Unilag Journal of Mathematics and Applications,
Volume 1, Issue 1 (2021), Pages 60--81.
ISSN: ***-****. URL: http://lagjma.edu.ng

CONJUGATION OF GENERALIZED GAMMA PRIOR WITH POISSON AND GENERALIZED POISSON LIKELIHOODS FOR DISEASE MAPPING

UGOCHUKWU MBATA*, ISMAIL ADELEKE AND KAYODE ADEKEYE

ABSTRACT. This article focused on the use of generalized Gamma distribution as conjugate prior with Poisson and generalized Poisson likelihoods to handle dispersion in small samples. Based on this conjugacy, Poisson-Generalized Gamma model (PGG) and Generalized Poisson-Generalized Gamma model (GPGG) are developed for Bayesian disease mapping and compared with the existing Poisson-Gamma model. The efficiency of these models was investigated using both simulated and real data applications. The deviance information criterion (DIC), dispersion test (DT), Monte Carlo error (MCE) and relative efficiency (reff) were used for comparison. All indicated that GPGG model provided the best precision and model efficiency to handle dispersion and relative risk estimation for disease mapping in small and large samples under uncontaminated and contaminated data. Thus, GPGG and PGG models served as alternative models in providing reliable mapping of disease.

1. INTRODUCTION

Disease mapping was derived from Clayton and Kaldor (1987), and defined as the investigation, estimation and visual representation of summary measures of health incidences across related regions. Disease mapping is mainly used for explanatory purposes, to survey high-risk areas and to help policy making and implementation (Koch, 2005). As discussed in Meza (2003), application of Empirical Bayes (EB) methods to improve relative risk estimation efficiency is becoming popular since the frequentist approaches are not fully satisfactory. Casella (1985) and Efron (1996) provided good introductions to EB methods. Adeleke *et al.* (2009), Mbata *et al.* (2010), Okafor *et al.* (2010), Okafor and Mbata (2012), Zou *et al.* (2018) provided good applications of EB methods. The hallmark of EB analysis is that it has the capacity to remove the random variability which is present in data from small population counts (Böhning *et al.*, 2000) as well as the capacity to combine independent but related studies.

2010 *Mathematics Subject Classification*. Primary: 22E30. Secondary: 58J05.
Key words and phrases. Disease Mapping, Dispersion, Empirical Bayes, Generalized Gamma, Relative Risk

Submitted: March 7, 2021. Revised: May 4, 2021. Accepted: May 11, 2021. *Correspondence

1.1. Literature Review. EB concept was first introduced by Robbins (1955), which employed the Thomas Bayes theory of 1740, in a non-parametric setting. But subsequent research developments by Morris (1983) have introduced parametric models; which the present study is based-on. Bayes techniques for mortality rates were introduced by Manton *et al.* (1981). However, a well-known EB model for disease mapping was introduced by Clayton and Kaldor (1987). Other EB models for disease mapping included the work of Azzalini (1985) which introduced Poisson-Normal (PN) model. As discussed in Militino *et al.* (2001), PN model is not a natural conjugate model. Clayton and Kaldor (1987) introduced Poisson-Lognormal model, which has potential for multicollinearity problem (Lawson, 2013).The Poisson-conditional autoregressive model (P-CAR) was first introduced by Besag (1974) and the Bayes disease-mapping models based on P-CAR was studied by Clayton and Kaldor (1987, 1989), Cressie and Chan (1989), Besag *et al.* (1991). Modifications of P-CAR model were carried out by Besag *et al.* (1991) popular called the BYM models. However, the P-CAR models have potential for multicollinearity problem (Lawson, 2013).

Meanwhile, the use of Poisson-Gamma (PG) model, as suggested by Clayton and Kaldor (1987), is the widely used EB model in disease mapping. Other notable works that used PG model in disease mapping include: Lord (2006), Lawson (2013), Clement (2014), Srinivasan and Venkatesan (2014), Mbata *et al.* (2018). According to Lord (2006), for point data estimation, PG model is usually preferred over other models because it offers a simple way of accommodating over-dispersion which usually features in disease mapping. But one major problem of PG modelas discussed in Lord and Geedipally (2016) and Famoye *et al.* (2011) is theincapability to address under-dispersion, and to handle over-dispersion in small samples. To resolve the problems inherent in the use of PG model, the present article is proposing using generalized Gamma (GG) distribution as prior with Poisson and generalized Poisson (GP) likelihoods. The efficiency of these models in handling dispersion is investigated under contaminated data for small and large samples. Conjugation of Generalized Gamma distribution with Poisson likelihood has been carried out by Mbata *et al.* (2018). However, the conjugation of Generalized Gamma distribution with Generalized Poisson likelihood has not been investigated in the literature, hence the motivation for this study. The study will deepen knowledge in Bayesian disease mapping for credible health risk assessment and management, and aid practitioners in epidemiology and Bio-medical fields.

The PGG model and the proposed GPG G model are based on the assumption that incidence of disease is being investigated in a population that is partitioned into K subpopulations or regions and each subpopulation (region) observed count assumes to follow a Poisson or Generalized Poisson (GP) distribution with unknown relative risk parameter which is assigned a GG distribution as prior. Hence, inference about the unknown parameter (relative risks) is generated empirically using the derived posterior distribution. A detailed Bayes theorem is found in Box and Tiao (1973).

2. MATERIALS AND METHODS

2.1 EB Modeling of Generalized Gamma (GG) Distribution with Generalized Poisson (GP) Likelihood

Definition 2.1. For EB disease mapping models: let $\boldsymbol{Y_i}$beobserved number of cases of disease in region $i\ (i = 1,\ \ldots, K)$ **,** let $\boldsymbol{E_i}$ beexpected number of cases of disease in region $i\ (i = 1,\ \ldots, K)$**,** let$\boldsymbol{N_i}$benumber of persons at risk for disease in region $i\ (i = 1,\ \ldots, K)$**,** let $\boldsymbol{\theta_i}$ bemaximum likelihood estimate of relative risk of disease in region $i\ (i = 1,\ \ldots, K)$**,** and let$\widetilde{\boldsymbol{\theta}}_{\boldsymbol{i}}$bePosterior estimatesof relative risk of disease in region $i\ (i = 1,\ \ldots, K)$.$\boldsymbol{Y_i}$ are observed random variables in region i, while $\boldsymbol{E_i}$ are known functions of $\boldsymbol{N_i}$**.** Hence, $\boldsymbol{E_i} = \boldsymbol{N_i}\bar{\boldsymbol{r}} = \boldsymbol{N_i}\left(\frac{\sum_{i=1}^{k}\boldsymbol{Y_i}}{\sum_{i=1}^{k}\boldsymbol{N_i}}\right)$ while estimated$\boldsymbol{\theta_i}$ is $\boldsymbol{\theta}_i = \frac{\boldsymbol{Y_i}}{\boldsymbol{E_i}}(thus\ \boldsymbol{Y}_i = \boldsymbol{E_i\theta_i})$. $\bar{\boldsymbol{r}}$is the totalrisk of disease in the whole study region (called internal standardization (IS)).

Based on Bayes' Theorem, Generalized Poisson-Generalized Gamma (GPGG) model is presented for disease mapping.

Definition 2.2. Let

$Y_i|\theta_i, \omega \sim GP(E_i\theta_i, \omega)$, and

$$\theta_i|\alpha, \beta, \lambda \sim GG(\alpha, \beta, \lambda).$$

Therefore, the posterior distribution is described as

$$p\left(\tilde{\theta}_i\middle|Y_i, \alpha, \beta, \omega, \lambda\right) = \frac{l(\theta_i|Y_i, \omega)\, p(\theta_i|\alpha, \beta, \lambda)}{\int_0^\infty l(\theta_i|Y_i, \omega)\, p(\theta_i|\alpha, \beta, \lambda) d\theta} \propto l(\theta_i|Y_i, \omega)\, p(\theta_i|\alpha, \beta, \lambda). \quad (2.1)$$

Where; the likelihood function is obtained from the pmf of GP distribution presented by Consul and Jain (1973) further studied by Famoye (2010). The re-parameterization with rate parameter $\eta_i = \epsilon_i\theta$, gives

$$p(Y_i|\theta_i, \omega) = \frac{\epsilon_i\theta(\epsilon_i\theta + \omega y)^{y-1}}{y!}e^{-(\epsilon_i\theta + \omega y)},\ y \geq 0,\ \theta > 0.\ \max\left(-1, -\frac{\theta}{4}\right) \leq \omega \leq 1.$$

The GP likelihood ignoring factors that are free of θ is jointly given as

$$l(\theta_i|Y_i, \omega) = \prod_{i=1}^{n}\frac{\epsilon_i\theta}{y_i!}(\epsilon_i\theta + \omega y_i)^{y_i-1}e^{-(\epsilon_i\theta + \omega y_i)} \propto \theta^n e^{-(\theta\sum\epsilon_i + \omega\sum y_i)}\prod_{i=1}^{n}(\epsilon_i\theta + \omega y_i)^{y_i-1}$$

$$l(\theta_i|Y_i, \omega) \propto \theta^n e^{-(E\theta + \omega Y)}\prod_{i=1}^{n}(\epsilon_i\theta + \omega y_i)^{y_i-1}$$

$since\ \prod_{i=1}^{n}(\epsilon_i + \beta y_i)^{y_i-1}$is free of θ and using the transformation $\omega = \beta\theta$, thus

$l(\theta_i|Y_i, \omega) \propto \theta^Y e^{-(E+\beta Y)\theta}$, (2.2)

where, $Y = \sum_{i=1}^{n} y_i$ is the sufficient statistic, and $E = \sum_{i=1}^{n}\epsilon_i$ is the random effect.

