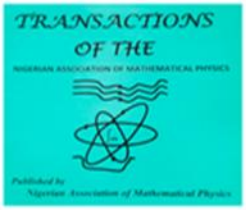
Transactions of NAMP 21, (2025) 183-204

Transactions of
The Nigerian Association of
Mathematical Physics
Journal homepage: https://nampjournals.org.ng



DESIGN-BASED GENERALIZED LINEAR MIXED MODEL FOR BINOMIAL OUTCOME TWO-STAGE SURVEY USING LAPLACE APPROXIMATION WITH APPLICATION TO 2021 NIGERIA MALARIA INDICATOR SURVEY DATA

Olakiigbe A.K[1,2,*], Dawodu G.A[2], Olayiwola O.M.[2], Olurode S.A.[3], Wale-Orojo O.A.[2], and Akintunde A.A.[2].

[1]*Monitoring & Evaluation Unit, Nigerian Institute of Medical Research 6, Edmund Crescent off Murtala Mohammed Way, P.M.B. 2013, Yaba, Nigeria*
[2]*Department of Statistics, Federal University of Agriculture, Abeokuta, P.M.B 2240, Abeokuta, Ogun State, Nigeria*
[3]*Department of Veterinary Surgery & Theriogenology, Federal University of Agriculture, Abeokuta, Ogun State, Nigeria*

ARTICLE INFO

Article history:

ABSTRACT

Applied statistical analyses have increasingly incorporated complex survey designs, though generalized linear mixed models (GLMMs) remain scarce. This study developed a GLMM for two-stage samples using integrated nested Laplace approximation (INLA) on the 2021 Nigeria Malaria Indicator Survey (NMIS) data. A binomial outcome GLMM was fitted, treating design sampling weights as a Gaussian latent model, and posterior estimates were obtained using INLA. Simulation and real data applications compared classical, weighted, MCMC, and INLA approaches, evaluated through accuracy and model diagnostics. Results indicated that incorporating design weights improved model fits, with the INLA GLMM showing superior performance. INLA also required the least computational time, highlighting its advantage for large survey datasets. The study recommends the design-based GLMM approach using INLA for analyzing large-scale survey data

Keywords:
Design-based model, Sampling weights, Laplace approximation, Malaria indicator, Two-stage design.

1. Introduction

Generalized linear mixed model (GLMM) is a pervasive analysis in applied statistics in areas such as health and medical fields, ecological and geographical fields, and even in economics and finance. This is not unconnected with its structure, which allows many complex experimental designs to be handled within the familiar linear model framework. With the progress in statistical methodologies and the availability of powerful computers, GLMMs have become widely used for data analysis in applied settings. GLMM in complex survey design is of interest here.

---

[*] Corresponding author: ABIODUN KANMI OLAKIIGBE
E-mail address: abiodunolakiigbe@hotmail.com

https://doi.org/10.60787/tnamp.v21.511

Not much recent literatures abound on GLMM under complex designs. Addressed both the computational issues and inflation of the variance of fixed effects for spatial generalized linear mixed models [1]. Proposed two methods relying on auxiliary variables for missing covariate data and unknown distribution assumptions for random effects to analyze the data using a penalized conditional likelihood (PCLE) method [2]. Examined the impact of misspecifying a linear mixed model for single-case experimental design data [3]. Proposed two mixed model methods to monitor profiles from the exponential family [4]. Compared 95% confidence intervals for small area estimation for health indicators via GLMM [5].

Modelled grouped responses using GLMM with many explanatory variables, redundancies, and collinearities [6]. Extended the GLMM to the simultaneous modelling of multiple mixed outcomes. [7]. A multivariate GLMM framework was presented, which allows the specification of a set of response variables for under-dispersed count data [8]. Presented GLMM for multivariate responses with Poisson, negative binomial, and Conway-Maxwell-Poisson distributions [9]. Adeniyi and Yahya (2020) proposed a class of Generalized Linear Mixed Models (GLMM) in which the assumption of normality was replaced by the class of normal-independent distributions introduced by Liu [10, 11]. Compared results from INLA, JAGS' Gibbs sampling, and Stan's Hamiltonian Monte Carlo for two longitudinal count data [12,13]. Introduced Laplace approximation to Bayesian inference for Dirichlet regression models, suitable for multivariate compositional data with skewness and heteroscedasticity, with no need for data transformation [14].

Proposed GLMM with a level-specific random effect for non-repeated item designs where items are only used in one level of experimental conditions [15]. Developed a semiparametric mixed-effect regression model for data from a two-stage design [16]. Studied a model-based simulation procedure using the Monte-Carlo EM algorithm [17]. Simulation study evaluated in respondent-driven clustering, the validity of different regression models, with and without weights [18]. Presented a two-stage estimator for clustered count correlated data with overdispersion [19]. Examined and detailed the primary inference methods utilized for analyzing data derived from complex surveys [20]. Applied regression analysis for a randomized response technique with quantitative variables in complex sample designs [21].

Available regression models for design-based survey analysis only cater to linear regression and generalized linear model (we refer to [22]) for material on regression models for survey data using R package Survey [23], which handles the regression models except for GLMM). This study aimed at GLMMs under complex design survey case of two-stage outcome, specifically binary outcome using integrated nested Laplace approximation (INLA) with application to malaria indicator data. Multistage sampling is a survey design that entails more than one step of random selection. For example, a two-stage sampling could entail sampling households within strata, and then at a second stage, individuals are sampled within the already sampled households. Taking into consideration the correlation among sampled observations when estimating sampling quantities and model coefficients becomes very crucial at this point. Usually, the interest of surveys is in estimating the characteristics of the finite population, and hence, the sampling design is important in the analysis. The design-based inference is centered on how the sampling design information is implemented in the estimation process for population parameters. Such a survey design involves complex surveys featuring multistage sampling.

