<lower=0> Ngroups;             // number of subjects
vector[Nobs] y;                   // response variable
matrix[Nobs,Npreds] X;            // fixed effect design matrix
vector[Nobs] time;                // measurement times or visit times
int<lower=1,upper=Ngroups> group[Nobs];  // group-wise observation
}
#parameter block
parameters {
real<lower=0.01> phi;       // dispersion parameter
vector[Npreds] beta;        // fixed effect regression coefficients
vector<lower=0>[2] sigma_zeta;  //  variance of random effects
cholesky_factor_corr[2] L_zeta;  // L,Choleskey factor of correlation matrix
matrix[2,Ngroups]z_zeta;
// matrix of randomly distributed random variabl
//used to generate correlated random effects
}
#transform parameter block
transformed parameters {
vector[Nobs] thetahat ;
```

```
// vector of estimated theta considered as parameters
matrix[2,Ngroups] zeta;      // matrix of random intercepts and slopes
zeta<-diag_pre_multiply(sqrt(sigma_zeta),L_zeta)*z_zeta;
// creating correlated random-
//-intercept and slope
# estimating thetas
for (i in 1:Nobs)
thetahat[i]<- exp(X[i]*beta+zeta[1,group[i]]+ zeta[2,group[i]]*time[i]);
}
# model block