The prior distribution is the Generalized Gamma distribution (GG) given by Stacy (1962) as

$$p(\theta_i|\alpha,\beta,\lambda) = \frac{\lambda\beta^{\alpha\lambda}}{\Gamma(\alpha)}\theta^{\alpha\lambda-1}e^{-(\beta\theta)^{\lambda}},\ \alpha,\beta,\lambda,\theta > 0. \quad (2.3)$$

64 U. MBATA, I. ADELEKE AND K. ADEKEYE

Whereα and λ are the shape parameters and β is the inverse scale parameter.

Posterior involves plugging likelihood and prior; hence, the joint posterior resulted to

$$p(\tilde{\theta}_i|Y_i,\alpha,\beta,\omega,\lambda) \propto \left[\theta^{Y}e^{-(E+\beta Y)\theta}\right]\left[\frac{\lambda\beta^{\alpha\lambda}}{\Gamma(\alpha)}\theta^{\alpha\lambda-1}e^{-(\beta\theta)^{\lambda}}\right]$$

$$= \frac{\lambda\beta^{\alpha\lambda}}{\Gamma(\alpha)}\theta^{Y+\alpha\lambda-1}e^{-\left((E+\beta Y)\theta+(\beta\theta)^{\lambda}\right)}$$

By applying Binomial transformation for the$e^{-(E+\beta Y)\theta}$ to factor out θ, GPGG model is derived as

$$p(\tilde{\theta}_i|Y_i,\alpha,\beta,\omega\lambda) = \frac{\lambda\beta^{\alpha\lambda}}{\Gamma(\alpha)}e^{-(\beta\theta)^{\lambda}}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\theta^{Y+\alpha\lambda+i-1},\ \alpha,\beta,\lambda,\theta > 0 \quad (2.4)$$

To obtain a proper posterior distribution of GPGG model, Equation (2.4) is multiplied with the constant of proportionality(c).The constant of proportionality (c) is the inverse of the marginal distribution (denominator) in Equation (2.1). So, the marginal distribution is derived as

$$\int_0^{\infty} l(\theta_i|Y_i,\omega)\,p(\theta_i|\alpha,\beta,\lambda)d\theta = \int_0^{\infty}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\frac{\lambda\beta^{\alpha\lambda}}{\Gamma(\alpha)}\theta^{Y+\alpha\lambda+i-1}e^{-(\beta\theta)^{\lambda}}\,d\theta$$

$$= \frac{\lambda\beta^{\alpha\lambda}}{\Gamma(\alpha)}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\int_0^{\infty}\theta^{Y+\alpha\lambda+i-1}e^{-(\beta\theta)^{\lambda}}d\theta$$

Integrating by substitution: let $W = (\beta\theta)^{\lambda} \implies \theta = \frac{W^{\frac{1}{\lambda}}}{\beta}. Hence\ d\theta = \frac{W^{\frac{1}{\lambda}-1}}{\lambda\beta}dW$, therefore

$$l(\theta_i|Y_i,\omega)\,p(\theta_i|\alpha,\beta,\lambda)d\theta$$

$$= \frac{\lambda\beta^{\alpha\lambda}}{\Gamma(\alpha)}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\int_0^{\infty}\left(\frac{W^{\frac{1}{\lambda}}}{\beta}\right)^{Y+\alpha\lambda+i-1}e^{-W}\left(\frac{W^{\frac{1}{\lambda}-1}}{\lambda\beta}\right)dW.$$

Therefore, the marginal distribution is described as

$$M(Y_i|\alpha,\beta,\omega,\lambda) = \frac{1}{\Gamma(\alpha)}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\frac{1}{\beta^{Y+i}}\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right). \quad (2.5)$$

Hence, the constant of proportionality is derived as

$$c = \frac{1}{\frac{1}{\Gamma(\alpha)}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\frac{1}{\beta^{Y+i}}\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)} = \frac{\Gamma(\alpha)}{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\frac{1}{\beta^{Y+i}}\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)} \quad (2.6)$$

Equations (2.4) and (2.6) are multiplied to obtain a proper posterior distribution of GPGG model derived as

$$p\left(\tilde{\theta}_i\middle|Y_i.\,\alpha,\beta,\omega,\lambda\right)=\frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\beta^i}{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\,\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)}\lambda\beta^{Y+\alpha\lambda}\theta^{Y+\alpha\lambda+i-1}e^{-(\beta\theta)^{\lambda}},$$

$\alpha,\beta,\lambda,\theta>0$■ (2.7)

To proof that the obtained expression in Equation (2.7) is a pdf, we have Theorem 2.1.

Theorem 2.1: Let$\tilde{\theta}_i\sim GPGG\ model$, as in Equation (2.7), then

$$\int_0^{\infty}p\left(\tilde{\theta}_i\middle|Y_i,\alpha,\beta,\omega,\lambda\right)d\theta$$

$$=\int_0^{\infty}\frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\beta^i}{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\,\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)}\lambda\beta^{Y+\alpha\lambda}\theta^{Y+\alpha\lambda+i-1}e^{-(\beta\theta)^{\lambda}}d\theta=1$$

Proof: GPGG model integrates to unity

Given

$$p\left(\tilde{\theta}_i\middle|Y_i,\alpha,\beta,\omega,\lambda\right)=\frac{l(\theta_i|Y_i,\omega)p(\theta_i|\alpha,\beta,\lambda)}{\int_0^{\infty}l(\theta_i|Y_i,\omega)p(\theta_i|\alpha,\beta,\lambda)d\theta}=$$

$$\frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\beta^i}{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)}\lambda\beta^{Y+\alpha\lambda}\theta^{Y+\alpha\lambda+i-1}e^{-(\beta\theta)^{\lambda}}.$$

Considering the probability axiom$\int_{-\infty}^{\infty}p(x)dx=1$, therefore,

$$\int_0^{\infty}p\left(\tilde{\theta}_i\middle|Y_i;\ \alpha,\beta,\omega,\lambda\right)d\theta$$

$$=\frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\beta^i}{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\,\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)}\int_0^{\infty}\lambda\beta^{Y+\alpha\lambda}\theta^{Y+\alpha\lambda+i-1}e^{-(\beta\theta)^{\lambda}}d\theta=1$$

Integrating by substitution: we have

$$=\frac{\lambda\beta^{Y+\alpha\lambda}\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\beta^i}{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\,\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)}\int_0^{\infty}\left(\frac{W^{\frac{1}{\lambda}}}{\beta}\right)^{Y+\alpha\lambda+i-1}e^{-\mathrm{W}}\left(\frac{W^{\frac{1}{\lambda}-1}}{\lambda\beta}\right)dW$$

$$=\frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i}{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\,\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)}\int_0^{\infty}W^{\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}-1}e^{-\mathrm{W}}dW$$

$$=\frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\,\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)}{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\,\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)}=1\blacksquare$$

Hence, the pdf of the GPGG distribution is a proper distribution.

Theorem 2.2: The rth Moment of GPGG Model is derived as

$$E\left(\tilde{\theta}^r\right) = \frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{r}{\lambda}+\frac{i}{\lambda}\right)}{\beta^r\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)} \tag{2.8}$$

Proof: Using the method of moment (MOM),

$$E\left(\tilde{\theta}^r\right) = \frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\,\lambda\beta^{Y+\alpha\lambda+i}}{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\,\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)}\int_0^{\infty}\theta^{Y+\alpha\lambda+r+i-1}e^{-(\beta\theta)^{\lambda}}d\theta.$$

Applying integration by substitution, we have

$$E\left(\tilde{\theta}^r\right) = \frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\,\lambda\beta^{Y+\alpha\lambda+i}}{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\,\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)}\int_0^{\infty}\left(\frac{W^{\frac{1}{\lambda}}}{\beta}\right)^{Y+\alpha\lambda+r+i-1}e^{-W}\left(\frac{W^{\frac{1}{\lambda}-1}}{\lambda\beta}\right)dW$$

$$E\left(\tilde{\theta}^r\right) = \frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{r}{\lambda}+\frac{i}{\lambda}\right)}{\beta^r\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)}\blacksquare \qquad \text{(The rth Moment of GPGG)}$$

The first moment and second moment are used to obtain the posterior expectation and the posterior variance respectively.As discussed in Mbata *et al.* (2018),the hyperparameters α and β of the prior are obtained from GG model,to completely specify the posterior distribution. Therefore, when $\lambda = 0.5$, as

$$\hat{\beta} = \frac{\alpha\,(\alpha+1)}{\mu}. \tag{2.9}$$

$$\hat{\alpha} = \left[6\left(\frac{\mu^2}{\sigma^2}-\frac{1}{K}\right)+\left(\frac{1}{2}+\frac{2}{K}-\frac{2\mu^2}{\sigma^2}\right)^2\right]^{\frac{1}{2}}-\left(\frac{1}{2}+\frac{2}{K}-\frac{2\mu^2}{\sigma^2}\right). \tag{2.10}$$

According to Marshall (1991), μ and $\sigma^2$are estimated as

$\hat{\mu} = \frac{\sum_{i=1}^{k}\theta_i E_i}{\sum_{i=1}^{k}E_i}$and $\hat{\sigma}^2 = S^2 - \frac{\hat{\mu}}{\frac{1}{K}\sum_{I=1}^{k}E_i}$, where $S^2 = \frac{\sum_{i=1}^{k}E_i\left(\hat{\theta}_i-\hat{\mu}\right)^2}{\sum_{i=1}^{k}E_i}$.