2 Methods

2.1 Sample Probabilities and Weights in the motivating dataset

We review the sampling procedure used in the motivating dataset, the 2021 Nigeria Malaria Indicator Survey (NMIS), which was a two-stage stratified cluster survey. Sampling weights were computed separately for each stage and cluster based on sampling probabilities. Let P_{1hi} denote the first-stage sampling probability of the i^{th} cluster in stratum h and P_{2hi} denote the second-stage sampling probability within the i^{th} cluster. Let a_h be the number of sampled clusters in stratum h, M_{hi} the number of households in the i^{th} cluster, and $\sum M_{hi}$the total number of households in the stratum. The probability of selecting the i^{th} cluster is

$$\frac{a_h M_{hi}}{\sum M_{hi}}$$

Also, let b_{hi} be the number of households in stratum h if the enumeration area (EA) is segmented and $b_{hi} = 1$ otherwise. Then the probability of selecting the i^{th} cluster is

$$P_{1hi} = \frac{a_h M_{hi}}{\sum M_{hi}} b_{hi}$$

Given L_{hi} as the number of households listed in cluster i and stratum h, and g_{hi} is the number of households selected in the cluster. The probability of selecting each household from the cluster is

$$P_{2hi} = \frac{g_{hi}}{L_{hi}}$$

The overall probability of selecting each household in cluster i of stratum h is then

$$P_{hi} = P_{1hi} \times P_{2hi}$$

The sampling weight for each household in cluster i of stratum h is given as

$$W_{hi} = 1/P_{hi}$$

The total number of the observation is then given by

$$N = \sum_{h=1}^{H} \sum_{i=1}^{a_h} \sum_{j=1}^{M_{hi}} \sum_{m=1}^{L_{hi}} n_{mjih} \tag{1}$$

2.2 Proposed GLMMs for binary two-stage survey with Laplace approximation inference.

Let y_{hi} be the binary response variable for each household in i^{th} cluster and h^{th} stratum, then we have that

$$y_{hi} \mid z_{hi}, x_{hi}, \beta_i \overset{i.d}{\sim} Bernolli\,(\mu_{hi}), \quad \text{logit}(\mu_{hi}) = x_{hi}^T \beta + z_{hi}^T u_i \tag{2}$$
$$u_i \sim N(0, G)$$

Equation (2) is an exponential family of distribution because its distribution can be written as thus

$$f(y_{hi}) = p^{y_{hi}}(1-p)^{1-y_{hi}}, \qquad \text{where } y_{hi} = 0, 1;\ \ 0 \le p \le 1$$
$$f(y_{hi}) = \exp[\log(p^{y_{hi}}(1-p)^{1-y_{hi}})]$$
$$= \exp[y_{hi} \log p + (1 - y_{hi}) \log(1-p)]$$
$$= \exp\left[y_{hi} \log \frac{p}{1-p} + \log(1-p)\right]$$
$$= \exp\left[y_{hi}\zeta_{hi} - \log\left(1 - e^{\zeta_{hi}}\right)\right],$$

with $\zeta = \log \frac{p}{1-p}$, $b(\zeta_{hi}) = \log\left(1 - e^{\zeta_{hi}}\right)$, $a_{hi}(\phi) = 1$, $c_i(y_{hi}, \phi) = 0$.

Given the $N \times 1$ vector $\boldsymbol{w}$ of sampling weights containing information on the sampling design employed in the survey, $\boldsymbol{w} = \boldsymbol{w}(\text{P})$. As already seen in the previous section, the sampling weight w_i is the inverse of the overall selection probability. The sampling weights are usually referenced to non-response and/or known population values of some auxiliary variable [17]. The sampling weight is introduced in the calculation of the likelihood for the parameter estimation.

The conditional expectation $E(y_{hi}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{U})$ is linked to a linear predictor η_{hi} with a link function $h(\cdot)^{-1}$

$$h^{-1}\{E(y_{hi}|\boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{U})\} = \eta_{hi} = \boldsymbol{x}'_{hi}\boldsymbol{\beta} + \boldsymbol{z}'_{hi} u_{hi} \tag{3}$$

where $\boldsymbol{x}'_{hi}$ is a vector of p covariates, $\boldsymbol{\beta}$ is a vector of covariates fixed effects, $\boldsymbol{z}'_{hi}$ is $N \times q$ design matrix of random effects and u_{hi} is vector of q random effects. In matrix notation, we have

$$\boldsymbol{Y} = \boldsymbol{X\beta} + \boldsymbol{ZU} + \boldsymbol{\varepsilon}$$

where $\boldsymbol{\varepsilon}$ is $N \times 1$ vector of residuals and

$$\boldsymbol{Y} \mid \boldsymbol{X\beta} + \boldsymbol{ZU} \sim f(\boldsymbol{0}, \boldsymbol{R})$$

In classical statistics, $\boldsymbol{U}$ is assumed as $\boldsymbol{U} \sim N(\boldsymbol{0}, \boldsymbol{G})$**,** with $\boldsymbol{G}$ being a covariance matrix of random effects which are joint deviations around the values in $\boldsymbol{\beta}$, so what is estimated is the covariance. $\boldsymbol{G}$ is a function of $\boldsymbol{\theta}$, which are unknown hyperparameters that influences the distribution of $\boldsymbol{G}$ and they are to be estimated, i.e., $\boldsymbol{G} = \boldsymbol{\psi}(\boldsymbol{\theta})$.With the assumption of independent random effects $\boldsymbol{G} = \text{diag}(\sigma^2_{intercept}, \sigma^2_{slope})$. A common structure for the covariance matrix of residuals is $\boldsymbol{R} = \boldsymbol{I}\sigma^2_{\varepsilon}$, which assumes homogeneous independent residuals variances.