model {
beta~ normal(0,10000);           // fixed effect regression coefficients
L_zeta~lkj_corr_cholesky(2.0);      // priors Choleskey factor
to_vector(z_zeta)~normal(0,1);      // prior for to-vector
phi~lognormal(0,15);             // prior for dispersion parameter
sigma_zeta~inv_gamma(0.1,0.1);      // prior for dispersion parameter
y ~ compoisson(thetahat, phi);
}'
#----- end of STAN coding block-------
# Rstan code for model running
 fit.cmpcrs_mf=stan(model_code=cmpcrs, data= standata, iter = 5000, chains = 4,
cores=4, thin=1, refresh = 1000, init_r=0.01)
#Output posterior means with 95% Credible intervals
print(fit.cmpcrs_mf, pars=c("phi","beta[1]", "beta[2]", "beta[3]",
"beta[4]","beta[5]",
'sigmaint'), probs=c(0.025,0.5,0.975))
```

Chapter 5

Conclusion and Future Research Directions

Biomedical studies, clinical trials, and observational studies generate abundance of count data. Often times, these data are measured longitudinally or cluster-wise. McCullough & Nelder (1989) mentions that over-dispersion is a rule rather than exception. Over-dispersion in the data arises for a variety of reasons. For instance, when mean and variance are related in generalized linear models, due to important predictor missing in the model, functional miss-specifications, correlation between responses, excess variation in counts, and violation of distributional assumptions. On the other hand under-dispersion is also arises when adjacent groups are correlated. It can occur by data generating or modeling process, and usually seen in small sample values. Failure to address over- and under-dispersion leads to bias inference and model over-fitting respectively. Poisson model is a benchmark in count data analysis, and is constrained by equi-dispersion assumption. Negative binomial model is dedicated to deal with over-dispersion. Literature reveals that in terms of model fit and predictive power, CMP model outperforms NB and Poisson as it can capture a wide spectrum of dispersion in a parsimonious way (K. F. Sellers & Shmueli, 2010). In addition, the

longer tail of the CMP distribution can capture extreme observations, and it enables us to fit a single model instead of fitting separate models in different dispersion conditions.

Besides, in modeling longitudinal count data, it is reasonable to think that there is natural heterogeneity among the study subjects not only on their baseline level but also in the change in the expected counts over time. In practice this situation happens a lot. To address the heterogeneity and dispersion in count data we include random effects in CMP regression model. The CMP model with subject-specific random intercept and slope, and generalized additive mixed effect model with CMP distributional assumption of count data were not studied in literature. Fitting such models in the classical approach may be cumbersome due to complex nature of the likelihood and integrational intractability. We propose a Bayesian approach for these models and attempt to assess model performance based on the simulation and real data analysis.

In Chapter 2, we fit a Bayesian generalized linear mixed effect model for longitudinal count data distributed as CMP by assuming a linear functional relationship between expected mean counts and covariate effects. We fit CMP, Poisson, NB model, and do model assessments by using LPML, LOO, WAIC, and DIC. From both simulated and real data analysis we experience that each of the model assessment criteria does not perform equally in all situations, rather their performances are data specific. The similar experience mentioned in Vehtari et al. (2017). Since we observe a consistent performance of DIC across dispersion conditions, we recommend to the use of DIC for Bayesian model assessments, although it suffers from some limitations mentioned in Vehtari et al. (2017); Plummer (2008); Van Der Linde (2005). In simulated study, based on DIC, we observed that CMP models fit better than other models. We use epilepsy data to check performance of our proposed model. Epilepsy dataset is well studied in literature. The data are over-dispersed, and no statistically significant

impact of study drug (progabide) was found in most of the analysis (Leppik et al., 1987). Our proposed model also conforms similar findings.

In Chapter 3, we fit a Bayesian generalized additive mixed effect model for longitudinal count data distributed as CMP by considering a non-linear functional relation between expected mean counts and covariate effects. Such a model is widely applicable in clinical trial data where functional relationship is not known in advance, rather is data driven. We use an under-dispersed hypothetical dataset to fit the proposed model. Our model could identify the under-dispersion in the data. Based on all four model assessment criteria (LMPL, LOO, WAIC, and DIC) the CMP model fits the best among others.

Whether the multivitamin supplementation is beneficial to explain symptomology in HIV patients receiving highly active antiretroviral therapy? By using CMP model, Chapter 4 deals with the above research question. We found, the CMP model fits the best, and model convergence is quite satisfactory. The model depicts that the incidence rate of suffering from different sign and symptom of HIV patients who consume multivitamin is lowed by 9% than those who consume placebo while age and gender are adjusted.

In this study we encounter some problems and could not explain the reason why they are arising, require further investigation. Some of the issues are: (1) although we do not encounter any problem in CPOs for simulated data, we experience some subjects retain abnormally high values of CPO for epilepsy data, and makes LPML questionable (2) in calculation of WAIC, variance of log posterior likelihood across simulations produce high values (when exceeds 0.04 are not reliable) and makes WAIC calculation unreliable for the real data (3) in Chapter 4 for a over-dispersed multivitamin dataset, DIC for NB model is substantially high in comparing to the other model.

Our proposed models, Bayesian generalized mixed effect model with higher order random effects, and Bayesian generalized additive mixed effect model are the extensions to the existing literature for longitudinal count data modeling. The CMP model saves us from model misspecification for count data. Due to unavailability of ready to use software, the application of the CMP model is limited, despite having the appealing properties of CMP distribution. We provide a tutorial for implementing ting CMP model in STAN and Rstan, and attach full code in Appendix. We hope, with the advent of computational procedure, the proposed model is easily extendable to study subject and cluster specific variability in multi-site clinical trials by adding cluster or site specific random effects in the model. While dealing with dispersed longitudinal count responses, our proposed models can account for both linear and non-linear relationship between expected counts and covariates arises from various biomedical, public health, and business research.

Some noted future research plans are:

1. Developing CPO using weight function and corresponding logarithm of pseudo marginal ikelihood (LMPL) as the Bayesian model assessment criteria with Conway-Maxwell Poisson distributional assumption for longitudinal count data. In Chapter 2, we used numerical calculation of CPO and LPML by using MCMC samples based on CPO identity-I described in (Zhang et al., 2017). However, we realized that CPO identity-II proposed in Zhang et al. (2017) remains unexplored in the context of CMP. CPO identity-II involves calculation of normalized weight function by using random effects, and approximation of covariance matrix of random effects which remains for further study.

2. Exploring posterior identifiability condition under non-informative priors on regression parameters (β_k) to develop a theoretical background. This involves advance inequality techniques along with matrix rank calculation, to be explored in future.

3. In dealing with Bayesian mixed effect model and GAM in Chapter 2 and Chapter 3 respectively, we did not deal with zero inflated and missing longitudinal counts, and duel links (modeling log link of both shape and dispersion parameter), and automated knot selection in CMP setting which remain for further extension.

4. Developing statistical joint model of time to event and longitudinal count data distributed as Conway-Maxwell Poisson and its application in drug development.

   There is a growing interest in statistical joint modeling in drug and medical instrument development industry. Literature reveals that joint model provides better insight for biomarkers and survival events, produce less bias, provides greater efficiency by lowering sample size in the drug development process (Lawrence Gould et al., 2015). A joint model includes a longitudinal model to deal with bio-marker (may be dispersed counts) trajectory and survival model to deal with time to events. However, research scope remains open in both components.

5. Methodology development to identify the distribution and shape of the longitudinal data.

   In practice, if longitudinal data is continuous, we assume normal distribution of the data, for continuous rates/ratios/proportion we consider beta distribution, categorical (binary, multinomial, ordinal) are dealt with binomial, multinomial distribution, and count data with Poisson, negative binomial, CMP distribution, etc;. But, rarely we pay attention to the shape of the distribution of the data. If we want to see the shape of the longitudinal data, how can we do that? Should we aggregate all data and plot them? Or should we plot data at each visit time and address them with some mixture distribution? This option may be feasible