2.2Estimation of Relative Risk (RR) and Variance from GPGG Model

From Equation (2.8), it follows that the estimation of relative risk (RR) in each Sub-population is obtained as

$$\hat{\tilde{\theta}}_i^{GPGG} = \frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{1}{\lambda}+\frac{i}{\lambda}\right)}{\beta\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)} \tag{2.11}$$

and the estimation of variance of relative risk (RR) in each sub-population is obtained as

$$Var\left(\hat{\tilde{\theta}}_i^{GPGG}\right) = \frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{2}{\lambda}+\frac{i}{\lambda}\right)}{\beta^2\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)} - \left(\frac{\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{1}{\lambda}+\frac{i}{\lambda}\right)}{\beta\sum_{i=0}^{\infty}\frac{(-1)^i}{i!}(E+\beta Y)^i\Gamma\left(\alpha+\frac{Y}{\lambda}+\frac{i}{\lambda}\right)}\right)^2 \tag{2.12}$$

$$SD\left(\hat{\hat{\theta}}_i^{GPGG}\right) = \sqrt{Var\left(\hat{\hat{\theta}}_i^{GPGG}\right)} \quad (2.13)$$

2.3Performance Indicators for Model Comparison

Beside the variance and standard deviations, the following indicators are employed for model comparison: Deviance Information Criterion (DIC), Dispersion Test with Adjusted Deviance (DT), Monte Carlo error (MCE) and Relative Efficiency (REFF). A good description of the performance indicators can be seen in Carlin and Louis (2009), Brooks and Gelman (1998), and Miller and Miller (2004).

3.RESULTS

The results of simulated and real data for GPGG model for disease mapping are presented in this section. The obtained results were compared to PGG and PG model to investigate the precision and efficiency of the EB models in handling dispersion and in relative risk parameter estimation for disease mapping under uncontaminated and contaminated data.

3.1 Simulation Study Results

The examination of dispersion effect on the efficiency of the different EB models in the estimation of relative risk parameter under uncontaminated data (UD) and contaminated data (CD) for small and large samples is presented using simulated data. This is carried out through a simulation study of 100000 samples using MCMC sampling technique by OpenBUGS statistical software. In order to consider variation, the efficiency of the EB models is studied in different simulated sample sizes Y = 5, 10, 20, 50, 100, 200.The simulation results for prior estimates are displayed in Tables 3.1 and3.2 while Posterior estimates are presented in Tables3.3 and3.4.The relative efficiency is presented in Table 3.5. The corresponding diagnostic plots are presented in Figures 3.1 respectively.

Table 3.1. Prior Estimates of Simulated Uncontaminated Data

				Gamma Prior		Generalized Gamma Prior			
Characteristics	**Sample Size (Y)**	μ	σ^2	α	β	α	β	λ	**Nature of Data**
Small Size Data	5	5.00	16.0000	1.3625	0.2725	5.8479	8.0092	0.5	Over-Dispersed
	10	10.40	10.2666	10.4350	1.0033	42.2231	175.4821	0.5	Equi-Dispersed
	20	20.10	7.2526	55.6553	2.7689	223.1179	2487.7960	0.5	Under-Dispersed
Large Size Data	50	50.18	65.4159	38.4726	0.7666	154.3858	478.0662	0.5	Over-Dispersed
	100	100.16	100.1358	100.1742	1.0001	401.1951	1611.0100	0.5	Equi-Dispersed
	200	200.35	167.2337	240.0192	1.1979	960.5759	4610.2650	0.5	Under-Dispersed

Table 3.2. Prior Estimates of Simulated Contaminated Data

				Gamma Prior		Generalized Gamma Prior			
Characteristics	**Sample Size (Y)**	μ	σ^2	α	β	α	β	λ	**Nature of Data**
Small Size Data	5	5.33	10.3333	2.5526	0.4786	10.6490	23.2594	0.5	Over-Dispersed
	10	10.40	10.3000	10.4009	1.0001	42.0866	174.3630	0.5	Equi-Dispersed
	20	20.16	7.9696	50.9801	2.5279	204.4168	2082.1800	0.5	Under-Dispersed
Large Size Data	50	50.32	87.0455	29.0751	0.5777	116.7944	273.3776	0.5	Over-Dispersed
	100	100.28	100.1816	100.3703	1.0008	401.9795	1615.3570	0.5	Equi-Dispersed
	200	200.02	156.3624	255.8821	1.2792	1024.0280	5247.5550	0.5	Under-Dispersed

The hyperparameters estimatesfor Gamma and Generalized Gamma priorsfrom the simulated uncontaminated data are presentedin Table 3.1 while contaminated data are presented in Table 3.2. The simulation is based on different sample sizes based on the work of Lord (2006) and $\boldsymbol{\lambda}$ fixed at 0.5. Sample sizes (5) and (50) are over dispersed, (10) and (100) are equi-dispersed while (20) and (200) are under dispersed.The prior estimates constitute the initial values for the MCMC sampling to generate the posterior estimates. The posterior results for uncontaminated data are presented in Table 3.3 while contaminated data are displayed in Table 3.4.

Table 3.3. Posterior Estimates of Simulated Uncontaminated Data

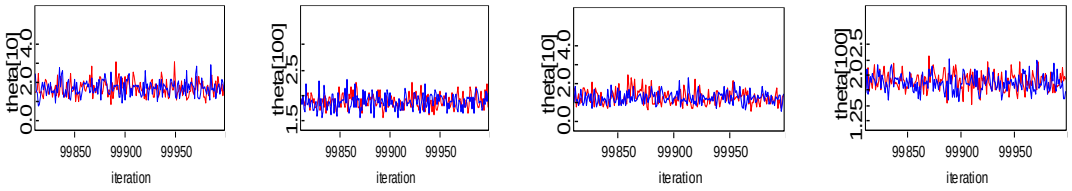
EB MODELS					**EB MODELS**				
Small Sample Size		**PG**	**PGG**	**GPGG**	**Large Sample Size**		**PG**	**PGG**	**GPGG**
5	$\widehat{\widehat{\boldsymbol{\theta}}}$	1.2067	1.3338	1.0142	**50**	$\widehat{\widehat{\boldsymbol{\theta}}}$	1.8134	2.1935	1.2077
	SD	0.7585	0.7414	0.4029		**SD**	0.1489	0.1436	0.0034
	DIC	27.03	26.07	16.09		**DIC**	1474.00	1469.00	1431.94
	DT	58.7912	26.8384	18.2796		**DT**	64.9753	62.9267	52.0265
10	$\widehat{\widehat{\boldsymbol{\theta}}}$	1.7941	2.0270	1.1515	**100**	$\widehat{\widehat{\boldsymbol{\theta}}}$	1.9371	2.3802	1.2366
	SD	0.2856	0.2378	0.0189		**SD**	0.1019	0.0997	0.0090
	DIC	119.20	117.20	109.03		**DIC**	6236.00	6230.00	6208.45
	DT	22.5842	21.2059	20.8800		**DT**	129.5831	127.4570	115.1463
20	$\widehat{\widehat{\boldsymbol{\theta}}}$	1.8154	2.0968	1.0867	**200**	$\widehat{\widehat{\boldsymbol{\theta}}}$	1.9692	2.4384	1.2443
	SD	0.1409	0.1145	0.0520		**SD**	0.0716	0.0708	0.0074
	DIC	633.00	568.90	418.00		**DIC**	24750.00	24590.00	23880.00
	DT	47.3289	41.2410	31.1526		**DT**	252.7179	248.4407	202.8271

Table 3.4. Posterior Estimates of Simulated Contaminated Data

EB MODELS					**EB MODELS**				
Small Sample Size		**PG**	**PGG**	**GPGG**	**Large Sample Size**		**PG**	**PGG**	**GPGG**
5	$\widehat{\widehat{\boldsymbol{\theta}}}$	3.8938	3.1116	1.1152	**50**	$\widehat{\widehat{\boldsymbol{\theta}}}$	11.4262	4.5778	3.2558
	SD	2.8832	1.5717	0.3283		**SD**	19.2174	4.5433	0.1715
	DIC	17.26	16.56	13.33		**DIC**	1367.00	1347.00	1277.50
	DT	7.8189	7.3125	5.7184		**DT**	46.7686	43.3815	38.7420
10	$\widehat{\widehat{\boldsymbol{\theta}}}$	5.1548	3.3608	1.8630	**100**	$\widehat{\widehat{\boldsymbol{\theta}}}$	12.7377	4.5271	3.2543
	SD	4.0601	1.9676	0.1979		**SD**	30.8777	6.1369	0.1547
	DIC	34.29	34.16	31.93		**DIC**	5577.00	5493.00	5295.00
	DT	5.3622	5.1225	4.9736		**DT**	103.3825	100.3256	84.6059
20	$\widehat{\widehat{\boldsymbol{\theta}}}$	6.6669	3.3235	3.0491	**200**	$\widehat{\widehat{\boldsymbol{\theta}}}$	21.8779	5.6626	4.2616
	SD	6.5674	2.3172	0.1378		**SD**	59.9140	9.7534	0.1560
	DIC	273.10	244.20	153.75		**DIC**	21780.00	21730.00	21412.50
	DT	18.5591	16.7933	10.4923		**DT**	202.7292	200.0370	170.9675