2.2.1 The Latent Gaussian modelling approach

The GLMM herein considered is a latent Gaussian model (LGM) when it can be expressed in the form below

$$\eta_{hi} = h(\mu_i) = \alpha + \sum_{h=1}^{p} \beta_h x_{hi} \sum_{l=1}^{r} u_l z_{li} + \varepsilon_i \tag{4}$$

where in addition to the already defined parameters, μ_i is the mean of the ith individual, α is intercept and ε_i are the residuals or unstructured random effects. The Latent model estimates the vector of parameters $\boldsymbol{\gamma} = \{\boldsymbol{\eta}, \boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{u}\}$. Assuming each parameter follows a Gaussian distribution, we note that this is parsimonious in most cases, then the model is an LGM and if the vector $\boldsymbol{\gamma}$ is assumed to be from a multivariate normal distribution, then it is referred to as a Gaussian latent field, i.e., $\boldsymbol{\gamma} \sim \boldsymbol{N}(\cdot, \boldsymbol{\Omega})$ where

$$\boldsymbol{\Omega} = \begin{pmatrix} \sigma^2 & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \sigma^2 \end{pmatrix}$$

If the Gaussian latent field satisfies Markov properties, then it is called a Gaussian Markov random field (GMRF), therefore $\boldsymbol{\gamma} \sim GMRF(\boldsymbol{\theta})$, where $\boldsymbol{\theta}$ are hyperparameters representing the covariance matrix $\boldsymbol{\Omega}$. A process producing the realizations of a quantity is known to have Markov properties if the expected future state of the process depends only on the present state and not the past state. The GLMM has the parameters of interest as $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \boldsymbol{u})$ with $\boldsymbol{u}$ and $\boldsymbol{\beta}$ following a normal distribution before that the covariance matrix has a diagonal with very large and equal entries for each component of $\boldsymbol{\beta}$ so that the resulting prior is uninformative prior. Then prior distribution is assigned to the parameters of the latent field $\boldsymbol{\gamma}$. The elements of $\boldsymbol{\gamma}$ are conditionally independent, given the stochastic dependence structure, such that it results in a GMRF with a sparse precision matrix $\boldsymbol{Q}(\boldsymbol{\theta})$. With the Markov property, the precision $\boldsymbol{Q} = \boldsymbol{\Omega}^{-1}$, has a large number of zeros making it very sparse and much simpler [24].

From section 2.2, assume the canonical link and $\log f(\cdot)$ with the shape

$$\log f(y_{hi} | \boldsymbol{\gamma}, \boldsymbol{\theta}) = y_{hi} - a(\eta_{hi}) + b(\eta_{hi})$$

with $\frac{\partial a(\eta_{hi})}{\eta_{hi}} = \mu_{hi}$. The finite population likelihood

$$L(\boldsymbol{y}, \boldsymbol{X}, \boldsymbol{Z}, \boldsymbol{\gamma}, \boldsymbol{\theta}) = \prod_{i=1}^{N} \int \prod_{h=1}^{H} w_i f(y_{hi}, \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\gamma}, \boldsymbol{\theta})$$

$$\propto \prod_{i=1}^{N} \prod_{h=1}^{H} f(y_{hi} | \boldsymbol{x}_i, \boldsymbol{z}_i, \boldsymbol{\gamma}, \boldsymbol{\theta}) \cdot \phi(\boldsymbol{\gamma} | \boldsymbol{\sigma}), \tag{5}$$

where $\phi(\cdot\,|\boldsymbol{\sigma})$ is the density of the multivariate normal $N(\mathbf{0},\boldsymbol{\Sigma})$, $\boldsymbol{\sigma}$ denote the vector of matrix elements in $\boldsymbol{\Sigma}$. Due to better readability, $\boldsymbol{X}$ and $\boldsymbol{Z}$ are left out in the writeup, though the functions are conditioned on them. Taking the logarithm of (5) we have

$$\log L(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta}) = \sum_{i=1}^{N}\left(\sum_{h=1}^{H} w_i \log f(y_{hi}|\boldsymbol{\gamma},\boldsymbol{\theta}) + \log\phi(\boldsymbol{\theta}|\boldsymbol{\sigma})\right) \tag{6}$$

Given the finite population parameter vector $\boldsymbol{\theta}^{pop}$ and subvectors $\boldsymbol{\beta}^{pop}$ and $\boldsymbol{\sigma}^{pop}$. $\boldsymbol{\theta}^{pop}$ is defined as

$$\boldsymbol{\theta}^{pop} = \arg\max_{\boldsymbol{\theta}} \log L(\boldsymbol{y},\boldsymbol{\gamma},\boldsymbol{\theta})$$

That means, $\boldsymbol{\theta}^{pop}$ is the maximum likelihood estimate of the super population parameter given that the total finite population was observed. $\boldsymbol{\beta}^{pop}$ exists and is unique for the common GLM such as logistic regression and standard linear regression. The population covariance matrix $\Sigma(\boldsymbol{\sigma}^{pop})$ is unique, too, thus $\boldsymbol{\theta}^{pop}$ is well defined as (6) is globally concave, which means convergence to the global maximum of (6). As the maximum is unique, it implies that the parameter estimate will converge to the maximizer and by extension $\boldsymbol{\theta}^{pop}$ [17].

As already noted, the solution of the marginal likelihood for the integral in (6) under GLMMs is not feasible analytically due to the q-multidimension vectors u_i and even at the finite population level. Examined Laplace's integral approximation using a Taylor series expansion around the mode [25]. The Gauss-Hermite approximation was discussed in Fitzmaurice et al., (2008), in which a Cholesky decomposition of the covariance of the random effect is used, such that $u_i = \boldsymbol{G}(\boldsymbol{\theta})^{1/2}u_i^*$ giving an independent standard normal distribution for $u_i^* \sim N(\mathbf{0},\boldsymbol{I})$ [26]. The penalized quasi-likelihood approach was developed, focusing on penalized quasi-likelihood estimation using Laplace approximation for the marginal likelihood [27, 28].

Given some extreme cases of $\boldsymbol{G}(\boldsymbol{\theta}) = 0$, it implies that for a GLMM with a binary outcome, the observations for a specific individual will always have the same value, and the penalization term tends to infinity such that the maximum likelihood estimate for $u_i$will no longer be defined [29]. In this parameterization, priors are defined for the latent Gaussian field and the hyperparameters and we present brief definitions of prior distributions for an integrated Laplace approximation.