for fixed time visits for patients, what about for continuous varying time points (different visit times for different patients)? May need attention.

6. In case of joint modeling with multiple bio-markers, there is scope for methodology development when multiple bio-markers are measured in different time points from the same subject.

7. Exploring both methodology and applied research windows relating to CMP model in Biomedical and Public Health Research.

   Use of CMP distribution is gaining popularity in biomedical and public health research in recent years. For example, some of the studies are: study of proportional hazard and proportional odds under CMP cure rate model (Balakrishnan et al., 2017; Balakrishnan & Feng, 2018; Pal et al., 2018), interval censored cure rate (Wiangnak & Pal, 2018), cure rate model computation (He & Emura, 2019), study of dispersion in positron emission therapy (Santarelli et al., 2016), modeling doctor's visit using right censored zero-inflated CMP (Saffari et al., 2018), modeling motor vehicle crash frequency (Abdella et al., 2019), zero inflated number of dental caries (Choo-Wosoba et al., 2016), to study fertility count data (Peluso et al., 2019), accident prediction in highway crossing (Lu & Tolliver, 2016), Weibull CMP to study survival data (Gupta & Huang, 2017), to model number of babies born alive in multiple pregnancy (Erkan et al., 2017), to study highway rail grade traffic hazard, to study annual reproduction rate. The above studies may be helpful to unveil potential research scopes, mostly in survival analysis and traffic safety studies.

Appendix A

This appendix includes full conditional for BGMM

$$\pi(\boldsymbol{\beta}|\boldsymbol{D},\boldsymbol{\zeta},\phi,\boldsymbol{y}) \propto \prod_{i=1}^{I}\prod_{j=1}^{n_i}\left(\frac{1}{y_{ij}!}\right)^{\phi} \times \left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{\zeta}_i\right)\right)^{y_{ij}} \times \left(\frac{\sum_{k=0}^{\infty}\left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{\zeta}_i\right)\right)^k}{(k!)^{\phi}}\right)^{-1} \times \frac{\exp\left(-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\beta_0})^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta_0})\right]\right)}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}}$$

$$\pi(\boldsymbol{\zeta}|\boldsymbol{\beta},\boldsymbol{D},\phi,\boldsymbol{y}) \propto \prod_{i=1}^{I}\prod_{j=1}^{n_i}\left(\frac{1}{y_{ij}!}\right)^{\phi} \times \left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{\zeta}_i\right)\right)^{y_{ij}} \times \left(\frac{\sum_{k=0}^{\infty}\left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{\zeta}_i\right)\right)^k}{(k!)^{\phi}}\right)^{-1} \times \prod_{i=1}^{I}\frac{\exp\left(-\frac{1}{2}[\boldsymbol{\zeta}_i^T\boldsymbol{D}^{-1}\boldsymbol{\zeta}_i]\right)}{|\boldsymbol{D}|^{\frac{1}{2}}}$$

$$\pi(\boldsymbol{D}|\boldsymbol{\beta},\boldsymbol{\zeta},\boldsymbol{y}) \propto \prod_{i=1}^{I}\frac{\exp\left(-\frac{1}{2}\left[\boldsymbol{\zeta}_i^T\boldsymbol{D}^{-1}\boldsymbol{\zeta}_i\right]\right)}{|\boldsymbol{D}|^{\frac{1}{2}}} \times \frac{|\boldsymbol{\omega}|^{\varphi/2}}{\Gamma q(\varphi/2)} \times |\boldsymbol{D}|^{-\frac{(\varphi+q+1)}{2}}\exp\left(-\frac{1}{2}\mathrm{Tr}\left(\boldsymbol{\omega}\boldsymbol{D}^{-1}\right)\right)$$

$$\pi(\phi|\boldsymbol{\beta},\boldsymbol{\zeta},\boldsymbol{y}) \propto \prod_{i=1}^{I}\prod_{j=1}^{n_i}\left(\frac{1}{y_{ij}!}\right)^{\phi} \times \left(\frac{\sum_{k=0}^{\infty}\left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{\zeta}_i\right)\right)^k}{(k!)^{\phi}}\right)^{-1} \times \frac{\exp\left(-\frac{1}{2}(\frac{\log\phi-\mu_{\phi}}{\psi})^2\right)}{\psi\phi}$$

Appendix B

Table B.1: 95% Credible Interval from simulated Under-dispersed ($\phi = 2.50$) data

Model	Param	True	Mean	95%Crl.	HPD
CMP	ϕ	2.50	2.82	(1.89, 3.87)	(2.79, 3.87)
	β_0	-0.80	$-.87$	$(-1.31, -0.45)$	$(-1.30, -0.4)$
	β_1	-0.05	0.03	$(-0.33,0.39)$	$(-0.33,0.39)$
	β_2	-0.01	0.01	$(-0.42,0.45)$	$(-0.42,0.45)$
	σ_0	0.05	0.13	(0.01,0.48)	(0.00,0.41)
	σ_1	0.02	0.14	(0.01,0.52)	(0.00,0.44)
	ρ	-0.50	-0.02	$(-0.83,0.80)$	$(-0.84,0.78)$
NB	ϕ	-	0.04	(-)	(-)
	β_0	-0.80	-1.12	$(-1.50,-0.76)$	$(-1.49,-0.75)$
	β_1	-0.05	0.02	$(-0.29,0.33)$	$(-0.29,0.33)$
	β_2	-0.01	0.02	$(-0.36,0.41)$	$(-0.36,0.40)$
	σ_0	0.06	0.07	(0.01,0.26)	(0.00,0.22)
	σ_1	0.02	0.08	(0.01,0.29)	(0.00,0.25)
	ρ	-0.50	-0.04	$(-0.83,0.80)$	$(-0.85,0.77)$
Poisson	β_0	-0.80	-1.11	$(-1.50,-0.77)$	$(-1.48,-0.76)$
	β_1	-0.05	0.02	$(-0.35,0.40)$	$(-0.36,0.40)$
	β_2	-0.01	0.02	$(-0.35,0.40)$	$(-0.36,0.40)$
	σ_0	0.05	0.07	(0.01,0.27)	(0.00,0.24)
	σ_1	0.02	0.07	(0.01,0.30)	(0.00,0.24)
	ρ	-0.50	-0.04	$(-0.83,0.80)$	$(-0.85,0.78)$

Table B.2: 95% Credible Interval from simulated Over-dispersed $\phi = 0.30$ data

Model	Param	True	Mean	95%Crl	HPD
CMP	ϕ	0.30	0.39	(0.10, 0.75)	(0.09, 0.71)
	β_0	-0.80	-0.83	$(-1.12,-0.53)$	$(-1.13,-0.53)$
	β_1	-0.05	-0.01	$(-0.23,0.22)$	$(-0.23,0.22)$
	β_2	-0.01	-0.02	$(-0.28,0.25)$	$(-0.29,0.25)$
	σ_0	0.05	0.16	(0.01,0.43)	(0.00,0.38)
	σ_1	0.02	0.17	(0.01,0.49)	(0.00,0.42)
	ρ	-0.50	-0.17	$(-0.89,0.75)$	$(-0.93,0.67)$
NB	ϕ	-	0.45	(0.01,0.12)	(0.01,0.09)
	β_0	-0.80	-0.58	$(-0.92,-0.25)$	$(-0.91,-0.24)$
	β_1	-0.05	-0.01	$(-0.30,0.28)$	$(-0.29,0.28)$
	β_2	-0.01	-0.02	$(-0.36,0.32)$	$(-0.36,0.32)$
	σ_0	0.05	0.22	(0.01,0.58)	(0.00,0.50)
	σ_1	0.02	0.23	(0.01,0.54)	(0.00,0.55)
	ρ	-0.50	-0.18	$(-0.89,0.75)$	$(-0.93,0.66)$
Poisson	β_0	-0.80	-0.64	$(-1.02,-0.32)$	$(-1.00,-0.30)$
	β_1	-0.05	-0.01	$(-0.28,0.27)$	$(-0.28,0.27)$
	β_2	-0.01	0.00	$(-0.32,0.39)$	$(-0.34,0.36)$
	σ_0	0.05	0.40	(0.04,0.82)	(0.02,0.75)
	σ_1	0.02	0.41	(0.05,0.91)	(0.03,0.84)
	ρ	-0.50	-0.33	$(-0.91,0.65)$	$(-0.95,0.53)$

Table B.3: 95% Credible Interval from simulated Equi-dispersed $\phi = 1.00$ data

Model	Param	True	Mean	95%Crl	HPD
CMP	ϕ	1.0	1.23	(0.69,1.82)	(0.68,1.80)
	β_0	-0.80	-0.81	$(-1.20,-0.44)$	$(-1.19,-0.44)$
	β_1	-0.05	0.00	$(-0.31,0.30)$	$(-0.30,0.31)$
	β_2	-0.01	-0.04	$(-0.40,0.33)$	$(-0.40,0.32)$
	σ_0	0.05	0.23	(0.01,0.64)	(0.00,0.55)
	σ_1	0.02	0.25	(0.01,0.71)	(0.00,0.62)
	ρ	-0.50	-0.18	$(-0.90,0.74)$	$(-0.94,0.66)$
NB	ϕ	-	0.07	(0.01,0.25)	(0.01,0.21)
	β_0	-0.80	-0.85	$(-1.20,-0.53)$	$(-1.19,-0.52)$
	β_1	-0.05	-0.00	$(-0.29,0.29)$	$(-0.29,0.29)$
	β_2	-0.01	-0.05	$(-0.39,0.31)$	$(-0.39,0.30)$
	σ_0	0.06	0.18	(0.01,0.48)	(0.00,0.42)
	σ_1	0.02	0.20	(0.01,0.56)	(0.00,0.47)
	ρ	-0.50	-0.16	$(-0.89,0.75)$	$(-0.93,0.68)$
Poisson	β_0	-0.80	-0.85	$(-1.20,-0.53)$	$(-1.19,-0.53)$
	β_1	-0.05	-0.00	$(-0.29,0.29)$	$(-0.29,0.29)$
	β_2	-0.01	-0.05	$(-0.38,0.31)$	$(-0.39,0.30)$
	σ_0	0.05	0.19	(0.01,0.51)	(0.00,0.44)
	σ_1	0.02	0.21	(0.01,0.56)	(0.00,0.49)
	ρ	-0.50	-0.16	$(-0.89,0.75)$	$(-0.93,0.68)$

Appendix C

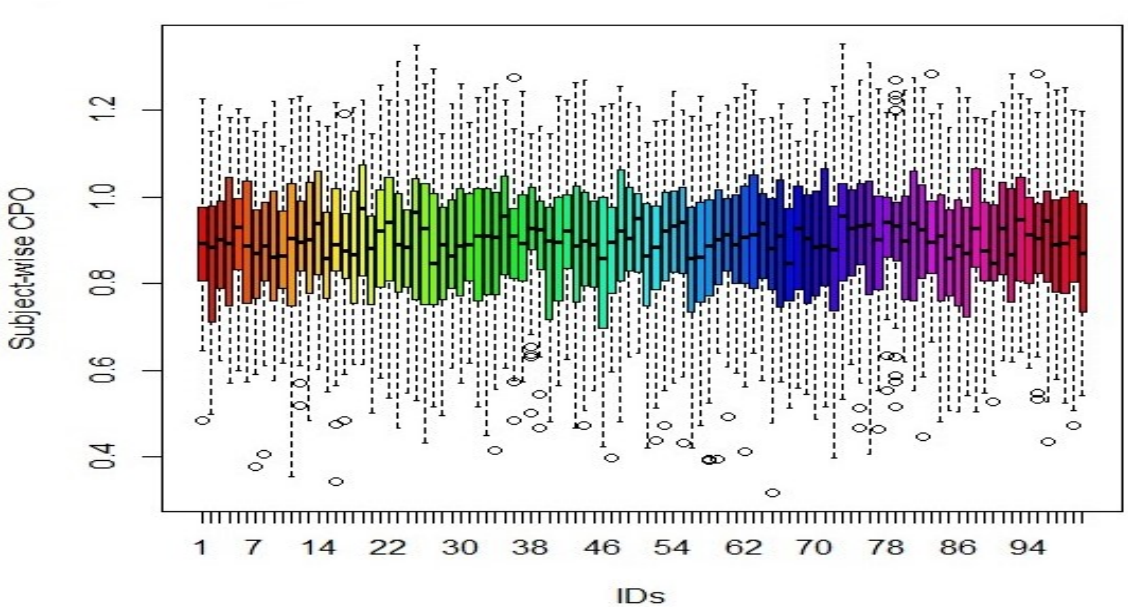

Figure C.1: Boxplot for subject-wise CPO for CMP model with $\phi = 0.30$

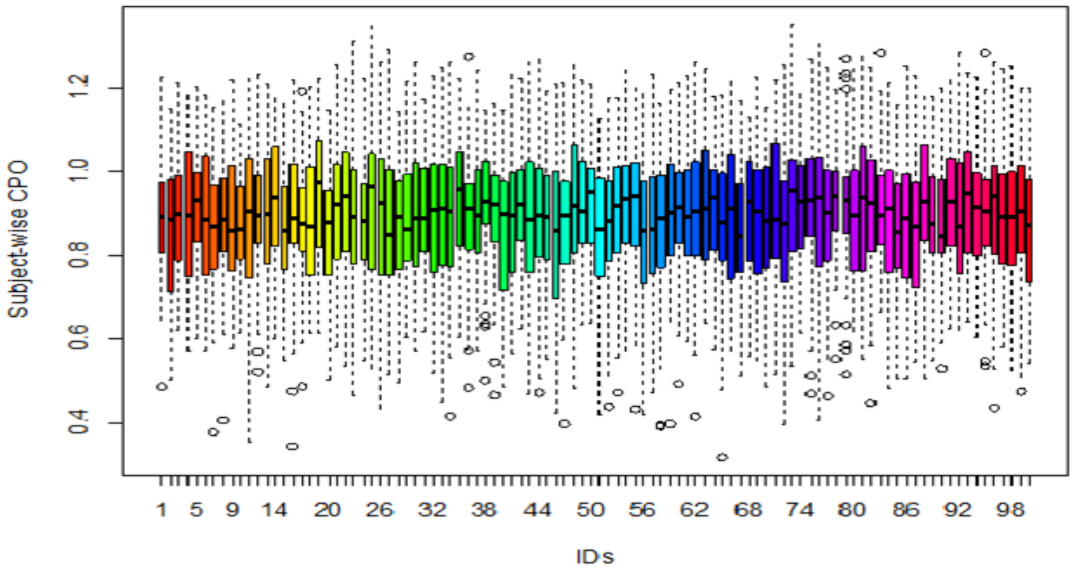

Figure C.2: Boxplot for subject-wise CPO for CMP model with $\phi = 1.00$

Appendix D

This appendix includes full conditional for BGAMM

$$\pi(\boldsymbol{\beta}|\boldsymbol{\gamma},\boldsymbol{\zeta},\phi,\boldsymbol{y}) \propto \prod_{i=1}^{I}\prod_{j=1}^{n_i}\left(\frac{1}{y_{ij}!}\right)^{\phi} \times \left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{\gamma}+\zeta_i\right)\right)^{y_{ij}} \times \left(\frac{\sum_{k=0}^{\infty}\left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{\gamma}+\zeta_i\right)\right)^k}{(k!)^{\phi}}\right)^{-1} \times \frac{\exp\left(-\frac{1}{2}\left[(\boldsymbol{\beta}-\boldsymbol{\beta_0})^T\boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}-\boldsymbol{\beta_0})\right]\right)}{|\boldsymbol{\Sigma}_0|^{\frac{1}{2}}} \times \frac{1}{C}$$

$$\pi(\boldsymbol{\zeta}|\boldsymbol{\beta},\boldsymbol{\gamma},\phi,\boldsymbol{y}) \propto \prod_{i=1}^{I}\prod_{j=1}^{n_i}\left(\frac{1}{y_{ij}!}\right)^{\phi} \times \left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{\gamma}+\zeta_i\right)\right)^{y_{ij}} \times \left(\frac{\sum_{k=0}^{\infty}\left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{\gamma}+\zeta_i\right)\right)^k}{(k!)^{\phi}}\right)^{-1} \times \prod_{i=1}^{I}(\sigma_\zeta)^{-(1/2)} \times \exp\frac{-\zeta_i^2}{2\sigma_\zeta^2} \times \frac{1}{C}$$

$$\pi(\boldsymbol{\gamma}|\boldsymbol{\beta},\boldsymbol{\zeta},\phi,\boldsymbol{y}) \propto \prod_{i=1}^{I}\prod_{j=1}^{n_i}\left(\frac{1}{y_{ij}!}\right)^{\phi} \times \left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{\gamma}+\zeta_i\right)\right)^{y_{ij}} \times \left(\frac{\sum_{k=0}^{\infty}\left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{\gamma}+\zeta_i\right)\right)^k}{(k!)^{\phi}}\right)^{-1} \times \frac{\exp\left(-\frac{1}{2}\left[(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)^T\boldsymbol{\Sigma}_\gamma^{-1}(\boldsymbol{\gamma}-\boldsymbol{\gamma}_0)\right]\right)}{|\boldsymbol{\Sigma}_\gamma|^{\frac{1}{2}}} \times \frac{1}{C}$$

$$\pi(\phi|\boldsymbol{\beta},\boldsymbol{\gamma},\boldsymbol{\zeta},\boldsymbol{y}) \propto \prod_{i=1}^{I}\prod_{j=1}^{n_i}\left(\frac{1}{y_{ij}!}\right)^{\phi} \times \left(\frac{\sum_{k=0}^{\infty}\left(\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}+\boldsymbol{z}_{ij}^T\boldsymbol{\gamma}+\zeta_i\right)\right)^k}{(k!)^{\phi}}\right)^{-1}$$
$$\times \frac{\exp\left(-\frac{1}{2}(\frac{\log\phi-\mu_\phi}{\psi})^2\right)}{\psi\phi} \times \frac{1}{C}$$

Appendix E