Table 3.5. Relative Efficiency (REFF) of Posterior Estimates of Standard Deviations

			Uncontaminated Data			Contaminated Data		
Model	**Sample Size**		**PG**	**PGG**	**GPGG**	**PG**	**PGG**	**GPGG**
PG	**Small samples**	5	1			1		
		10	1			1		
		20	1			1		
	Large Samples	50	1			1		
		100	1			1		
		200	1			1		
PGG	**Small samples**	5	0.9775	1		0.5451	1	
		10	0.8326	1		0.4846	1	
		20	0.8126	1		0.3528	1	
	Large Samples	50	0.9644	1		0.2364	1	
		100	0.9784	1		0.1987	1	
		200	0.9888	1		0.1628	1	
GPGG	**Small samples**	5	0.5312	0.5434	1	0.1139	0.2089	1
		10	0.0662	0.0795	1	0.0487	0.1006	1
		20	0.3691	0.4541	1	0.0210	0.0595	1
	Large Samples	50	0.0228	0.0237	1	0.0089	0.0377	1
		100	0.0883	0.0903	1	0.0050	0.0252	1
		200	0.1034	0.1045	1	0.0026	0.0160	1



PG Model under Uncontaminated Data **PG Model under Contaminated Data**

PGG Model under Uncontaminated Data **PGG Model under Contaminated Data**

GPGG Model under Uncontaminated Data **GPGG Model under Contaminated Data**

Figure 3.1.Dynamic Trace Plot of Posterior Convergence of the Different EB Models for Small and Large Sample Sizes under uncontaminated data and contaminated data

3.2 Real-Life Application Study Results

This section presents the application to real-life data to investigate model efficiency of the different EB models in relative risk estimation for disease mapping under different data conditions such as uncontaminated data (UD) and contaminated data (CD). This is achieved by using people living with HIV (Human Immunodeficiency Virus) cases by State in Nigeria(2014) as uncontaminated data and Malaria cases by State in Nigeria (2014) as contaminated data due to missingness.The Nigeria national population census figures areused as standardization. The data are available with the National Bureau of Statistics (2016 and 2012) respectively and, the expected counts are available with the authors.The investigation is carried out through a simulation study of 100000 samples using MCMC sampling technique by OpenBUGS statistical software. The EB models results are presented in Tables 3.6 and 3.7, and the corresponding disease maps are presented inFigures 3.2, 3.3, 3.4,3.6,3.7 and 3.8. For further investigation, the comparative line graphs of the standard deviations are depicted in Figures3.5 and 3.9.

Table 3.6 Posterior Relative Risk Estimates and SD of Different EB Models by State under Uncontaminated Data using HIV Cases by State in Nigeria (2014)

	PG Model		**PGG Model**		**GPGG Model**	
State	$\widehat{\theta}_i^{PG}$	**SD**	$\widehat{\theta}_i^{PGG}$	**SD**	$\widehat{\theta}_i^{GPGG}$	**SD**
Abia	1.2430*	0.004356	1.2430*	0.004346	0.9138	0.001152
Adamawa	1.0260*	0.003749	1.0260*	0.003718	0.7143	0.001135
AkwaIbom	2.7030*	0.005454	2.7030*	0.005438	2.3188*	0.000776
Anambra	0.5470	0.002377	0.5470	0.002376	0.3008	0.001070
Bauchi	0.3227	0.001729	0.3228	0.001724	0.1351	0.001003
Bayelsa	0.7821	0.004446	0.7821	0.004449	0.4973	0.001618
Benue	2.0080*	0.004526	2.0080*	0.004533	1.6419*	0.000828
Borno	0.7840	0.002854	0.7840	0.002849	0.4990	0.001042
Cross River	1.2490*	0.004322	1.2490*	0.004335	0.9194	0.001141
Delta	0.3657	0.001959	0.3657	0.001964	0.1644	0.001084
Ebonyi	0.5791	0.003389	0.5791	0.003380	0.3266	0.001480
Edo	0.3061	0.002022	0.3061	0.002028	0.1243	0.001205
Ekiti	0.3255	0.002428	0.3255	0.002426	0.1370	0.001406
Enugu	0.4542	0.002456	0.4542	0.002453	0.2287	0.001221
Gombe	0.9056	0.004077	0.9056	0.004074	0.6060	0.001348
Imo	1.0290*	0.003362	1.0290*	0.003372	0.7170	0.001017
Jigawa	0.7053	0.002650	0.7053	0.002645	0.4314	0.001031
Kaduna	2.4510*	0.004176	2.4510*	0.004179	2.0724*	0.000644
Kano	0.7615	0.001871	0.7615	0.001872	0.4795	0.000694
Katsina	0.3936	0.001715	0.3936	0.001713	0.1841	0.000912
Kebbi	0.2257	0.001732	0.2257	0.001731	0.0756	0.001151
Kogi	0.8507	0.003350	0.8507	0.003357	0.5573	0.001157
Kwara	0.6648	0.003467	0.6648	0.003480	0.3972	0.001399
Lagos	0.9427	0.002123	0.9427	0.002127	0.6391	0.000684
Nassarawa	2.4600*	0.007558	2.4600*	0.007531	2.0812*	0.001160
Niger	0.8240	0.003010	0.8240	0.002999	0.5339	0.001060
Ogun	0.8036	0.003051	0.8036	0.003050	0.5160	0.001090
Ondo	0.6183	0.002789	0.6183	0.002787	0.3586	0.001170
Osun	0.6140	0.002790	0.6140	0.002784	0.3551	0.001174
Oyo	1.3350*	0.003209	1.3350*	0.003211	0.9997	0.000808
Plateau	1.1100*	0.003894	1.1100*	0.003890	0.7910	0.001117
Rivers	1.0100*	0.002892	1.0100*	0.002898	0.6997	0.000893
Sokoto	1.3200*	0.003921	1.3200*	0.003914	0.9856	0.000996
Taraba	2.0150*	0.006160	2.0150*	0.006161	1.6486*	0.001116
Yobe	1.1140*	0.004552	1.1140*	0.004542	0.7946	0.001306
Zamfara	0.2474	0.001813	0.2474	0.001812	0.0880	0.001163
Abuja	3.4330*	0.010280	3.4330*	0.010275	3.0369*	0.001186
Nigeria	**1.0000**		**1.0000**		**1.0000**	
DIC	553.31		553.31		139.70	
DT	-511.19		-511.19		-138.32	
Dbar	516.30		516.30		139.70	
P^D	37.01		37.01		37.01	
MCE	0.000007892		0.000005550		0.000001725	
5% SD	0.000176432		0.000176324		0.000054703	

Note: **Asterisk implies Relative Risk ≥ 1 (High Risk Area). Non-asterisk implies Relative Risk < 1 (Low Risk Area). SD (Standard Deviation). MCE (Monte Carlo error), DT (Dispersion Test), DIC (Deviance Information Criterion), Dbar (Adjusted Deviance), P^D (Number of Parameters).*

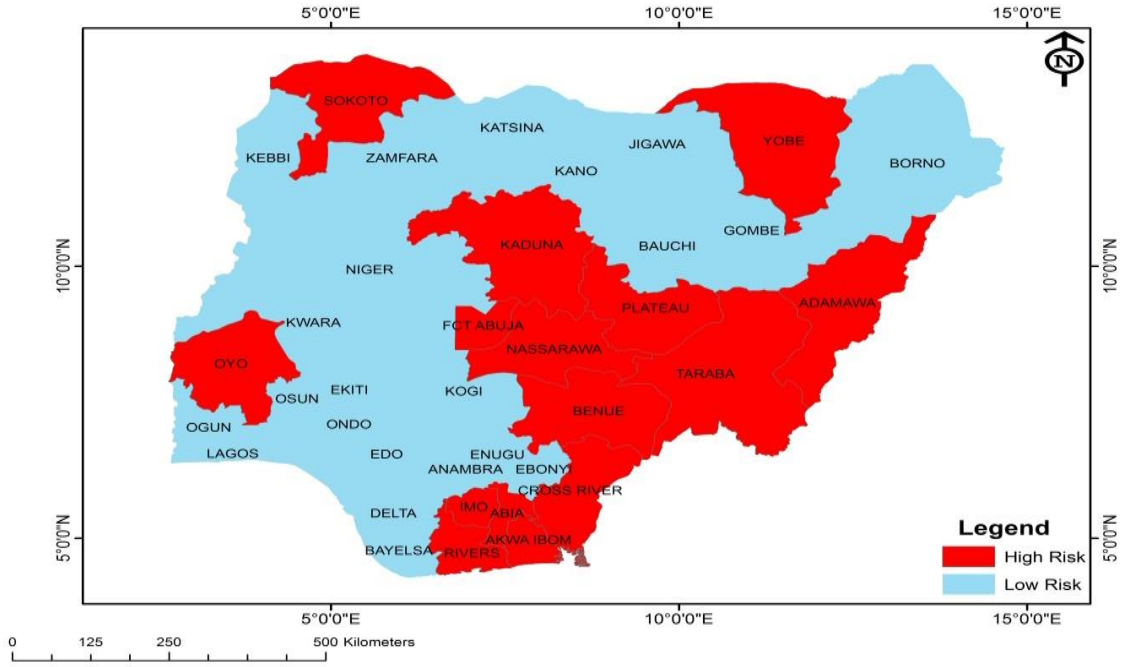


Figure 3.2.HIV Incidence Mapping By State in Nigeria (2014) for PG Model Estimates under Uncontaminated Data