2.3 Laplace Approximation to Marginal Likelihood for GLMMs for binary outcome for two-stage survey

From section 2.2.1, the latent field $\boldsymbol{\gamma}$ is assumed to have a multivariate Gaussian prior with zero mean, forming a GMRF with sparse precision matrix matrix $\boldsymbol{Q}(\boldsymbol{\theta}_2)$, i.e. $\boldsymbol{\gamma} \sim N(\mathbf{0},\boldsymbol{Q}^{-1}(\boldsymbol{\theta}_2))$, and a prior on the hyperparameter vector $\phi(\boldsymbol{\theta})$ is assumed for the set of hyperparameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1,\boldsymbol{\theta}_2)$, to follow the penalized complexity (PC) prior [30], which results in penalizing more complex models to reduce model complexity. The PC priors have been noted to improve point estimation and credible intervals [31]. The marginal likelihood function in equation (6) is then evaluated using Laplace approximation by expressing the integral term as

$$\int \prod_{j=1}^{n_i} f\left(y_{ij}|\boldsymbol{\beta},u_i\right) f\left(u_i|\boldsymbol{G}(\boldsymbol{\theta})\right) du_i = \int f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{u}) f\left(\boldsymbol{u}|\boldsymbol{G}(\boldsymbol{\theta})\right) d\boldsymbol{u}$$

$$= \int e^{\log\left\{f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{u}) f\left(\boldsymbol{u}|\boldsymbol{G}(\boldsymbol{\theta})\right)\right\}} d\boldsymbol{u} = \int e^{h(u)}\, du$$

$$h(u) = \log\left\{f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{u}) f\left(\boldsymbol{u}|\boldsymbol{G}(\boldsymbol{\theta})\right)\right\} \qquad (7).$$

The aim is to choose $\hat{u}$ such that $h(u)$ is maximized and fulfils the necessary and sufficient conditions, $h'(u) = 0$ and $h''(\hat{u}) < 0$. The second order Taylor expansion around $\hat{u}$ for $h(u)$ is given as

$$h(u) \approx \tilde{h}(u) = h(\hat{u}) + (u-\hat{u})h'(\hat{u}) + \frac{1}{2}(u-\hat{u})^2 h''(\hat{u})$$
$$= h(\hat{u}) - \frac{1}{2}(u-\hat{u})^2\left(-h''(\hat{u})\right) \tag{8}$$

From here $e^{\tilde{h}(u)}$ is proportional to the normal density with mean $\hat{u}$ and variance $\frac{1}{-h''(\hat{u})}$. This follows that the Laplace approximation for the likelihood becomes

$$L(\boldsymbol{\beta},\boldsymbol{U},\boldsymbol{\theta}) = \int e^{h(u)}\,du \approx \int e^{\tilde{h}(u)}\,du = \exp\left(h(\hat{u})\int \exp\left(\frac{h''(\hat{u})}{2}(u-\hat{u})^2\right)\right)du$$
$$= \exp\left(-h(\hat{u})\sqrt{\frac{2\pi}{h''(\hat{u})}}\right) \tag{9}$$

It follows that

$$(\boldsymbol{u},\boldsymbol{\beta}|\boldsymbol{y}) = \frac{f(\boldsymbol{y}|\boldsymbol{\beta},\boldsymbol{u})f\left(\boldsymbol{u}\middle|\boldsymbol{G}(\boldsymbol{\theta})\right)}{f(\boldsymbol{y})} \propto \exp\left(h(\boldsymbol{u})\right) \approx C \times \exp\left(\frac{h''(\hat{u})}{2}(u-\hat{u})^2\right) \tag{10}$$

which implies that $\boldsymbol{U}|\boldsymbol{Y} = y \approx N\left(\hat{u}, \frac{1}{-h''(\hat{u})}\right)$.

The Laplace approximation performs better for larger clusters with its precision increases when higher order of Taylor expansion is used. However, the approximation is less accurate if the variance of random effects is large [32]. The expression in equation (10) is applied to obtain the posterior expression for the latent field and hyperparameters.

2.3.1 Posterior approximation with INLA

The posterior distribution of the model hyperparameters and the latent field under the Bayesian framework, using Bayes' theorem, can be obtained from the conditional posterior distribution

$$p(\boldsymbol{\theta}_i,\boldsymbol{\gamma}_i|\boldsymbol{D}_i) = \frac{p(\boldsymbol{D}_i|\boldsymbol{\theta}_i,\boldsymbol{\gamma}_i)p(\boldsymbol{\theta}_i,\boldsymbol{\gamma}_i)}{p(\boldsymbol{D}_i)} \propto p(\boldsymbol{D}_i|\boldsymbol{\theta}_i,\boldsymbol{\gamma}_i)p(\boldsymbol{\gamma}_i|\boldsymbol{\theta}_i)p(\boldsymbol{\theta}_i) \tag{11}$$

where $p(\boldsymbol{\gamma}_i|\boldsymbol{\theta}_i)$ and $p(\boldsymbol{\theta}_i)$ are prior distributions and the focus is on approximating the multidimensional integral from the marginal likelihood $p(\boldsymbol{D}_i|\boldsymbol{\theta}_i,\boldsymbol{\gamma}_i)$ and approximation technique of INLA has been shown to provide exact approximations to the posterior estimates at faster rates than sampling-based methods such as Markov Chain Monte Carlo (MCMC), especially for complex and hierarchical models [33].

We consider the Laplace transformation using a second-order Taylor series expansion for the integral of the density function $p(\boldsymbol{\chi})$ by taking the form of [34].

$$\int_{-\infty}^{\infty} p(\boldsymbol{\gamma})\,d\boldsymbol{\gamma} = \int_{-\infty}^{\infty} \exp(\log p(\boldsymbol{\gamma}))\,d\boldsymbol{\gamma} = \int_{-\infty}^{\infty} \exp\left(g(\boldsymbol{\gamma})\right)d\boldsymbol{\gamma} \tag{12}$$

The integral value is a function of the behaviour in the neighbourhood of the mode of $g(\boldsymbol{\gamma})$ and hence $g(\boldsymbol{\gamma})$ can be replaced by a second-order Taylor approximation of $g(\boldsymbol{\gamma})$ to compute the integral's approximate value.