```
#setwd("C:....")


setwd("C:/Users/Desktop/Simulation_dt_1.5_cont_woi")

rm(list=ls())

# ----------------------Library----------------------------------------#


library(compoisson)

library(ggplot2)

library(MASS)

library(StanHeaders)

library(rstan)

library(shapefiles)

library(BayesX)

library(dplyr)

library(tidyverse)

library(MCMCvis)

library(parallel)
```

```
#--------------------------CMP longitudinal data simulation----------------#

#---coefficients------

b0=-0.80

b1=-0.05

b2= -0.01

#---------

N=100 #number of subjects

J=5 # number of measurements


#-----Data generation------

id=rep(1:N, rep(J,N))

x<-rep(c(0,1),N)

x1=x[rep(1:N,rep(J,N))]

tm=c(0,1,2,3,4)

t<-rep(tm,N)

t1=ifelse(t>0,1,0)


#---correlated random effects simulation--------

set.seed(123)

library(mvtnorm)

q = 0.05

r = -0.50

s = 0.02


cov <- matrix(c(q^2, r * q * s, r * q * s, s^2), nrow = 2,

              byrow = TRUE)
```

```
re <- rmvnorm(N, mean = c(0, 0), sigma = cov)
u1<-round(rep(re[,1], rep(J,N)),4)
u2<-round(rep(re[,2], rep(J,N)),4)