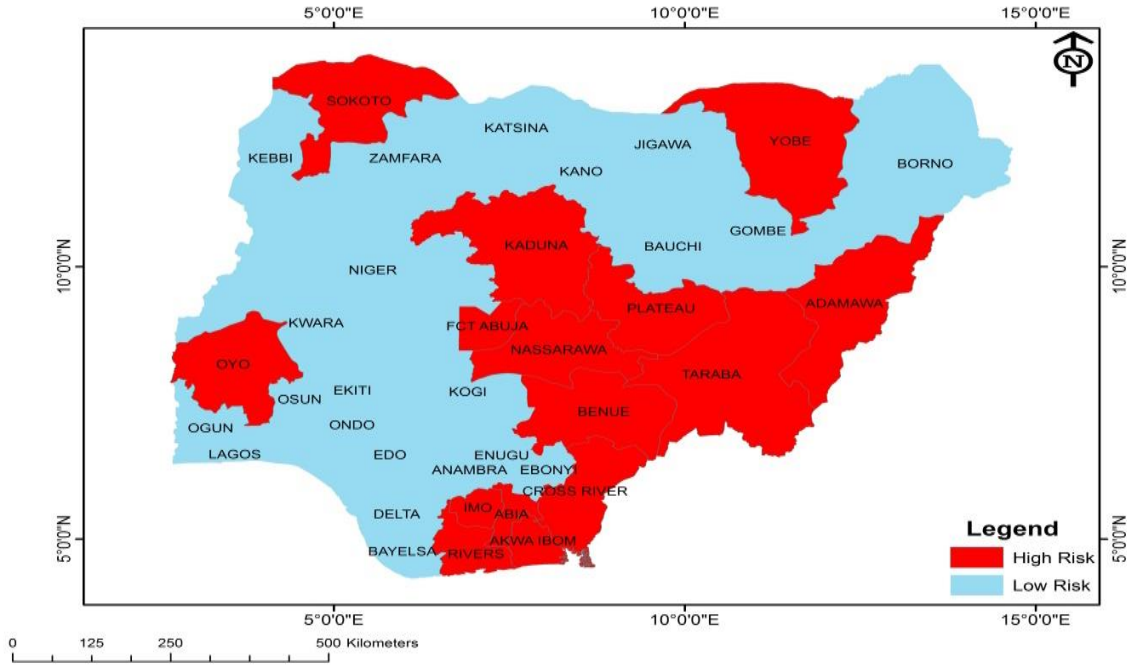


Figure 3.3.HIV Incidence Mapping By State in Nigeria (2014) for PGG Model Estimates under Uncontaminated Data

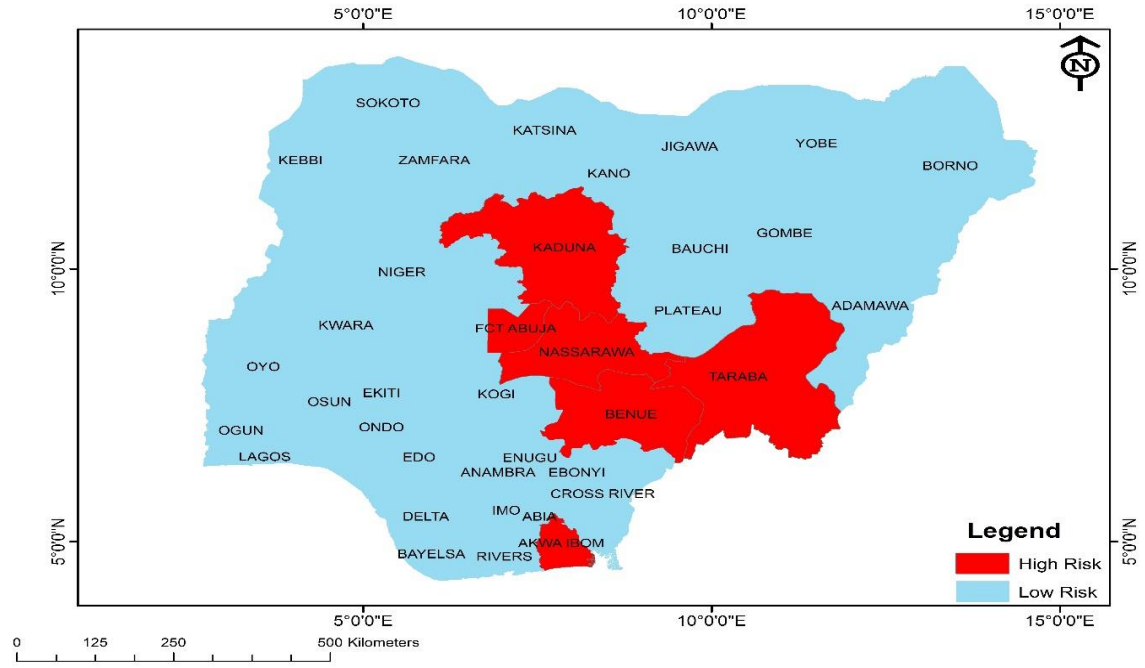


Figure 3.4.HIV Incidence Mapping By State in Nigeria (2014) for GPGG Model Estimates under Uncontaminated Data

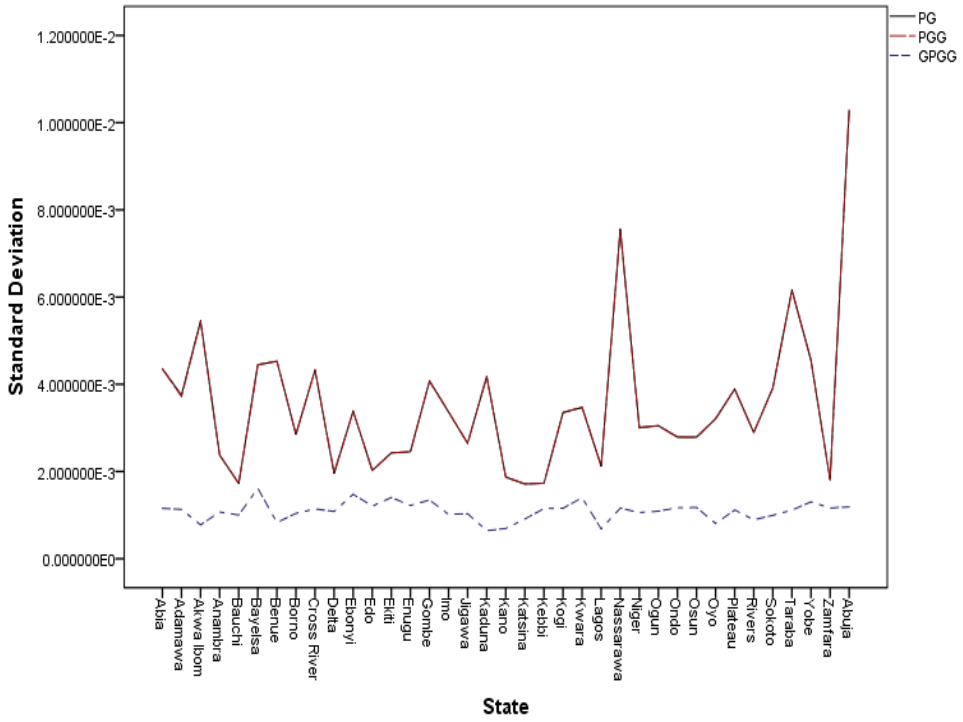


Figure 3.5. Line Graph of Standard Deviations of Different EB models under Uncontaminated Data

Table 3.7. Posterior Relative Risk Estimates and SD of Different EB Models by State under Contaminated Data using Malaria cases by State in Nigeria (2014)