Let $\boldsymbol{\gamma}^*$ be the global maximum of $\boldsymbol{\gamma}$ given as

$$\boldsymbol{\gamma}^* = \operatorname{argmax}_{\boldsymbol{\gamma}} g(\boldsymbol{\gamma}),$$

then

$$\left.\frac{\partial g(\boldsymbol{\gamma})}{\partial \boldsymbol{\gamma}}\right|_{\boldsymbol{\gamma}=\boldsymbol{\gamma}^*} = 0$$

for $g(\boldsymbol{\gamma})$ to be approximated as

$$g(\boldsymbol{\gamma}) \approx g(\boldsymbol{\gamma}^*) + 0.5(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)'\mathbf{H}g(\boldsymbol{\gamma}^*)(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)$$

where $\mathbf{H}g(\boldsymbol{\gamma}^*)$ is the Hessian of $g(\boldsymbol{\gamma}^*)$, and equation (9) can be written as

$$\int_{-\infty}^{\infty} p(\boldsymbol{\gamma})\,d\boldsymbol{\gamma} = \int_{-\infty}^{\infty} \exp\big(g(\boldsymbol{\gamma}^*) + 0.5(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)'\mathbf{H}g(\boldsymbol{\gamma}^*)(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)\big)\,d\boldsymbol{\gamma}$$

$$= \exp\big(g(\boldsymbol{\gamma}^*)\big)\int_{-\infty}^{\infty} \exp\big(0.5(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)'\mathbf{H}g(\boldsymbol{\gamma}^*)(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)\big)\,d\boldsymbol{\gamma}$$

$$= \exp\big(g(\boldsymbol{\gamma}^*)\big)\int_{-\infty}^{\infty} \exp\big(-0.5(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)'\{-\mathbf{H}g(\boldsymbol{\gamma}^*)\}(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)\big)\,d\boldsymbol{\gamma}$$

$$= \exp\big(g(\boldsymbol{\gamma}^*)\big)\,(2\pi)^{\frac{nm}{2}}|\mathbf{H}g(\boldsymbol{\gamma}^*)|^{-\frac{1}{2}} \times$$

$$\int_{-\infty}^{\infty} (2\pi)^{-\frac{nm}{2}}|\mathbf{H}g(\boldsymbol{\gamma}^*)|^{-\frac{1}{2}} \exp\big(-0.5(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)'\{-\mathbf{H}g(\boldsymbol{\gamma}^*)\}(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)\big)\,d\boldsymbol{\gamma}$$

The integral follows multivariate Normal and putting $-\mathbf{H}g(\boldsymbol{\gamma}^*) = \boldsymbol{Q}(\boldsymbol{\gamma}^*)$, the precision matrix for the random vector $\boldsymbol{\gamma}^*$ yields

$$\int_{-\infty}^{\infty} p(\boldsymbol{\gamma})\,d\boldsymbol{\gamma} \approx \exp\big(g(\boldsymbol{\gamma}^*)\big)\,(2\pi)^{\frac{nm}{2}}|\mathbf{H}g(\boldsymbol{\gamma}^*)|^{-\frac{1}{2}} \times$$

$$\int_{-\infty}^{\infty} (2\pi)^{-\frac{nm}{2}}|\boldsymbol{Q}(\boldsymbol{\gamma}^*)|^{-\frac{1}{2}} \exp\big(-0.5(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)'\boldsymbol{Q}(\boldsymbol{\gamma}^*)(\boldsymbol{\gamma}-\boldsymbol{\gamma}^*)\big)\,d\boldsymbol{\gamma}$$

$$\approx (2\pi)^{\frac{nm}{2}}|\boldsymbol{Q}(\boldsymbol{\gamma}^*)|^{-\frac{1}{2}}\exp\big(g(\boldsymbol{\gamma}^*)\big).$$

The posterior distribution of $p(\boldsymbol{\gamma},\boldsymbol{\theta}|\boldsymbol{D})$ is obtained from the joint posterior distribution as

$$p(\boldsymbol{\gamma},\boldsymbol{\theta}|\boldsymbol{D}) \propto p(\boldsymbol{\theta})|\boldsymbol{Q}(\boldsymbol{\theta})|^{\frac{1}{2}}\exp\left(-\frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{\gamma} + \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} \log p(\boldsymbol{D}_i|\boldsymbol{\gamma}_i,\boldsymbol{\theta})\right)$$

which can be rewritten as, ignoring elements with $\boldsymbol{\gamma}$

$$p(\boldsymbol{\gamma}|\boldsymbol{\theta},\boldsymbol{D}) \propto \exp\left(-\frac{1}{2}\boldsymbol{\gamma}'\boldsymbol{Q}(\boldsymbol{\theta})\boldsymbol{\gamma} + \sum_{\boldsymbol{i}=\mathbf{1}}^{\boldsymbol{n}} g_i(\boldsymbol{\gamma}_i)\right) \tag{13}$$

Gaussian approximation

The Gaussian approximation of Equation (13), $p_G(\boldsymbol{\gamma}|\boldsymbol{\theta},\boldsymbol{D})$ is obtained by combining the mode and the curvature at the mode of $p(\boldsymbol{\gamma}|\boldsymbol{\theta},\boldsymbol{D})$. A Newton-Raphson method is used to calculate the mode in iteration. Let $\mu^{(0)}$ be the mode's initial value, and expand $g_i(\boldsymbol{\gamma}_i)$ around $\boldsymbol{\mu}_i^{(0)} = \left(\mu_{i1}^{(0)}, \dots, \mu_{iN}^{(0)}\right)$ to the second order Taylor expansion,

$$g_i(\boldsymbol{\gamma}_i) \approx g_i\left(\boldsymbol{\mu}_i^{(0)}\right) + \boldsymbol{b}_i'\boldsymbol{\gamma}_i - \frac{1}{2}\boldsymbol{c}_i'\boldsymbol{\gamma}_i'\boldsymbol{\gamma}_i \tag{14}$$