#---shape parameter estimation-------

theta<-cbind(exp(b1+u1+b2*x1+(b3+u2)*t1))

phi=1.5 # dispersion parameter

#View(cdata)

cmp.DIC=NB.DIC=pois.DIC=rep(NA,100)
cmp.loglik=NB.loglik=pois.loglik=rep(NA,100)
cmp.EP=NB.EP=pois.EP=rep(NA,100)
# Saving output
cmp.out=matrix(NA, nrow = 100,ncol = 63);
colnames(cmp.out)=c('phi', 'sdphi', 'cbphi2.5', 'cphi.5', 'cphi97.5',
'hphi95l', 'hphi95u', 'rphi',  'esphi',
'b0',' sdb0', 'cb02.5', 'cb0.5', 'cb097.5', 'hb095l', 'hb095u', 'rb0', 'esb0',
'b1', 'sdb1', 'cb12.5', 'cb1.5', 'cb197.5', 'hb195l', 'hb195u', 'rb1', 'esb1',
'b2', 'sdb2', 'cb22.5', 'cb2.5', 'cb297.5', 'hb295l', 'hb295u', 'rb2', 'esb2'
'sigu0', 'sdsigu0', 'csigu02.5', 'csigu0.5', 'csigu097.5', 'hsigu095l',
'hsigu095u', 'rsigu0', 'essigu0',
'sigu1', 'sdsigu1', 'csigu12.5', 'csigu1.5', 'csigu197.5', 'hsigu195l',
'hsigu195u', 'rsigu1', 'essigu1',
 'crr', 'sdcrr', 'ccrr2.5', 'ccrr.5', 'ccrr97.5', 'hcrr95l', 'hcrr95u',
```

```
 'rcrr','escrr')

NB.out=matrix(NA, nrow = 100,ncol = 63);
colnames(NB.out)=c('phi', 'sdphi', 'cbphi2.5', 'cphi.5', 'cphi97.5',
'hphi95l', 'hphi95u', 'rphi',  'esphi',
'b0',' sdb0', 'cb02.5', 'cb0.5', 'cb097.5', 'hb095l', 'hb095u', 'rb0', 'esb0',
'b1', 'sdb1', 'cb12.5', 'cb1.5', 'cb197.5', 'hb195l', 'hb195u', 'rb1', 'esb1',
'b2', 'sdb2', 'cb22.5', 'cb2.5', 'cb297.5', 'hb295l', 'hb295u', 'rb2', 'esb2',
'sigu0', 'sdsigu0', 'csigu02.5', 'csigu0.5', 'csigu097.5', 'hsigu095l',
'hsigu095u', 'rsigu0', 'essigu0',
'sigu1', 'sdsigu1', 'csigu12.5', 'csigu1.5', 'csigu197.5', 'hsigu195l',
'hsigu195u', 'rsigu1', 'essigu1',
'crr', 'sdcrr', 'ccrr2.5', 'ccrr.5', 'ccrr97.5', 'hcrr95l', 'hcrr95u',
'rcrr','escrr')

pois.out=matrix(NA, nrow = 100,ncol = 54);
colnames(pois.out)=c('b0',' sdb0', 'cb02.5', 'cb0.5', 'cb097.5', 'hb095l',
'hb095u', 'rb0','esb0',
'b1', 'sdb1', 'cb12.5', 'cb1.5', 'cb197.5', 'hb195l', 'hb195u', 'rb1', 'esb1',
'b2', 'sdb2', 'cb22.5', 'cb2.5', 'cb297.5', 'hb295l', 'hb295u', 'rb2', 'esb2',
'sigu0', 'sdsigu0', 'csigu02.5', 'csigu0.5', 'csigu097.5', 'hsigu095l',
'hsigu095u', 'rsigu0', 'essigu0',
'sigu1', 'sdsigu1', 'csigu12.5', 'csigu1.5', 'csigu197.5', 'hsigu195l',
'hsigu195u', 'rsigu1', 'essigu1',
'crr', 'sdcrr', 'ccrr2.5', 'ccrr.5', 'ccrr97.5', 'hcrr95l', 'hcrr95u',
'rcrr','escrr')
```

```
#---------------------- Stan data preparation-------------------------

m=5000
T=1  #thin
C=4 # of chain
k=(m*C/(2*T))

for (ir in 1: 100) {

  y=rcom(N*J,theta,phi) # CMP response
  mean(y)
  var(y)
  #-----Final data---------------------------

 cdata<-data.frame(id,x1, t1, tm,t,u1,u2, y)

  X<-(model.matrix(~1+x1+t1, cdata ))
  standat <- list(N = nrow(cdata),
                  P = ncol(X),
                  J = length(unique(cdata$id)),
                  y = cdata$y,
                  X = X,
                  x1=(cdata$x1),
                  t=(cdata$t1),
                  d = as.integer(factor(cdata$id)))
  standat
```

```
#-----------------------------------STAN code: CMP_U-------------

cat("----------------CMP model is running-----\n")
cmp = '
functions{
real Z(real theta, real phi){
real sm;
real sm_prev;
int i;
real diff;
sm=0;
i=0;
diff=1;
while (diff > 0.0001){
sm_prev=sm;
sm=sm+exp(i*log(theta)-phi*lgamma(i+1));
diff=sm-sm_prev;
i=i+1;
}
return(sm);
return(i);
}
vector Zv(vector theta, real phi){
int N = rows(theta);
vector[N] zs;
for (i in 1:N){
```

```
zs[i] = Z(theta[i], phi);
}
return(zs);
}
// define log likelihood
real compoisson_lpmf(int  y, vector theta, real phi){
int N = rows(y);
return-sum(log(Zv(theta,phi)))-phi*sum(lgamma(y+1)) + sum(y.*log(theta));
}
}

data {
int<lower=0> N;
int<lower=0> P;
int<lower=0> J;
int y[N];
matrix[N,P] X;
vector[N] t;
int<lower=1,upper=J>id[N];
}
parameters {
real<lower=0.01> phi;
vector[P] beta;
vector<lower=0>[2] sigma_u;
cholesky_factor_corr[2] L_u;
matrix[2,J] z_u;
}
```

```
transformed parameters {
matrix[2,J] u;
u<-diag_pre_multiply(sigma_u,L_u)*z_u;
}
model {
vector[N] thetahat ;
beta~ normal(0,10000);
L_u~lkj_corr_cholesky(2.0);
to_vector(z_u)~normal(0,1);
phi~lognormal(0,15);
sigma_u~inv_gamma(0.1,0.1);

for (i in 1:N)
thetahat[i]<- exp(X[i]*beta+u[1,id[i]]+ u[2,id[i]]*t[i]);
y ~ compoisson(thetahat, phi);
}'

#---------------Stan Model fitting:CMP_U--------

fit.cmp_u=stan(model_code=cmp, data= standat,
iter = m, chains = C, cores=4, thin=T,
refresh = 1000,init_r=0.01)

#--------------HPD calulation: CMP_U-----------
mc<-as.matrix(fit.cmp_u)
mc<-data.frame(mc)
cm<-read.csv("C:/Users/morshed.alam/Desktop/mc.csv")
```

```
cm <- mc%>% select(-contains("z"))
%>%select(-contains("L"))
%>%select(-contains("sigma"))
%>%select(-contains("l"))
cm<-data.frame(subset(cm,select=c("phi","beta.1.","beta.2.","beta.3.")),
select(cm,contains("u.1.")),
select(cm,contains("u.2.")),
select(cm,contains("sigma_u.1.")),
select(cm,contains("sigma_u.2.")))

#----------DIC calculation: CMP_U--------

y <-cdata$y
t<-cdata$t1
loglikc <- NULL

for (i in 1:k){
  Betac <- cm[i,2:4]
  uic <- rep(unlist(cm[i,5:104]), each=5)
  usc <- rep(unlist(cm[i,105:204]), each=5)
  phic <- cm[i,1]
  lambdac <- exp(X%*%t(Betac) + uic+usc*t)
  likc<- sum(com.log.density(y,lambdac,phic, log.z=NULL))
 loglikc <- c(loglikc,likc)
}

mean_simc<-mean(loglikc)
```

```
# mean parameter calculation
theta_hatc = as.matrix(apply(cm, 2, mean))
Beta_hatc <- theta_hatc[2:4,1]
phi_hatc<-theta_hatc[1]
ui_hatc<-rep(unlist(theta_hatc[5:104,1]), each=5)
us_hatc<-rep(unlist(theta_hatc[105:204,1]), each=5)
lambda_hatc <- exp(X%*%Beta_hatc+ui_hatc+us_hatc*t)
likc_hat<- sum(com.log.density(y,lambda_hatc,phi_hatc, log.z=NULL))
Pc<-2*(likc_hat-mean_simc)
cmp.loglik[ir]<-likc_hat
cmp.EP[ir]<-2*(likc_hat-mean_simc)
cmp.DIC[ir]<--2*(likc_hat-Pc)

cat("----------------NB model is running-----\n")
#-----------------------------------------------

nb='
data {
  int<lower=0> N;
  int<lower=0> P;
  int<lower=0> J;
  int y[N];
  matrix[N,P] X;
  vector[N] t;
   int<lower=1,upper=J>id[N];
}
parameters {
```

```
  real<lower=0.01> phi;
  vector[P] beta;
  vector<lower=0>[2] sigma_u;
  cholesky_factor_corr[2] L_u;
  matrix[2,J] z_u;
}
transformed parameters {
  vector[N] thetahat;
  matrix[2,J]u;
  u<-diag_pre_multiply(sigma_u,L_u)*z_u;
  for (i in 1:N)
    thetahat[i]<-(X[i]*beta+u[1,id[i]]+ u[2,id[i]]*t[i]);
}
model {

beta~normal(0,10000);
phi~lognormal(0,15);
L_u~lkj_corr_cholesky(2.0);
to_vector(z_u)~normal(0,1);
sigma_u~inv_gamma(0.1,0.1);
y ~ neg_binomial_2_log(thetahat, 1/phi);
}'


fit.nb_u=stan(model_code=nb, data= standat, iter = m, chains = C,
cores=4, thin=T, refresh = 1000, init_r=0.01)
#---------------------------------HPD calulation: NB_U------------
```

```
mn<-as.matrix(fit.nb_u)
mn<-data.frame(mn)
nm <- mn%>% select(-contains("z"))
%>%select(-contains("L"))
%>%select(-contains("sigma"))
%>%select(-contains("l"))
nm<-data.frame(subset(nm,select=c("phi","beta.1.","beta.2.","beta.3.")),
select(nm,contains("u.1.")),
select(nm,contains("u.2.")),select(nm,contains("sigma_u.1.")),
select(nm,contains("sigma_u.2.")))