	PG Model		PGG Model		GPGG Model	
State	$\widehat{\theta}_i^{PG}$	**SD**	$\widehat{\theta}_i^{PGG}$	**SD**	$\widehat{\theta}_i^{GPGG}$	**SD**
Abia	0.9985	1.567000	1.1000*	1.225000	3.43E-05	0.000046
Adamawa	1.0020*	1.565000	1.0970*	1.223000	3.08E-05	0.000041
AkwaIbom	0.9992	1.569000	1.0980*	1.225000	2.50E-05	0.000033
Anambra	0.0598	0.001565	0.0598	0.001565	0.0585	0.001532
Bauchi	0.9999	1.573000	1.1010*	1.228000	2.09E-05	0.000028
Bayelsa	0.5400	0.007360	0.5400	0.007360	0.4481	0.006124
Benue	1.0020*	1.570000	1.0970*	1.219000	2.30E-05	0.000031
Borno	0.1509	0.002498	0.1509	0.002498	0.1427	0.002352
Cross River	5.6970*	0.018380	5.6970*	0.018340	1.8010*	0.005810
Delta	0.9965	1.566000	1.0970*	1.222000	2.37E-05	0.000031
Ebonyi	3.5700*	0.016790	3.5700*	0.016740	1.5160*	0.007133
Edo	1.0060*	1.575000	1.0960*	1.222000	3.06E-05	0.000041
Ekiti	1.1340*	0.009005	1.1340*	0.008997	0.7928	0.006303
Enugu	0.0984	0.002275	0.0984	0.002275	0.0949	0.002190
Gombe	1.0020*	1.568000	1.1000*	1.225000	4.13E-05	0.000055
Imo	0.9957	1.552000	1.0940*	1.220000	2.47E-05	0.000033
Jigawa	2.2490*	0.009405	2.2490*	0.009407	1.2130*	0.005074
Kaduna	0.9984	1.563000	1.0970*	1.222000	1.61E-05	0.000021
Kano	0.8588	0.003969	0.8588	0.003955	0.6476	0.002986
Katsina	1.0030*	1.579000	1.0970*	1.223000	1.69E-05	0.000022
Kebbi	4.5500*	0.015500	4.5490*	0.015520	1.6680*	0.005682
Kogi	1.0000*	1.569000	1.0940*	1.222000	2.96E-05	0.000039
Kwara	0.9995	1.560000	1.0950*	1.221000	4.10E-05	0.000055
Lagos	1.6710*	0.005626	1.6710*	0.005623	1.0230*	0.003446
Nassarawa	0.8746	0.008960	0.8746	0.008960	0.6566	0.006739
Niger	1.0010*	1.561000	1.0980*	1.222000	2.47E-05	0.000033
Ogun	1.0040*	1.570000	1.0960*	1.224000	2.61E-05	0.000035
Ondo	6.5080*	0.017990	6.5080*	0.017990	1.8750*	0.005165
Osun	5.7720*	0.016960	5.7720*	0.016960	1.8090*	0.005324
Oyo	1.0010*	1.562000	1.0970*	1.223000	1.74E-05	0.000023
Plateau	0.3136	0.004114	0.3136	0.004100	0.2802	0.003671
Rivers	0.5592	0.004297	0.5593	0.004285	0.4613	0.003541
Sokoto	3.0470*	0.011890	3.0470*	0.011870	1.4130*	0.005508
Taraba	1.4270*	0.010290	1.4270*	0.010320	0.9256	0.006667
Yobe	0.4294	0.005634	0.4294	0.005631	0.3692	0.004837
Zamfara	0.0321	0.001298	0.0321	0.001292	0.0317	0.001281
Abuja	0.9969	1.566000	1.0960*	1.218000	0.0001	0.000091
Nigeria	**1.0000**		**1.0000**		**1.0000**	
DIC	271.70		270.66		118.00	
DT	15.73		15.69		7.37	
Dbar	251.70		251.70		118.00	
P^D	20.00		19.96		19.99	
MCE	0.001612193		0.000894493		0.000003969	
5% SD	0.036228162		0.028321243		0.000124432	

Note: **Asterisk implies Relative Risk ≥ 1 (High Risk Area). Non-asterisk implies Relative Risk < 1 (Low Risk Area). SD (Standard Deviation). MCE (Monte Carlo error), DT (Dispersion Test), DIC (Deviance Information Criterion), Dbar (Adjusted Deviance), P^D (Number of Parameters).*

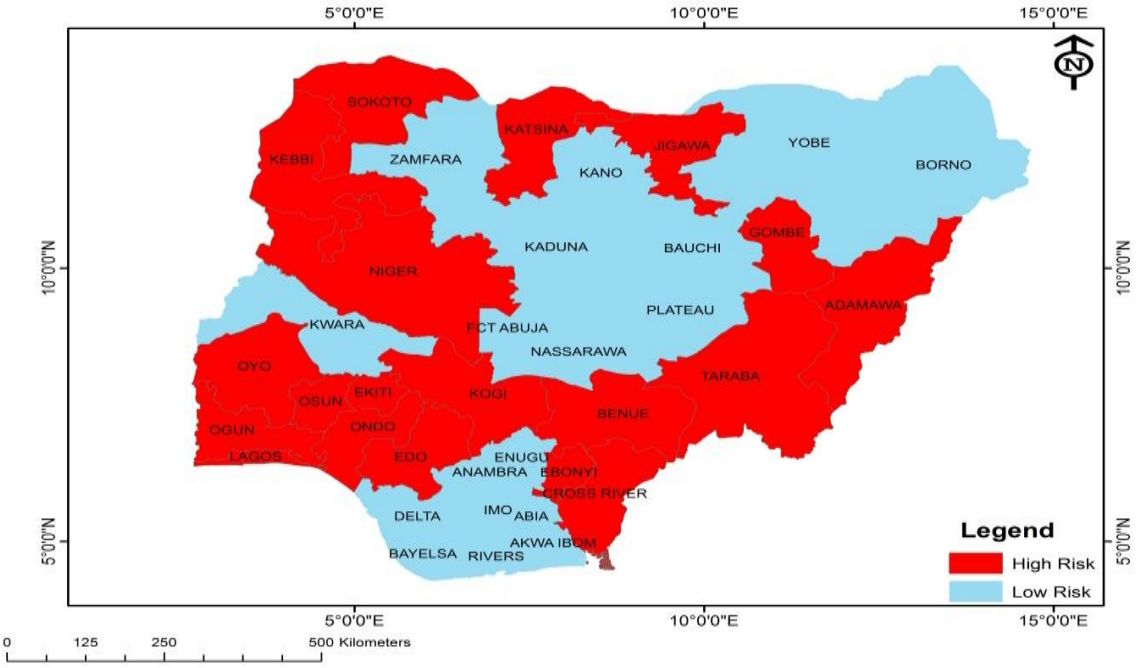


Figure 3.6. Malaria Incidence Mapping By State in Nigeria (2014) for PG Model Estimates under Contaminated Data

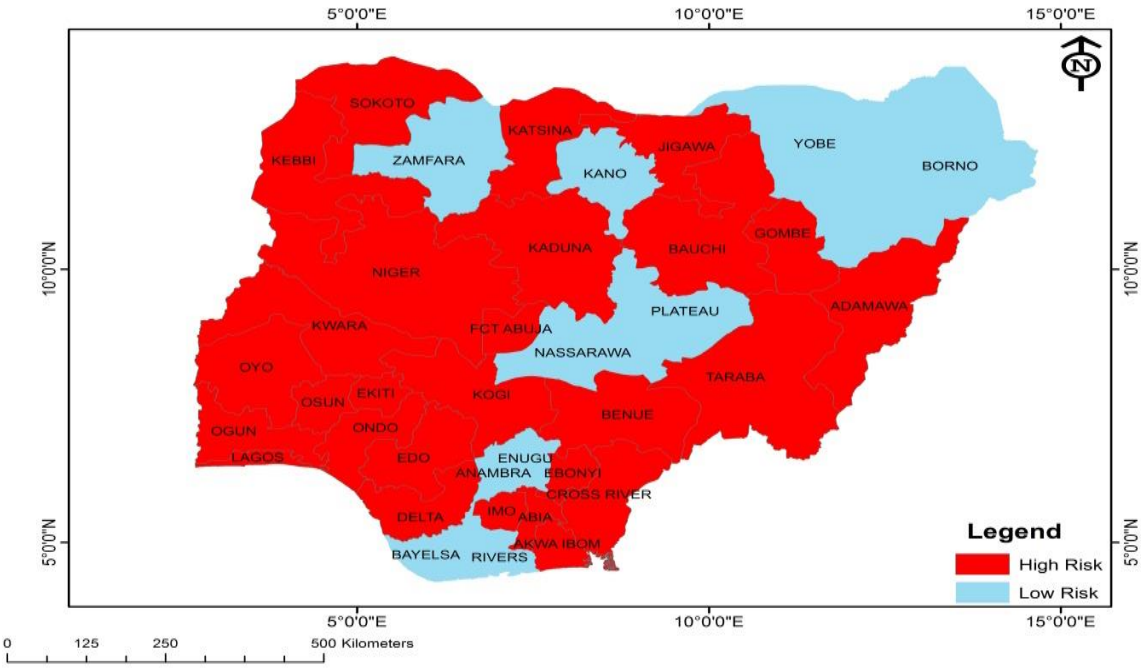


Figure 3.7. Malaria Incidence Mapping By State in Nigeria (2014) for PGG Model Estimates under Contaminated Data

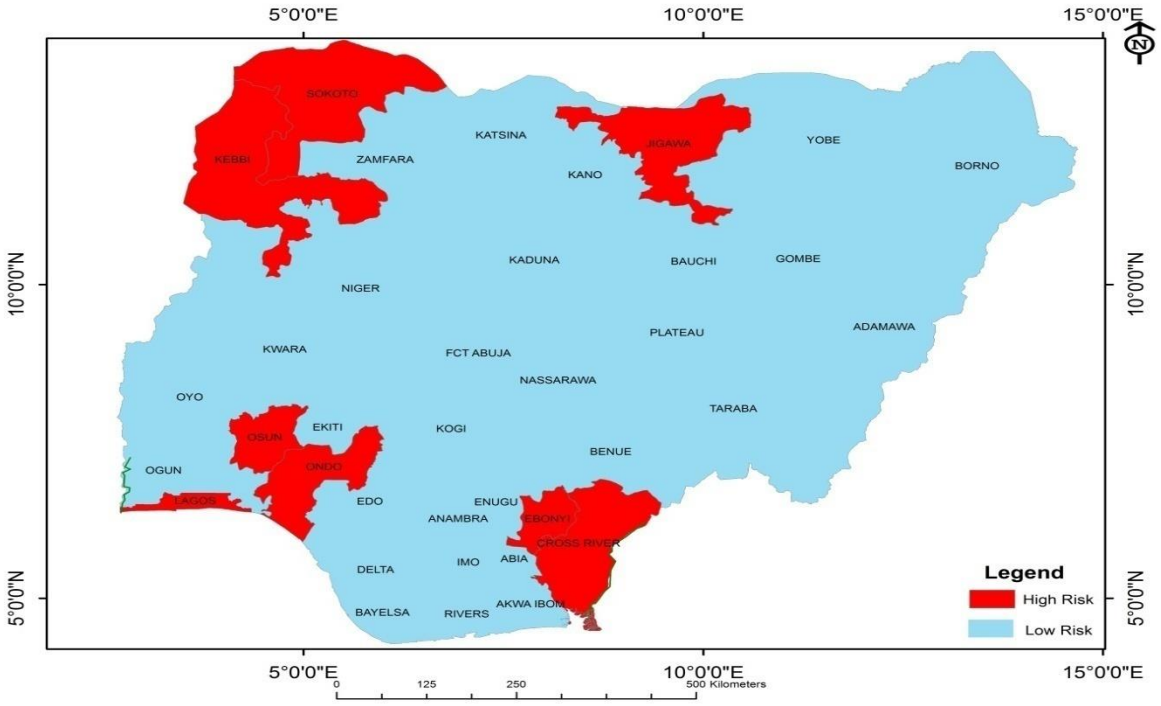


Figure 3.8. Malaria Incidence Mapping By State in Nigeria (2014) for GPGG Model Estimates under Contaminated Data