where $\boldsymbol{b}_i$ and $\boldsymbol{c}_i$ depend on $\boldsymbol{\mu}^{(0)}$. Substituting equation (14) into equation (13) yields

$$p_G(\boldsymbol{\gamma}|\boldsymbol{\theta},\boldsymbol{D}) \approx g_i\left(\boldsymbol{\mu}_i^{(0)}\right)\exp\left(-\frac{1}{2}\boldsymbol{\gamma}'(\boldsymbol{Q}+\boldsymbol{c})\boldsymbol{\gamma}+\boldsymbol{b}'\boldsymbol{\gamma}\right)$$
$$\propto \exp\left(-\frac{1}{2}\boldsymbol{\gamma}'(\boldsymbol{Q}+\boldsymbol{c})\boldsymbol{\gamma}+\boldsymbol{b}'\boldsymbol{\gamma}\right).$$

A Gaussian approximation of $p_G(\boldsymbol{\gamma}|\boldsymbol{\theta},\boldsymbol{D})$ is arrived at with precision matrix $(\boldsymbol{Q}+\mathrm{diag}(\boldsymbol{c}))$ and mode $\boldsymbol{\mu}^{(1)}$, which gives the solution of $\left(\boldsymbol{Q}+\mathrm{diag}(\boldsymbol{c})\right)\boldsymbol{\mu}^{(1)}=\boldsymbol{b}$. The procedure is then iterated, with $\boldsymbol{\mu}^{(1)}$as the new starting value, until it converges to a Gaussian distribution with, say, mean $\boldsymbol{\mu}^{(j)}\to\boldsymbol{\mu}^{(*)}=\boldsymbol{\gamma}^{*}$ and precision matrix $\boldsymbol{Q}^{(j)}\to\boldsymbol{Q}^{(*)}=\boldsymbol{Q}+\mathrm{diag}(\boldsymbol{c}^{*}), j=1,2,\dots$, using a suitable convergence criterion.

At this point, the obtained approximation will be [35]:

$$p_G(\boldsymbol{\gamma}|\boldsymbol{\theta},\boldsymbol{D}) \propto \exp\left(-\frac{1}{2}\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}^{*}(\boldsymbol{\theta})\right)'\left(\boldsymbol{Q}(\boldsymbol{\theta})+\mathrm{diag}(\boldsymbol{c})\right)\left(\boldsymbol{\gamma}-\boldsymbol{\gamma}^{*}(\boldsymbol{\theta})\right)\right),$$

where $\boldsymbol{c}$ is the 2nd-order term in the Taylor expansion of $\sum_{\boldsymbol{i}=\boldsymbol{1}}^{\boldsymbol{n}}\log p(\boldsymbol{D}_i|\boldsymbol{\gamma}_i,\boldsymbol{\theta})$ at modal value $\boldsymbol{\gamma}^{*}(\boldsymbol{\theta})$.

For the marginal posterior conditional distribution $p(\boldsymbol{\gamma}_i|\boldsymbol{\theta},\boldsymbol{D})$ included in the computation of the marginal posterior $p(\boldsymbol{\gamma}_i|\boldsymbol{D})$, [36] discussed three approximations $\tilde{p}(\gamma_i|\boldsymbol{\theta}_k,\boldsymbol{D})$ where $\boldsymbol{\theta}_k$ are weighted points to be used in the integration, the approximations are Gaussian, full Laplace, and simplified Laplace approximation. The Gaussian approximation has been noted to generally not be the best if the true density of $p(\boldsymbol{\gamma}_i|\boldsymbol{\theta},\boldsymbol{D})$ is not symmetric, hence, the full Laplace approximation is a correction of Gaussian approximation and accurate but at a very expensive computational cost. However, the simplified Laplace approximation which is based on the Taylor series expansion of the full Laplace approximation is modestly accurate in most real application settings [34].

If the mean of $\boldsymbol{\gamma}$ is $\boldsymbol{\mu}$, the density of $\boldsymbol{\gamma}$ is

$$p(\boldsymbol{\gamma})=(2\pi)^{-n/2}|\boldsymbol{Q}|^{1/2}\exp\left[-\frac{1}{2}(\boldsymbol{\gamma}-\boldsymbol{\mu})^{T}\boldsymbol{Q}(\boldsymbol{\gamma}-\boldsymbol{\mu})\right] \tag{15}$$

The sparse matrix $\boldsymbol{Q}$ is factorized as Cholesky triangle product $\boldsymbol{L}\boldsymbol{L}^{T}$, and only non-zero terms are computed due to the Markov property and $L_{ji}=0$. Let $\boldsymbol{L}^{T}\boldsymbol{\gamma}=\boldsymbol{r}$ where $\boldsymbol{r}\sim N(\boldsymbol{0},\boldsymbol{1})$**,** then we have that $L_{ii}\chi_i=r_i-\sum_{k=i+1}^{n}L_{ki}\chi_k$ for $i=n,\dots,1$. Multiplying each side with $\chi_j, j\geq i$ and taking the expectation yields the recursion

$$\Sigma_{ij}=\frac{\partial_{ij}^{2}}{L_{ii}}-\frac{1}{L_{ii}}\sum_{k=i+1}^{n}L_{ki}\Sigma_{kj}\qquad j\geq i,\quad i=n,\dots,1$$

where $\boldsymbol{\Sigma}=\boldsymbol{Q}^{-1}$ is the covariance matrix and $\partial_{ij}=1$ if $i=j$ and $\partial_{ij}=0$ otherwise. These recursion results in Gaussian approximations $\tilde{p}_G(\boldsymbol{\gamma}|\boldsymbol{\theta},\boldsymbol{D})$ with mean $\mu_i(\boldsymbol{\theta})$ and marginal variance $\sigma_i^2(\boldsymbol{\theta})$.