#------DIC_NB-----------------#
y <-cdata$y
t<-cdata$t1
loglikn <- NULL
for (i in 1:k){
  Betan <- nm[i,2:4]
  uin <- rep(unlist(nm[i,5:104]), each=5)
  usn <- rep(unlist(nm[i,105:204]), each=5)
  phin <- nm[i,1]
  lambdan <- exp(X%*%t(Betan) + uin+usn*t)
  likn <- sum(y * log(phin*lambdan) - (y + 1/phin)*log(1+phin*lambdan) +
                lgamma(y + 1/phin) - lgamma(1/phin) - lgamma(y+1))
  loglikn <- c(loglikn,likn)
}
mean_simn<-mean(loglikn)
```

```
# mean parameter calculation
theta_hatn = as.matrix(apply(nm, 2, mean))
Beta_hatn <- theta_hatn[2:4,1]
phi_hatn<-theta_hatn[1]
ui_hatn<-rep(unlist(theta_hatn[5:104,1]), each=5)
us_hatn<-rep(unlist(theta_hatn[105:204,1]), each=5)
lambda_hatn <- exp(X%*%Beta_hatn + ui_hatn+us_hatn*t)
likn_hat <- sum(y * log(phi_hatn*lambda_hatn) -
(y + 1/phi_hatn)*log(1+phi_hatn*lambda_hatn)
+lgamma(y + 1/phi_hatn) - lgamma(1/phi_hatn) - lgamma(y+1))
likn_hat
Pn<-2*(likn_hat-mean_simn)

DICn<--2*(likn_hat-Pn)

NB.loglik[ir]<-likn_hat

NB.EP[ir]<-2*(likn_hat-mean_simn)

NB.DIC[ir]<--2*(likn_hat-Pn)



cat("----------------Poisson model is running-----\n")
#-----------------------------------------------------------------#
pos='
data {
  int<lower=0> N;
```

```
  int<lower=0> P;
  int<lower=0> J;
  int y[N];
  matrix[N,P] X;
  vector[N] t;
  int<lower=1,upper=J>id[N];
}
parameters {
  vector[P] beta;
  vector<lower=0>[2] sigma_u;
  cholesky_factor_corr[2] L_u;
  matrix[2,J] z_u;
}
transformed parameters {
  vector[N] thetahat;
  matrix[2,J]u;
  u<-diag_pre_multiply(sigma_u,L_u)*z_u;

  for (i in 1:N)
    thetahat[i]<- exp(X[i]*beta+u[1,id[i]]+ u[2,id[i]]*t[i]);
}

model {
 beta~normal(0,10000);
  L_u~lkj_corr_cholesky(2.0);
  to_vector(z_u)~normal(0,1);
  sigma_u~inv_gamma(0.01,0.01);
```

```
   y ~ poisson(thetahat);
  }'

#---------------------- Stan Model fitting:Pos_U-----------------

fit.pos_u=stan(model_code=pos, data= standat, iter = m, chains = C,
cores=4, thin=T, refresh = 1000,init_r=0.01)

#---------------------------------HPD calulation: NB_U------------
mp<-as.matrix(fit.pos_u)
mp<-data.frame(mp)
pm <- mp%>% select(-contains("z"))
%>%select(-contains("L"))
%>%select(-contains("sigma"))
%>%select(-contains("l"))

pm<-data.frame(subset(pm,select=c("beta.1.","beta.2.","beta.3.")),
select(pm,contains("u.1.")),
select(pm,contains("u.2.")),select(pm,contains("sigma_u.1.")),
select(pm,contains("sigma_u.2.")))



#----------------DIC Poisson--------------------------#
y <-cdata$y
t<-cdata$t1

  # Calculation of mean loglikelihood from simulation
```

```
  loglikp <- NULL
   for (i in 1:k){
    Betap <- pm[i,1:3]
    uip <- rep(unlist(pm[i,4:103]), each=5)
    usp <- rep(unlist(pm[i,104:203]), each=5)
    lambdap <- exp(X%*%t(Betap)+uip+usp*t)
    likp <- sum(y * log(lambdap) - lambdap - lgamma(y+1))
    loglikp <- c(loglikp,likp)
  }

mean_simp<-mean(loglikp) #mean of the loglikehood for all simulation
  # Calculation of posterior mean estimation and loglikelihood
  theta_hatp = as.matrix(apply(pm, 2, mean))
  Beta_hatp <- theta_hatp[1:3,1]
   ui_hatp<-rep(unlist(theta_hatp[4:103,1]), each=5)
  us_hatp<-rep(unlist(theta_hatp[104:203,1]), each=5)
  lambda_hatp <- exp(X%*%Beta_hatp + ui_hatp+ us_hatp*t)
  likp_hat <- sum(y * log(lambda_hatp) - lambda_hatp - lgamma(y+1))
  likp_hat # loglikelihood at poterior mean

  Pp<-2*(likp_hat-mean_simp) # number of effective parameters
  pois.loglik[ir]<-likp_hat
  pois.EP[ir]<-2*(likp_hat-mean_simp)
  pois.DIC[ir]<--2*(likp_hat-Pp)

#--------Results need to save----------------------#
```

```
  #-----------------------CMP Model output--------------------#
  M0=as.matrix(MCMCvis::MCMCsummary(fit.cmp_u,
                                params = c('phi','beta', 'sigma_u','L_u'),
                                probs = c(0.025, 0.5, 0.975),
                                round = 3))
   M0=t(M0)
  M1=as.matrix(MCMCvis::MCMCsummary(fit.cmp_u,
                                params = c('phi','beta', 'sigma_u','L_u'),
                                HPD = TRUE,
                                hpd_prob = 0.95,
                                round = 3))
  M1=t(M1)

result_c<-c(M0[1,1],M0[2,1],M0[3,1],M0[4,1],M0[5,1],M1[3,1],M1[4,1],
M0[6,1],M0[7,1],M0[1,2],M0[2,2],M0[3,2],M0[4,2],M0[5,2],M1[3,2],M1[4,2],
M0[6,2],M0[7,2],M0[1,3],M0[2,3],M0[3,3],M0[4,3],M0[5,3],M1[3,3],M1[4,3],
M0[6,3],M0[7,3],M0[1,4],M0[2,4],M0[3,4],M0[4,4],M0[5,4],M1[3,4],M1[4,4],
M0[6,4],M0[7,4],M0[1,5],M0[2,5],M0[3,5],M0[4,5],M0[5,5],M1[3,5],M1[4,5],
M0[6,5],M0[7,5],M0[1,6],M0[2,6],M0[3,6],M0[4,6],M0[5,6],M1[3,6],M1[4,6],
M0[6,6],M0[7,6],M0[1,9],M0[2,9],M0[3,9],M0[4,9],M0[5,9],M1[3,9],M1[4,9],
M0[6,9],M0[7,9])
 cmp.out[ir,]<-result_c

  #-----------------------NB  Model output--------------------#
  M0=as.matrix(MCMCvis::MCMCsummary(fit.nb_u,
                                params = c('phi','beta', 'sigma_u','L_u'),
                                probs = c(0.025, 0.5, 0.975),
```

```
                              round = 3))

  M0=t(M0)


  M1=as.matrix(MCMCvis::MCMCsummary(fit.nb_u,
                              params = c('phi','beta', 'sigma_u','L_u'),
                              HPD = TRUE,
                              hpd_prob = 0.95,
                              round = 3))
  M1=t(M1)

result_n<-c(M0[1,1],M0[2,1],M0[3,1],M0[4,1],M0[5,1],M1[3,1],M1[4,1],
M0[6,1],M0[7,1],M0[1,2],M0[2,2],M0[3,2],M0[4,2],M0[5,2],M1[3,2],M1[4,2],
M0[6,2],M0[7,2],M0[1,3],M0[2,3],M0[3,3],M0[4,3],M0[5,3],M1[3,3],M1[4,3],
M0[6,3],M0[7,3],M0[1,4],M0[2,4],M0[3,4],M0[4,4],M0[5,4],M1[3,4],M1[4,4],
M0[6,4],M0[7,4],M0[1,5],M0[2,5],M0[3,5],M0[4,5],M0[5,5],M1[3,5],M1[4,5],
M0[6,5],M0[7,5],M0[1,6],M0[2,6],M0[3,6],M0[4,6],M0[5,6],M1[3,6],M1[4,6],
M0[6,6],M0[7,6],M0[1,9],M0[2,9],M0[3,9],M0[4,9],M0[5,9],M1[3,9],M1[4,9],
M0[6,9],M0[7,9])
NB.out[ir,]<-result_n