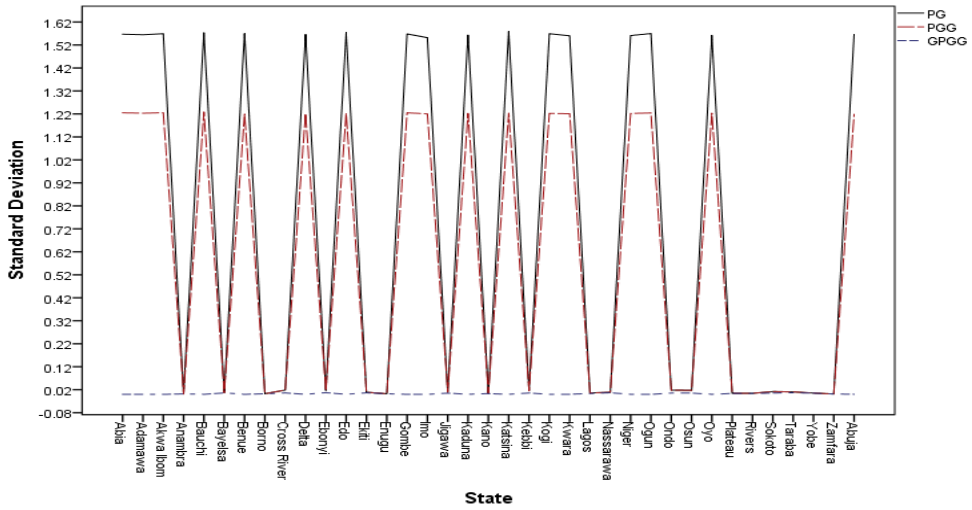


Figure 3.9. Line Graph of Standard Deviations of Different EB models under Contaminated Data

4. DISCUSSION OF RESULTS

4.1 Discussion of Simulation Results under Uncontaminated Data

From the simulation resultsof the different EB models under uncontaminated data in Table 3.3, for both small and large samples of 5, 10, 20 and 50, 100, 200 (that is, overdispersed, equidispersed and underdispersed data respectively), the PGG and GPGG models did better

than the existing PG model in handling dispersion, since they have smaller SD, DIC and DT values. In addition, GPGG model has the smallest values, hence the most efficient in handling dispersion in small and large samples under uncontaminated data. The results suggest that GPGG and PGG models have better precision in handling dispersion than PG model. These results are supported by the relative efficiency presented in Table 3.5; where GPGG and PGG models are relatively more efficient than PG model as the ratios are less than 1 respectively. The simulation results of the posterior estimates are reliable, since there is convergence of MCMC sampling as the chains overlap each other, as shown in the stationary dynamic trace plots in Figure 3.1 for each model.The results indicated under uncontaminated data, for both small and large samples characterize with overdispersion, equidispersion or underdispersion, GPGG provided the best precision and model efficiencyin handling dispersion.

4.2 Discussion of Simulation Results under Contaminated Data

From the simulation resultsof the different EB models under contaminated data in Table 3.4, for both small and large samples size of 5, 10, 20 and 50, 100, 200 (that is, overdispersed, equidispersed and underdispersed data respectively), the new EB models (PGG and GPGG) outperformed the existing PG model in handling dispersion, since they havesmaller SD values and better goodness of fit with lesser DIC and DT values compared to PG model. The results further revealed that GPGG model has the smallest SD, DIC and DT values, hence the most efficient in handling dispersion insmall and large samples under contaminated data. The results suggest that GPGG and PGG models have better precision in handling dispersion than PG model. These results are supported by the relative efficiency presented in Table 3.5; where GPGG and PGG models are relatively more efficient than PG model as the ratios are less than 1 respectively. The simulation results of the posterior estimates are reliable, since there is convergence of MCMC sampling as the chains overlap each other, as shown in the stationary dynamic trace plots in Figure 3.1 for each model.The results indicated under contaminated data, for both small and large sample data characterize with overdispersion, equidispersion or underdispersion, GPGG provided the best precision and model efficiencyin handling dispersion.

4.3 Discussion of Real Data Application Results under Uncontaminated Data

The results from Table 3.6, indicate that under uncontaminated data the estimates of relative risk (RR) of HIV cases by State in Nigeria range from 0.2257 (Kebbi State) to 3.4330 (Abuja FCT) respectively for PG and PGG EB models while for GPGG model, it ranges from 0.0756 (Kebbi State) to 3.0369 (Abuja FCT) respectively. RR≥1 imply higher risk while RR<1 imply lower risk. Fifteen (15) States (Abia, Adamawa, Akwa-Ibom, Benue, Cross-River, Imo, Kaduna, Nassarawa, Oyo, Plateau, Rivers, Sokoto, Taraba, Yobe and FCT Abuja) were reported to be at higher risk by PG and PGG models. While six (6) States (Akwa-Ibom, Benue, Kaduna, Nassarawa, Taraba and FCT Abuja) were reported to be at higher risk by GPGG model. The HIV disease mappings are in Figures 3.2, 3.3, and 3.4 respectively. The PG and PGG models produce similar estimates of relative risk. Also, the two EB models exhibited similar characteristics based on the DIC results (Table 3.6) under an uncontaminated data that is underdispersed. To evaluate the precision of the posterior estimates and the convergence of the EB models, the Monte Carlo error (MCE) is calculated (Table 3.6). According to Brooks and Gelman (1998), the simulation is done until MCE<5% of the sampled standard deviation. Hence, the results indicate that there is accuracy of the

posterior estimates for each EB model and convergence of MCMC sampling, since MCE<5%SD respectively. The dispersion test for each model (Table 3.6) suggests that the count data across the regions are highly underdispersed at -511.19,-511.19, and -138.32 respectively. In general, the GPGG model has the least SDs and DIC value at 139.70, hence provided the best model efficiency in handling dispersion in relative risk estimation for disease mapping under uncontaminated data. Standard deviations of the PGG model are mostly close to the PG model, which indicates that there is no much gain in precision by PGG model under uncontaminated data. This is supported by DIC values and line graphof standard deviations in Figure 3.5.Based on the results, GPGG is comparatively the best modelin handling dispersion under uncontaminated.

4.4 Discussion of Real Data Application Results under Contaminated Data

The results from Table 3.7, indicate that under contaminated data characterized with missing values, the estimates of relative risk (RR) of Malaria cases by State in Nigeria range from 6.5080 (Ondo State) to 0.0321 (Zamfara) for PG and PGG models, and 1.8750 (Ondo State) to 1.61E-05 (Kaduna) for GPGG model. RR≥1 imply higher risk while RR<1 imply lower risk. Therefore, nineteen (19) States (Adamawa, Benue, Cross-River, Ebonyi, Edo, Ekiti, Gombe, Jigawa, Katsina, Kebbi, Kogi, Lagos, Niger, Ogun, Ondo, Osun, Oyo, Sokoto and Taraba) were reported to be at higher risk by PG model. Twenty-Seven (27) States (Abia, Adamawa, Akwa-Ibom, Bauchi, Benue, Cross-River, Delta, Ebonyi, Edo, Ekiti, Gombe, Imo, Jigawa, Kaduna, Katsina, Kebbi, Kogi, Kwara, Lagos, Niger, Ogun, Ondo, Osun, Oyo, Sokoto, Taraba and FCT Abuja) were reported to be at higher risk by PGG model. While eight (8) States (Cross-River, Ebonyi, Jigawa, Kebbi, Lagos, Ondo, Osun and Sokoto) were reported to be at higher risk by GPGG model. Also GPGG and PGG models provided a better prediction of the relative risk (RR) parameter with minimum variance than PG model under a contaminated data characterized with missing values (States Malaria cases were not reported). The PG and PGG models produce similar estimates of relative risk at the States with reported cases of Malaria. But at the States where Malaria cases were not reported, the relative risk estimations are dissimilar. Since the standard deviations of the PGG model is mostly less than the PG model, the PGG model is equally competitive in relative risk estimation for disease mapping under contaminated data. Generally, GPGG model is the least variable and has the smallest DIC, SD and DT values hence the most efficient model to handle overdispersion in relative risk estimation for disease mapping under contaminated data. These are supported by the line graph plot of the standard deviations in Figure 3.9; indicating GPGG and PGG are comparatively better models than PG model. To assess convergence and the accuracy of the posterior estimates for each EB model, the Monte Carlo error (MCE) is calculated (Table 3.7). The results indicate that there is accuracy of the posterior estimates for each EB model, since MCE<5%SD respectively and convergence of MCMC sampling. The dispersion test for PG, PGG, and GPGG (Table 3.7) suggests that there is sign of overdispersion in the count data across the regions at 15.73, 15.69, and 7.37 respectively. Therefore, based on model efficiency in relative risk parameter estimation, GPGG model provided the most reliable disease mapping (Figures 3.6, 3.7, and 3.8).