Results and discussion

Performance of designed-based GLMM in different sample sizes using a simulation study

To compare our model to other existing models, we shall examine their performances under different sample sizes and variable random effect by simulating data set in the similitude of the design used for the 2021 NMIS, where the stratification variable was the place of residence with two strata, namely urban and rural. PSU are the EA clusters; samples were chosen separately in each stratum using a two-stage selection process. A predetermined number of households were chosen in each cluster using equal probability systematic sampling in the second step of selection, while EAs were chosen in the first stage with a probability proportional to their size.

To start, we run data simulations for four predictor factors: *ea* (for clusters the number of subject), *hh* (identifying the household in each cluster), *slpnet* (if children sleep under a mosquito net: no or yes), *resid* (place of residence stratification). These factors are crucial indicators to the incidence of malaria fever as have been captured in the 2021 NMIS report. The variable of interest is the event of fever denoted as y, (if the child had a fever, yes = 1 or no = 0) as a result of mosquito was simulated using a random number generator for the binomial distribution. We set the probability for the binomial distribution to vary based on the place of residence of respondents and whether they slept under mosquito nets or not. We set the probabilities thus; rural/no: p = 0.75, rural/yes: p = 0.47, urban/no: p = 0.65, urban/yes: p = 0.39. These are the fixed effects as we assume the effects of place of residence and sleeping under the mosquito net and their interaction are fixed. The effect of sleeping under mosquito net decreases in rural areas by a probability of 0.38 and decreases in urban areas by a probability of 0.26.

We assume that observation in the same cluster will be correlated thus we have a random effect for each EA, hence we have a mixed effects approach. A normal distribution with a mean of 0 and a standard deviation of 0.012 is used to generate *n* random samples from which we include random effect probabilities and use the probabilities to generate the response variables with the model

$$Prob(y = 1) = \text{logit}^{-1}(\beta_0 + \beta_1 \text{slpnet} - \beta_2 \text{resid} - \beta_3 slpnet * resid + u * \text{z})$$

where

- β_0 = log-odds of living in a rural area and not sleeping under a mosquito net
- β_1 = difference in log-odds between living in a rural area and sleeping under a mosquito net and living in a rural area, not sleeping under a mosquito net
- β_2 = difference in log-odds between living in an urban area, not sleeping under a mosquito net, and living in a rural area, not sleeping under a mosquito net
- β_3 = difference in log-odds for urban residents sleeping under mosquito net effect versus the rural residents sleeping under mosquito net effect
- $u \sim N(0, (\tau_{ea}{}^{-1})^2)$ is random effect probability for each household

The fixed effects values were obtained by transforming the probabilities to log-odds and obtained the fixed values as β_0 = 1.735, β_1 = -1.855, β_2 = -1.116, β_3 = 0.788 and $\tau_{ea}{}^{-1} = 0.012$. We simulate population sizes of N = 150, 500, 2000, 5000, and 10000. In order to oversample extreme random effect parameters to bias random-intercept variance we follow the specification of [37] pairwise sampling probabilities from sampling probabilities at each stage in calculating the design weights for each household given the values of N and resulted in the sample sizes n respectively as n = 90, 340, 1230, 2920 and 5965.

The simulated dataset is combined with all the survey design information including the primary sampling unit, the enumeration area, the household number, and design weights using the R *survey* package [23]. The purpose of this is that the necessary design information is bonded to the dataset for accurate analytical adjustments to be carried out automatically and dependably. Also, it helps to provide valid variance estimates for statistics computed on the *survey* objects. The models applied to the simulated dataset were the model-based GLMM with Adaptive Gauss-Hermite Quadrature (AGHQ) approximation, weighted Adaptive Gauss-Hermite Quadrature (wtd AGHQ) approximation, GLMM with MCMC and the design-based GLMM with INLA as presented in this study. The models were fitted for each value of N and the corresponding n values and presented the tables that follow with the parameters true and estimated values and accuracy measures such as BIAS, mean absolute error (MAE), mean squared error (MSE) and root mean squared error (RMSE).

Table 1: Comparison of models estimates for simulated data with N = 150

			N = 150			
AGHQ	true value	prediction	MSE	RMSE	MAE	BIAS
beta0	1.734601	0.863636	0.758580	0.870965	0.870965	0.870965
beta1	-1.854745	0.346154	4.843957	2.200899	2.200899	-2.200899

beta2	-1.115562	0.703704	3.309727	1.819266	1.819266	-1.819266
beta3	0.788394	0.533333	0.065056	0.255060	0.255060	0.255060
$\tau_{realpsu}$	0.012000	0.000000	0.000144	0.012000	0.012000	0.012000
wtd AGHQ						
beta0	1.734601	0.928108	0.650431	0.806493	0.806493	0.806493
beta1	-1.854745	0.367825	4.939819	2.222570	2.222570	-2.222570
beta2	-1.115562	0.711649	3.338698	1.827211	1.827211	-1.827211
beta3	0.788394	0.676472	0.012527	0.111922	0.111922	0.111922
$\tau_{realpsu}$	0.012000	1.596300	2.510006	1.584300	1.584300	-1.584300
MCMC						
beta0	1.734601	2.429484	0.482863	0.694883	0.694883	-0.694883
beta1	-1.854745	-3.365495	2.282366	1.510750	1.510750	1.510750
beta2	-1.115562	-1.350667	0.055274	0.235105	0.235105	0.235105
beta3	0.788394	2.608121	3.311406	1.819727	1.819727	-1.819727
$\tau_{realpsu}$	0.012000	0.346218	0.111702	0.334218	0.334218	-0.334218
INLA						
beta0	1.734601	1.977682	0.059088	0.243081	0.243081	-0.243081
beta1	-1.854745	-2.638078	0.613610	0.783333	0.783333	0.783333
beta2	-1.115562	-1.077267	0.001467	0.038295	0.038295	-0.038295
beta3	0.788394	1.879115	1.189673	1.090721	1.090721	-1.090721
$\tau_{realpsu}$	0.012000	0.012207	0.000000	0.000207	0.000207	-0.000207