  #-----------------------Poisson  Model output--------------------#

 M0=as.matrix(MCMCvis::MCMCsummary(fit.pos_u,
                              params = c('beta', 'sigma_u','L_u'),
                              probs = c(0.025, 0.5, 0.975),
```

```
                                round = 3))
  M0=t(M0)
 M1=as.matrix(MCMCvis::MCMCsummary(fit.pos_u,
                                params = c('beta', 'sigma_u','L_u'),
                                HPD = TRUE,
                                hpd_prob = 0.95,
                                round = 3))

  M1=t(M1)

result_p<-c(M0[1,1],M0[2,1],M0[3,1],M0[4,1],M0[5,1],M1[3,1],M1[4,1],
M0[6,1],M0[7,1],M0[1,2],M0[2,2],M0[3,2],M0[4,2],M0[5,2],M1[3,2],M1[4,2],
M0[6,2],M0[7,2],M0[1,3],M0[2,3],M0[3,3],M0[4,3],M0[5,3],M1[3,3],M1[4,3],
M0[6,3],M0[7,3],M0[1,4],M0[2,4],M0[3,4],M0[4,4],M0[5,4],M1[3,4],M1[4,4],
M0[6,4],M0[7,4],M0[1,5],M0[2,5],M0[3,5],M0[4,5],M0[5,5],M1[3,5],M1[4,5],
M0[6,5],M0[7,5],M0[1,8],M0[2,8],M0[3,8],M0[4,8],M0[5,8],M1[3,8],M1[4,8],
M0[6,8],M0[7,8])

pois.out[ir,]<-result_p

}
#------COMP output--------------------#
cmp.out1<-cbind(cmp.out,cmp.DIC,cmp.loglik,cmp.EP)

write.csv(cmp.out1,
file ="C:/Users/morshed.alam/Desktop
/Simulation_dt_1.5_cont_woi/cmp.out_dt_1.5_cont_woi_1_uni_bc.csv")
```

```
#------NB output--------------------#
NB.out1<-cbind(NB.out,NB.DIC,NB.loglik,NB.EP)

write.csv(NB.out1,
file = "C:/Users/morshed.alam/Desktop
/Simulation_dt_1.5_cont_woi/NB.out_dt_1.5_cont_woi_1_uni_bc.csv")

#------Pois output--------------------#
pois.out1<-cbind(pois.out, pois.DIC, pois.loglik, pois.EP)

write.csv(pois.out1,
file= "C:/Users/morshed.alam/Desktop
/Simulation_dt_1.5_cont_woi/pois.out_dt_1.5_cont_woi_1_uni_bc.csv")