5.CONCLUSION

The current study aimed at handling dispersion towards enhancing the efficiency of relative risk estimation in disease mapping using generalized Gamma distribution aspriorwith Poisson and generalized Poisson likelihoods, for health risk assessment.The study is as a

result of recent needs for development of Bayesian priors and modeling using generalized distributions. The developed EB models using GG distribution as prior revealed a proper conjugate prior with Poisson and generalized Poisson likelihoods, since the posterior and prior distributions are in the same class of distribution**.**The study has shown that the use of generalized Gamma distribution as Bayesian prior with Poisson and generalized Poisson likelihoods improved efficiency in relative risk estimations in small and large samples underuncontaminated and contaminated data characterized with either overdispersion or underdispersion. This is clearly evident in GPGG EB model which showed to be the most efficient, based on the current study, in handling dispersion in relative risk estimation for disease mapping when data is uncontaminated and contaminated. PGG is comparatively a better model than PG model in handling dispersion, especially when data is contaminated. As showed in the simulation results, PG EB model is inept to handle underdispersion; therefore upholding the view expressed inZou *et al.*,(2018), Lord and Geedipally (2016), and Lord (2006).Hence,GPGG and PGG models can serve as potential alternatives in providing reliable disease mapping. Key suggestions for further study includes deriving spatial correlation models, including covariates which were not considered in this paper due to the nature of data collected and deriving the bivariate and the multivariate cases of GPGG and PGG EB models for mapping of two or more diseases at the same time and the implementation to aid practitioners.

Acknowledgment: The authors use this medium to thank the National Bureau of Statistics (NBS) for making the study data readily available online.

Authors Contributions: The research was initiated by the first author who carried out the initial and final draft of the paper. The proof-reading and necessary corrections of the initial and final draft of the paper were carried out by the second and third authors as supervisors of the main PhD thesis.

Authors' Conflicts of interest: The authors declare that there are no conflicts of interest regarding the publication of this paper.

Funding Statement: This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

REFERENCES

[1] I. Adeleke,A. Ibiwoye, R. Ogundeji. Bayesian Estimation of the Proportion of Subscribers to Nigeria's National Health Insurance Scheme. Journal of Modern Mathematics and Statistics. 3(3) (2009), 56-59.

[2] A. Azzalini. A class of distributions which includes the normal ones. Scandinavian Journal of Statistics. 12 (1985), 171-178.

[3] J. Besag, J. York, A. Mollie. Bayesian Image Restoration, with Two Applications in Spatial Statistics (with discussion). Annals of the Institute of Statistical Mathematics.43 (1991),1-59.

[4] J. Besag. Spatial Interaction and the Statistical Analysis of Lattice Systems, Journal of the Royal Statistical Society. B36 (1974), 192-236.

[5] D. Böhning,E. Dietz, P. Schlattmann. Space-time mixture modelling of public health data.Stat Med. 19 (2000), 2333-2344.

[6] G. E. Box,G. C.Tiao. Bayesian Inference in Statistical Analysis.Addison-Wesley Publishing Company. Massachusetts (1973).

[7] S. Brooks,A. Gelman. Alternative methods for monitoring convergence of iterative simulations.Journal of Computational and Graphical Statistics.7(1998),434-455.

[8] B. P. Carlin,T. A. Louis.Bayesian Methods for Data Analysis.Chapman and Hall/CRC.Boca Raton (2009).

[9] G. Casella.An introduction to Empirical Bayes Data Analysis.The American Statistician. 39(1985),83-87.

[10] D. Clayton, J. Kaldor. Role of advanced statistical techniques in Cancer mapping.Recent Results in Cancer Research. 114 (1989), 87-98.

[11] D. Clayton,J. Kaldor.Empirical Bayes estimates of Age-standardized Relative Risks for use in Disease Mapping.Biometrics. 43 (1987), 671-681.

[12] E. P. Clement. Small Area Estimation with Application to Disease Mapping.International Journal of Probability and Statistics. 3(1) (2014), 15-22.

[13] P. C. Consul,G. C. Jain.A generalization of the Poisson distribution.Technometrics. 15(4) (1973), 791–799.

[14] N. Cressie, N. Chan. Spatial Modeling of Regional Variables. Journal of the American Statistical Association. 84 (1989), 393-401.

[15] B. Efron.Empirical Bayes Methods for combining Likelihoods (with discussion). Journal of the American Statistical Association. 91 (1996),538–565.

[16] F. Famoye,R. Okafor, M. Adamu. A multivariate generalized Poisson distribution. Journal of Statistical Theory andApplications.10(3) (2011), 519-531.

[17] F. Famoye. A new bivariate generalized Poisson distribution.StatisticaNeerlandica. 64 (2010), 112-124.

[18] A. Gelman,B. Carlin,S. Stern, D. Rubin.Bayesian Data Analysis.Chapman and Hall/CRC.Boca Raton, Florida (2004).

[19] T. Hwang,P. Huang, P. On new moment estimation of parameters of the Generalized Gamma distribution using its characterization.Taiwanese Journal of Mathematics. 10 (4) (2006), 1083-1093.

[20] T. Koch. Cartographies of Disease: Maps, Mapping, and Medicine. ESRI Press. Redlands, California (2005).

[21] A. B. Lawson. Bayesian Disease Mapping: Hierarchical Modeling in Spatial Epidemiology. Chapman and Hall/CRC. Boca Raton (2013).

[22] D. Lord,S. Geedipally. Safety Prediction with Datasets Characterized with Excess Zeros and Long Tails. Emerald Group Publishing Limited. Bingley UK (2016).

[23] D. Lord. Modeling motor vehicle crashes using Poisson-gamma models: Examining the effects of low sample mean values and small sample size on the estimation of the fixed dispersion parameter. Accident Analysis and Prevention. 38 (2006), 751-766.

[24] K. G. Manton, M. A. Woodbury, M. A, Stallard. A variance components approach to categorical data models with heterogeneous cell populations: Analysis of spatial gradients in lung cancer mortality rates in North Carolina counties. Biometrics. 37 (1981), 259-269.

[25] R. J. Marshall. Mapping Disease and Mortality rates Using Empirical Bayes Estimators. Applied Statistics. 40 (1991), 283-294.

[26] U. Mbata,R. Okafor, I. Adeleke. Poisson-Generalized Gamma Empirical Bayes Model for Disease Mapping.International Journal of Science and Technology. 7 (1) (2018), 34-56.

[27] U. A. Mbata,G. A. Olalude, B. E. Ahani,S. A. Amusan,H. A. Adetigha.Parametric Empirical Bayesian Procedure for Estimating the Mean Age and Weekly Allowance of Undergraduate Students at University of Lagos.International Journal of Statistics and Systems. 4(5) (2010), 497-504.

[28] J. L. Meza. Empirical Bayes estimation smoothing of relative risks in disease mapping.Journal of Statistical Planning and Inference. 112 (2003), 43 – 62.

[29] I. Miller, M. Miller. John E. Freund's Mathematical Statistics with Applications. Pearson Prentice Hall. Upper Saddle River New Jersey (2004).

[30] A. F. Militino, M. D. Ugarte, C. B. Dean. The use of mixture models for identifying high risks in disease mapping. Statistics in Medicine. 20 (13)(2001), 2035-2049.

[31] C. N. Morris. Parametric Empirical Bayes Inference: Theory and Applications. Journal of the American Statistical Association.78 (1983), 47–65.

[32] R. O. Okafor, U. A. Mbata. A Bayesian Model for Inference on Population Proportions.WIREs Computational Statistics. 4 (2012), 482–488.

[33] R. O. Okafor, A. I. Opara, R. K. Ogundeji, U. A. Mbata, G. A. Olalude. Empirical Bayes Model for Inference on Vehicular Traffic Density at the Main Campus of University of Lagos, Lagos, Nigeria.International Journal of Statistical System. 5 (2010), 53–62.

[34] H. Robbins, An empirical Bayes approach to statistics. Journal of Mathematical Statistics and Probability. 1(1955),157-163.

[35] R. Srinivasan,P. Venkatesan. Bayesian random effects model for disease mapping of relative risks. Annals of Biological Research. 5 (1) (2014),23-31.

[36] E. W. Stacy. A generalization of the Gamma distribution.The Annals of Mathematical Statistics. 33 (1962), 1187-1192.

[37] Y. Zou, J. Ash, B. Park, D. Lord,L. Wu. Empirical Bayes estimates of finite mixture of negative binomial regression models and its application to highway safety. Journal of Applied Statistics. 45(9)(2018), 1652-1669.

UGOCHUKWU MBATA*
MATHEMATICS DEPARTMENT, UNIVERSITY OF LAGOS, LAGOS, NIGERIA.
E-mail address: mbataugochukwu@gmail.com

ISMAIL ADELEKE
ACTURIAL SCIENCE AND INSURANCE DEPARTMENT, UNIVERSITY OF LAGOS, LAGOS NIGERIA.
E-mail address: adeleke22000@yahoo.ca

KAYODE ADEKEYE
MATHEMATICS DEPARTMENT, UNIVERSITY OF LAGOS, LAGOS NIGERIA.
E-mail address: adekeyek@run.edu.ng