#----------------End-------------------#
```

Bibliography

Abdella, G. M., Kim, J., Al-Khalifa, K. N., & Hamouda, A. M. (2019). Penalized conway-maxwell-poisson regression for modelling dispersed discrete data: The case study of motor vehicle crash frequency. *Safety Science*, *120*, 157–163.

Akaikei, H. (1973). Information theory and an extension of maximum likelihood principle. In *Proc. 2nd int. symp. on information theory* (pp. 267–281).

Allard, J. P., Aghdassi, E., Chau, J., Tam, C., Kovacs, C. M., Salit, I. E., & Walmsley, S. L. (1998). Effects of vitamin e and c supplementation on oxidative stress and viral load in hiv-infected subjects. *Aids*, *12*(13), 1653–1659.

Alvarez, I., Niemi, J., & Simpson, M. (2014). Bayesian inference for a covariance matrix. *arXiv preprint arXiv:1408.4050*.

Balakrishnan, N., Barui, S., & Milienos, F. (2017). Proportional hazards under conway–maxwell-poisson cure rate model and associated inference. *Statistical methods in medical research*, *26*(5), 2055–2077.

Balakrishnan, N., & Feng, T. (2018). Proportional odds under conway-maxwell-poisson cure rate model and associated likelihood inference. *Statistics, Optimization & Information Computing*, *6*(3), 305–334.

Barnard, J., McCulloch, R., & Meng, X.-L. (2000). Modeling covariance matrices in terms of standard deviations and correlations, with application to shrinkage. *Statistica Sinica*, 1281–1311.

Baum, M. K., Campa, A., Lai, S., Martinez, S. S., Tsalaile, L., Burns, P., ... others (2013). Effect of micronutrient supplementation on disease progression in asymptomatic, antiretroviral-naive, hiv-infected adults in botswana: a randomized clinical trial. *Jama*, *310*(20), 2154–2163.

Benguella, L., Arbault, A., Fillion, A., Blot, M., Piroth, C., Denimal, D., ... others (2018). Vitamin d supplementation, bone turnover, and inflammation in hiv-infected patients. *Medecine et maladies infectieuses*, *48*(7), 449–456.

Boatwright, P., Borle, S., & Kadane, J. B. (2003). A model of the joint distribution of purchase quantity and timing. *Journal of the American Statistical Association*, *98*(463), 564–572.

Borle, S., Boatwright, P., & Kadane, J. B. (2006). The timing of bid placement and extent of multiple bidding: An empirical investigation using ebay online auctions. *Statistical Science*, 194–205.

Borle, S., Boatwright, P., Kadane, J. B., Nunes, J. C., & Galit, S. (2005). The effect of product assortment changes on customer retention. *Marketing science*, *24*(4), 616–622.

Brown, J. C., Schall, J. I., Rutstein, R. M., Leonard, M. B., Zemel, B. S., & Stallings, V. A. (2015). The impact of vitamin d3 supplementation on muscle function among hiv-infected children and young adults: a randomized controlled trial. *Journal of musculoskeletal & neuronal interactions*, *15*(2), 145.

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., ... Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of statistical software*, *76*(1).

Chanialidis, C., Evers, L., Neocleous, T., & Nobile, A. (2014). Retrospective sampling in mcmc with an application to com-poisson regression. *Stat*, *3*(1), 273–290.

Chanialidis, C., Evers, L., Neocleous, T., & Nobile, A. (2018). Efficient bayesian inference for com-poisson regression models. *Statistics and computing*, *28*(3), 595–608.

Chen, M.-H., Shao, Q.-M., & Ibrahim, J. G. (2012). *Monte carlo methods in bayesian computation*. Springer Science & Business Media.

Choo-Wosoba, H., & Datta, S. (2018). Analyzing clustered count data with a cluster-specific random effect zero-inflated conway–maxwell–poisson distribution. *Journal of applied statistics*, *45*(5), 799–814.

Choo-Wosoba, H., Gaskins, J., Levy, S., & Datta, S. (2018). A bayesian approach for analyzing zero-inflated clustered count data with dispersion. *Statistics in medicine*, *37*(5), 801–812.

Choo-Wosoba, H., Levy, S. M., & Datta, S. (2016). Marginal regression models for clustered count data based on zero-inflated conway–maxwell–poisson distribution with applications. *Biometrics*, *72*(2), 606–618.

Chu, C., & Selwyn, P. A. (2011). An epidemic in evolution: the need for new models of hiv care in the chronic disease era. *Journal of Urban Health*, *88*(3), 556–566.

Consul, P. C., & Jain, G. C. (1973). A generalization of the poisson distribution. *Technometrics*, *15*(4), 791–799.

Conway, R. W., & Maxwell, W. L. (1962). A queuing model with state dependent service rates. *Journal of Industrial Engineering*, *12*(2), 132–136.

Daly, F., & Gaunt, R. E. (2015). The conway-maxwell-poisson distribution: distributional theory and approximation. *arXiv preprint arXiv:1503.07012*.

Deeks, S. G., Lewin, S. R., & Havlir, D. V. (2013). The end of aids: Hiv infection as a chronic disease. *The Lancet*, *382*(9903), 1525–1533.

del Castillo, J., & Pérez-Casany, M. (2005). Overdispersed and underdispersed poisson generalizations. *Journal of Statistical Planning and Inference*, *134*(2), 486–500.

Eckard, A. R., O'RIORDAN, M. A., Rosebush, J. C., LEE, S. T., Habib, J. G., Ruff, J. H., ... others (2018). Vitamin d supplementation decreases immune activation and exhaustion in hiv-1-infected youth. *Antiviral therapy*, *23*(4), 315.

Erickson, K. L., Medina, E. A., & Hubbard, N. E. (2000). Micronutrients and innate immunity. *The Journal of infectious diseases*, *182*(Supplement_1), S5–S10.

Erkan, G., Türkan, S., & Evkaya, O. (2017). The investigation of the number of babies born alive in multiple pregnancies using quasi poisson model. In *3 rd international researchers, statisticians and young statisticians congress 24-26 may 2017 selçuk university.*

Famoye, F. (1993). Restricted generalized poisson regression model. *Communications in Statistics-Theory and Methods*, *22*(5), 1335–1354.

Famoye, F., Wulu, J. T., & Singh, K. P. (2004). On the generalized poisson regression model with an application to accident data. *Journal of Data Science*, *2*(2004), 287–295.

Fawzi, W. W., Msamanga, G. I., Kupka, R., Spiegelman, D., Villamor, E., Mugusi, F., ... Hunter, D. (2007). Multivitamin supplementation improves hematologic status in hiv-infected women and their children in tanzania. *The American journal of clinical nutrition*, *85*(5), 1335–1343.

Fawzi, W. W., Msamanga, G. I., Spiegelman, D., Wei, R., Kapiga, S., Villamor, E., ... others (2004). A randomized trial of multivitamin supplements and hiv disease progression and mortality. *New England Journal of Medicine*, *351*(1), 23–32.

Fitzmaurice, G., Davidian, M., Verbeke, G., & Molenberghs, G. (2008). *Longitudinal data analysis*. CRC press.

Fitzmaurice, G. M., Laird, N. M., & Ware, J. H. (2012). *Applied longitudinal analysis* (Vol. 998). John Wiley & Sons.

for Disease Control, C., Prevention, et al. (2020). Estimated hiv incidence and prevalence in the united states, 2014–2018. *HIV Surveillance Supplemental Report*, *25*(1), 1–77.

Gaunt, R. E., Iyengar, S., Daalhuis, A. B. O., & Simsek, B. (2019). An asymptotic expansion for the normalizing constant of the conway–maxwell–poisson distribution. *Annals of the Institute of Statistical Mathematics*, *71*(1), 163–180.

Gautam, N., Dayal, R., Agarwal, D., Kumar, R., Singh, T., Hussain, T., & Singh, S. (2014). Role of multivitamins, micronutrients and probiotics supplementation in management of hiv infected children. *The Indian Journal of Pediatrics*, *81*(12), 1315–1320.

Geisser, S., & Eddy, W. F. (1979). A predictive approach to model selection. *Journal of the American Statistical Association*, *74*(365), 153–160.

Gelfand, A. E., Dey, D. K., & Chang, H. (1992). *Model determination using predictive distributions with implementation via sampling-based methods* (Tech. Rep.). STANFORD UNIV CA DEPT OF STATISTICS.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian data analysis*. CRC press.

Gelman, A., et al. (2006). Prior distributions for variance parameters in hierarchical models (comment on article by browne and draper). *Bayesian analysis*, *1*(3), 515–534.

Gelman, A., Rubin, D. B., et al. (1992). Inference from iterative simulation using multiple sequences. *Statistical science*, *7*(4), 457–472.

Guikema, S. D., & Goffelt, J. P. (2008). A flexible count data regression model for risk analysis. *Risk Analysis: An International Journal*, *28*(1), 213–223.

Gupta, R. C., & Huang, J. (2017). The weibull–conway–maxwell–poisson distribution to analyze survival data. *Journal of Computational and Applied Mathematics*, *311*, 171–182.

Guwatudde, D., Ezeamama, A. E., Bagenda, D., Kyeyune, R., Wabwire-Mangen, F., Wamani, H., ... others (2012). Multivitamin supplementation in hiv infected adults initiating antiretroviral therapy in uganda: the protocol for a randomized double blinded placebo controlled efficacy trial. *BMC infectious diseases*, *12*(1), 304.

Guwatudde, D., Wang, M., Ezeamama, A. E., Bagenda, D., Kyeyune, R., Wamani, H., ... Fawzi, W. W. (2015). The effect of standard dose multivitamin supplementation on disease progression in hiv-infected adults initiating haart: a randomized double blind placebo-controlled trial in uganda. *BMC infectious diseases*, *15*(1), 348.

Harezlak, J., Ruppert, D., & Wand, M. P. (2018). *Semiparametric regression with r*. Springer.

Harrell Jr, F. E. (2013). rms: Regression modeling strategies. r package version 4.0-0. *City*.

Harrell Jr, F. E. (2015). *Regression modeling strategies: with applications to linear models, logistic and ordinal regression, and survival analysis*. Springer.

Hastie, T. (2008). gam: Generalized additive models. *R package version*, *1*(2).

Hastie, T. J., & Tibshirani, R. J. (1990). *Generalized additive models* (Vol. 43). CRC press.

He, Z., & Emura, T. (2019). The com-poisson cure rate model for survival data-computational aspects. , *57*(1), 1–42.

Hemsworth, J. C., Hekmat, S., & Reid, G. (2012). Micronutrient supplemented probiotic yogurt for hiv-infected adults taking haart in london, canada. *Gut Microbes*, *3*(5), 414–419.

Herwartz, H., Klein, N., & Strumann, C. (2016). Modelling hospital admission and length of stay by means of generalised count data models. *Journal of Applied Econometrics*, *31*(6), 1159–1182.

Hoffman, M. D., & Gelman, A. (2014). The no-u-turn sampler: adaptively setting path lengths in hamiltonian monte carlo. *J. Mach. Learn. Res.*, *15*(1), 1593–1623.

Huang, A., Wand, M. P., et al. (2013). Simple marginally noninformative prior distributions for covariance matrices. *Bayesian Analysis*, *8*(2), 439–452.

Huang, H.-Y., Caballero, B., Chang, S., Alberg, A., Semba, R., Schneyer, C., ... others (2006). Multivitamin/mineral supplements and prevention of chronic disease. *Evid Rep Technol Assess (Full Rep)*, *139*, 1–117.

Jiamton, S., Pepin, J., Suttent, R., Filteau, S., Mahakkanukrauh, B., Hanshaoworakul, W., ... Jaffar, S. (2003). A randomized trial of the impact of multiple micronutrient supplementation on mortality among hiv-infected individuals living in bangkok. *Aids*, *17*(17), 2461–2469.

Kadane, J. B., Shmueli, G., Minka, T. P., Borle, S., Boatwright, P., et al. (2006). Conjugate analysis of the conway-maxwell-poisson distribution. *Bayesian analysis*, *1*(2), 363–374.

Kaiser, J. D., Campa, A. M., Ondercin, J. P., Leoung, G. S., Pless, R. F., & Baum, M. K. (2006). Micronutrient supplementation increases cd4 count in hiv-infected individuals on highly active antiretroviral therapy: a prospective, double-blinded, placebo-controlled trial. *JAIDS Journal of Acquired Immune Deficiency Syndromes*, *42*(5), 523–528.

Kalyanam, K., Borle, S., & Boatwright, P. (2007). Deconstructing each item's category contribution. *Marketing Science*, *26*(3), 327–341.

Khan, N. M., & Jowaheer, V. (2013). Comparing joint gql estimation and gmm adaptive estimation in com-poisson longitudinal regression model. *Communications in Statistics-Simulation and Computation*, *42*(4), 755–770.

Kongnyuy, E. J., Wiysonge, C. S., & Shey, M. S. (2009). A systematic review of randomized controlled trials of prenatal and postnatal vitamin a supplementation of hiv-infected women. *International Journal of Gynecology & Obstetrics*, *104*(1), 5–8.

Kooperberg, C. (2015). Polspline: Polynomial spline routines. *R package version*, *1*, 12.

Lawrence Gould, A., Boye, M. E., Crowther, M. J., Ibrahim, J. G., Quartey, G., Micallef, S., & Bois, F. Y. (2015). Joint modeling of survival and longitudinal non-survival data: current methods and issues. report of the dia bayesian joint modeling working group. *Statistics in medicine*, *34*(14), 2181–2195.

Leppik, I., Dreifuss, F., Porter, R., Bowman, T., Santilli, N., Jacobs, M., ... others (1987). A controlled study of progabide in partial seizures: methodology and results. *Neurology*, *37*(6), 963–963.

Lewandowski, D., Kurowicka, D., & Joe, H. (2009). Generating random correlation matrices based on vines and extended onion method. *Journal of multivariate analysis*, *100*(9), 1989–2001.

Lord, D., Geedipally, S. R., & Guikema, S. D. (2010). Extension of the application of conway-maxwell-poisson models: Analyzing traffic crash data exhibiting underdispersion. *Risk Analysis: An International Journal*, *30*(8), 1268–1276.

Lord, D., Guikema, S. D., & Geedipally, S. R. (2008). Application of the conway–maxwell–poisson generalized linear model for analyzing motor vehicle crashes. *Accident Analysis & Prevention*, *40*(3), 1123–1134.

Lu, P., & Tolliver, D. (2016). Accident prediction model for public highway-rail grade crossings. *Accident Analysis & Prevention*, *90*, 73–81.

Makinde, O., Rotimi, K., Ikumawoyi, V., Adeyemo, T., & Olayemi, S. (2017). Effect of vitamin a and vitamin c supplementation on oxidative stress in hiv and hiv-tb co-infection at lagos university teaching hospital (luth) nigeria. *African health sciences*, *17*(2), 308–314.

McCullough, P., & Nelder, J. (1989). Generalized linear models chapman and hall. *New York*.

McElreath, R. (2020). *Statistical rethinking: A bayesian course with examples in r and stan*. CRC press.

Mda, S., Raaij, J., De Villiers, F. P., Kok, F. J., et al. (2013). Impact of multi-micronutrient supplementation on growth and morbidity of hiv-infected south african children. *Nutrients*, *5*(10), 4079–4092.

Mda, S., van Raaij, J. M., MacIntyre, U. E., de Villiers, F. P., & Kok, F. J. (2010). Improved appetite after multi-micronutrient supplementation for six months in hiv-infected south african children. *Appetite*, *54*(1), 150–155.

Minka, T. P., Shmueli, G., Kadane, J. B., Borle, S., & Boatwright, P. (2003). Computing with the com-poisson distribution. *PA: Department of*, *776*.

Morris, D., Sellers, K., & Menger, A. (2017). Fitting a flexible model for longitudinal count data using the nlmixed procedure. In *Sas global forum, paper* (pp. 202–2017).

Motswagole, B. S., Mongwaketse, T. C., Mokotedi, M., Kobue-Lekalake, R. I., Bulawayo, B. T., Thomas, T. S., ... Kwape, L. D. (2013). The efficacy of micronutrient-fortified sorghum meal in improving the immune status of hiv-positive adults. *Annals of Nutrition and Metabolism*, *62*(4), 323–330.

Nadarajah, S. (2009). Useful moment and cdf formulations for the com–poisson distribution. *Statistical papers*, *50*(3), 617–622.

Nanayakkara, D. D., Sun, X., Morris, S., Louie, S., Mulligan, K., Overton, T., ... Dubé, M. P. (2019). Effect of vitamin d supplementation on bone turnover markers during hiv pre-exposure prophylaxis using tenofovir disoproxil fumarate-emtricitabine in men who have sex with men. *AIDS research and human retroviruses*, *35*(7), 608–614.

O'Malley, A. J., & Zaslavsky, A. M. (2008). Domain-level covariance analysis for multilevel survey data with structured nonresponse. *Journal of the American Statistical Association*, *103*(484), 1405–1418.

Pal, S., Majakwara, J., & Balakrishnan, N. (2018). An em algorithm for the destructive com-poisson regression cure rate model. *Metrika*, *81*(2), 143–171.

Peluso, A., Vinciotti, V., & Yu, K. (2019). Discrete weibull generalized additive model: an application to count fertility data. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *68*(3), 565–583.

Plummer, M. (2008). Penalized loss functions for bayesian model comparison. *Biostatistics*, *9*(3), 523–539.

PrayGod, G., Range, N., Faurholt-Jepsen, D., Jeremiah, K., Faurholt-Jepsen, M., Aabye, M. G., ... others (2011). Daily multi-micronutrient supplementation during tuberculosis treatment increases weight and grip strength among hiv-uninfected but not hiv-infected patients in mwanza, tanzania. *The Journal of nutrition*, *141*(4), 685–691.

Ridout, M. S., & Besbeas, P. (2004). An empirical model for underdispersed count data. *Statistical Modelling*, *4*(1), 77–89.

Rodrigues, J., de Castro, M., Cancho, V. G., & Balakrishnan, N. (2009). Com–poisson cure rate survival models and an application to a cutaneous melanoma data. *Journal of Statistical Planning and Inference*, *139*(10), 3605–3611.

Ruli, E., Sartori, N., Ventura, L., et al. (2016). Improved laplace approximation for marginal likelihoods. *Electronic Journal of Statistics*, *10*(2), 3986–4009.

Ruppert, D., Wand, M. P., & Carroll, R. J. (2003). *Semiparametric regression* (No. 12). Cambridge university press.

Saffari, S. E., Adnan, R., Hashemi Fard, M., Marchi, V. A., Louzada, F., Carson Allen, J., ... Greene, W. (2018). Modeling doctor visits using a right-censored zero-inflated conway–maxwell poisson regression. *Communications in Statistics: Case Studies, Data Analysis and Applications*, *4*(2), 96–107.

Santarelli, M. F., Della Latta, D., Scipioni, M., Positano, V., & Landini, L. (2016). A conway–maxwell–poisson (cmp) model to address data dispersion on positron emission tomography. *Computers in biology and medicine*, *77*, 90–101.

Sellers, K., Lotze, T., Raim, A., & Raim, M. A. (2019). Package 'compoissonreg'.

Sellers, K. F., Borle, S., & Shmueli, G. (2012). The com-poisson model for count data: a survey of methods and applications. *Applied Stochastic Models in Business and Industry*, *28*(2), 104–116.

Sellers, K. F., Morris, D. S., & Balakrishnan, N. (2016). Bivariate conway–maxwell–poisson distribution: formulation, properties, and inference. *Journal of Multivariate Analysis*, *150*, 152–168.

Sellers, K. F., & Raim, A. (2016). A flexible zero-inflated model to address data dispersion. *Computational Statistics & Data Analysis*, *99*, 68–80.

Sellers, K. F., & Shmueli, G. (2010). A flexible regression model for count data. *The Annals of Applied Statistics*, 943–961.

Sellers, K. F., & Shmueli, G. (2013). Data dispersion: Now you see it... now you don't. *Communications in Statistics-Theory and Methods*, *42*(17), 3134–3147.

Sellers, K. F., Swift, A. W., & Weems, K. S. (2017). A flexible distribution class for count data. *Journal of Statistical Distributions and Applications*, *4*(1), 22.

Semba, R. D., Kumwenda, J., Zijlstra, E., Ricks, M., Van Lettow, M., Whalen, C., ... others (2007). Micronutrient supplements and mortality of hiv-infected adults with pulmonary tb: a controlled clinical trial. *The international journal of tuberculosis and lung disease*, *11*(8), 854–859.

Shmueli, G., Minka, T. P., Kadane, J. B., Borle, S., & Boatwright, P. (2005). A useful distribution for fitting discrete data: revival of the conway–maxwell–poisson distribution. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, *54*(1), 127–142.

Singh, S. S., Borle, S., & Jain, D. C. (2009). A generalized framework for estimating customer lifetime value when customer lifetimes are not observed. *Qme*, *7*(2), 181–205.

Spiegelhalter, D. J., Best, N. G., Carlin, B. P., & Van Der Linde, A. (2002). Bayesian measures of model complexity and fit. *Journal of the royal statistical society: Series b (statistical methodology)*, *64*(4), 583–639.

Stasinopoulos, M., Rigby, B., Voudouris, V., Akantziliotou, C., Enea, M., & Kiose, D. (2017). *gamlss: Generalised additive models for location scale and shape. r package version 5.0.*

Stone, C. J., & Koo, C.-Y. (1985). Additive splines in statistics. *Proceedings of the American Statistical Association. Original pagination is p*, *45*, 48.

Sudjaritruk, A. L. B. T. K. S. C. T. T. S. O. P. S. P., T, & Puthanakit, T. (2017). Effects of calcium and vitamin d supplementation on bone health in hiv-infected youth. *Topics in antiviral medicine*, *25*(1), 348s-349s.

Thall, P. F., & Vail, S. C. (1990). Some covariance models for longitudinal count data with overdispersion. *Biometrics*, 657–671.

Tokuda, T., Goodrich, B., Van Mechelen, I., Gelman, A., & Tuerlinckx, F. (2011). Visualizing distributions of covariance matrices. *Columbia Univ., New York, USA, Tech. Rep*, 18–18.

Van Der Linde, A. (2005). Dic in variable selection. *Statistica Neerlandica*, *59*(1), 45–56.

Van Houwelingen, J., & Le Cessie, S. (1990). Predictive value of statistical models. *Statistics in medicine*, *9*(11), 1303–1325.

Vehtari, A., Gelman, A., & Gabry, J. (2017). Practical bayesian model evaluation using leave-one-out cross-validation and waic. *Statistics and computing*, *27*(5), 1413–1432.

Viard, J.-P., Souberbielle, J.-C., Kirk, O., Reekie, J., Knysz, B., Losso, M., ... others (2011). Vitamin d and clinical disease progression in hiv infection: results from the eurosida study. *Aids*, *25*(10), 1305–1315.

Villamor, E., Msamanga, G., Spiegelman, D., Antelman, G., Peterson, K. E., Hunter, D. J., & Fawzi, W. W. (2002). Effect of multivitamin and vitamin a supplements on weight gain during pregnancy among hiv-1-infected women. *The American journal of clinical nutrition*, *76*(5), 1082–1090.

Wang, Z., Wu, Y., & Chu, H. (2018). On equivalence of the lkj distribution and the restricted wishart distribution. *arXiv preprint arXiv:1809.04746*.

Watanabe, S. (2013). A widely applicable bayesian information criterion. *Journal of Machine Learning Research*, *14*(Mar), 867–897.

Wiangnak, P., & Pal, S. (2018). Gamma lifetimes and associated inference for interval-censored cure rate model with com–poisson competing cause. *Communications in Statistics-Theory and Methods*, *47*(6), 1491–1509.

Wimmer, G., & Altmann, G. (1996). The theory of word length: Some results and generalizations. *Glottometrika*, *15*, 112–133.

Wimmer, G., Köhler, R., Grotjahn, R., & Altmann, G. (1994). Towards a theory of word length distribution. *Journal of Quantitative Linguistics*, *1*(1), 98–106.

Winkelmann, R., & Zimmermann, K. F. (1995). Recent developments in count data modelling: theory and application. *Journal of economic surveys*, *9*(1), 1–24.

Wood, S., & Wood, M. S. (2015). Package 'mgcv'. *R package version*, *1*, 29.

Wu, G., Holan, S. H., & Wikle, C. K. (2013). Hierarchical bayesian spatio-temporal conway–maxwell poisson models with dynamic dispersion. *Journal of Agricultural, Biological, and Environmental Statistics*, *18*(3), 335–356.

Yee, T. W., Yee, M. T., & VGAMdata, S. (2020). Package 'vgam'.

Zhang, D., Chen, M.-H., Ibrahim, J. G., Boye, M. E., & Shen, W. (2017). Bayesian model assessment in joint modeling of longitudinal and survival data with applications to cancer clinical trials. *Journal of Computational and Graphical Statistics*, *26*(1), 121–133.

Zhao, F., Wang, Z., & Li, W. (2009). Impact of micronutrients on immunocyte and microelements status among people living with hiv. *Zhonghua liu Xing Bing xue za zhi= Zhonghua Liuxingbingxue Zazhi*, *30*(5), 452–454.