Table 2: Comparisons of models estimates for simulated data with $N = 500$

			N = 500			
AGHQ	true value	prediction	MSE	RMSE	MAE	BIAS
beta0	1.734601	0.855422	0.772957	0.879179	0.879179	0.879179
beta1	-1.854745	0.395062	5.061632	2.249807	2.249807	-2.249807
beta2	-1.115562	0.725490	3.389473	1.841052	1.841052	-1.841052
beta3	0.788394	0.297297	0.241176	0.491097	0.491097	0.491097

$\tau_{realpsu}$	0.012000	0.000000	0.000144	0.012000	0.012000	0.012000
wtd AGHQ						
beta0	1.734601	0.891870	0.710196	0.842731	0.842731	0.842731
beta1	-1.854745	0.391893	5.047386	2.246639	2.246639	-2.246639
beta2	-1.115562	0.775975	3.577912	1.891537	1.891537	-1.891537
beta3	0.788394	0.256024	0.283417	0.532369	0.532369	0.532369
$\tau_{realpsu}$	0.012000	1.038400	1.053497	1.026400	1.026400	-1.026400
MCMC						
beta0	1.734601	2.259958	0.276000	0.525357	0.525357	-0.525357
beta1	-1.854745	-2.781554	0.858974	0.926809	0.926809	0.926809
beta2	-1.115562	-1.121129	0.000031	0.005568	0.005568	0.005568
beta3	0.788394	0.690650	0.009554	0.097743	0.097743	0.097743
$\tau_{realpsu}$	0.012000	0.055756	0.001915	0.043756	0.043756	-0.043756
INLA						
beta0	1.734601	1.811701	0.005944	0.077100	0.077100	-0.077100
beta1	-1.854745	-2.243014	0.150753	0.388269	0.388269	0.388269
beta2	-1.115562	-0.828733	0.082271	0.286829	0.286829	-0.286829
beta3	0.788394	0.386822	0.161260	0.401571	0.401571	0.401571
$\tau_{realpsu}$	0.012000	0.012139	0.000000	0.000139	0.000139	-0.000139

Table 3: Comparisons of models estimates for simulated data with *N* = 2000

			N = 2000			
AGHQ	true value	prediction	MSE	RMSE	MAE	BIAS
beta0	1.734601	0.883206	0.724874	0.851396	0.851396	0.851396
beta1	-1.854745	0.479626	5.449288	2.334371	2.334371	-2.334371
beta2	-1.115562	0.646955	3.106466	1.762517	1.762517	-1.762517
beta3	0.788394	0.376625	0.169553	0.411769	0.411769	0.411769
$\tau_{realpsu}$	0.012000	0.000000	0.000144	0.012000	0.012000	0.012000
wtd AGHQ						
beta0	1.734601	0.943839	0.625304	0.790762	0.790762	0.790762
beta1	-1.854745	0.471910	5.413323	2.326655	2.326655	-2.326655

beta2	-1.115562	0.703921	3.310519	1.819483	1.819483	-1.819483
beta3	0.788394	0.314330	0.224737	0.474064	0.474064	0.474064
$\tau_{realpsu}$	0.012000	1.517900	2.267735	1.505900	1.505900	-1.505900
MCMC						
beta0	1.734601	2.420892	0.470995	0.686291	0.686291	-0.686291
beta1	-1.854745	-2.558532	0.495316	0.703787	0.703787	0.703787
beta2	-1.115562	-1.667474	0.304607	0.551912	0.551912	0.551912
beta3	0.788394	1.204320	0.172995	0.415927	0.415927	-0.415927
$\tau_{realpsu}$	0.012000	0.098815	0.007537	0.086815	0.086815	-0.086815
INLA						
beta0	1.734601	2.008520	0.075031	0.273919	0.273919	-0.273919
beta1	-1.854745	-2.087734	0.054284	0.232989	0.232989	0.232989
beta2	-1.115562	-1.405468	0.084046	0.289906	0.289906	0.289906
beta3	0.788394	0.998801	0.044271	0.210407	0.210407	-0.210407
$\tau_{realpsu}$	0.012000	0.012408	0.000000	0.000408	0.000408	-0.000408

Table 4: Comparisons of models estimates for simulated data with $N = 5000$

N = 5000						
AGHQ	true value	prediction	MSE	RMSE	MAE	BIAS
beta0	1.734601	0.856948	0.770274	0.877653	0.877653	0.877653
beta1	-1.854745	0.475366	5.429420	2.330112	2.330112	-2.330112
beta2	-1.115562	0.660811	3.155500	1.776373	1.776373	-1.776373
beta3	0.788394	0.359712	0.183768	0.428682	0.428682	0.428682
$\tau_{realpsu}$	0.012000	0.000000	0.000144	0.012000	0.012000	0.012000
wtd AGHQ						
beta0	1.734601	0.913252	0.674614	0.821349	0.821349	0.821349
beta1	-1.854745	0.478917	5.445981	2.333663	2.333663	-2.333663
beta2	-1.115562	0.684563	3.240448	1.800124	1.800124	-1.800124

beta3	0.788394	0.322673	0.216896	0.465721	0.465721	0.465721
$\tau_{realpsu}$	0.012000	1.253000	1.540081	1.241000	1.241000	-1.241000
MCMC						
beta0	1.734601	2.403566	0.447514	0.668965	0.668965	-0.668965
beta1	-1.854745	-2.529783	0.455676	0.675037	0.675037	0.675037
beta2	-1.115562	-1.497472	0.145855	0.381910	0.381910	0.381910
beta3	0.788394	0.842585	0.002937	0.054191	0.054191	-0.054191
$\tau_{realpsu}$	0.012000	0.006057	0.000035	0.005943	0.005943	0.005943
INLA						
beta0	1.734601	1.794182	0.003550	0.059580	0.059580	-0.059580
beta1	-1.854745	-1.892915	0.001457	0.038170	0.038170	0.038170
beta2	-1.115562	-1.126270	0.000115	0.010708	0.010708	0.010708
beta3	0.788394	0.647476	0.019858	0.140918	0.140918	0.140918
$\tau_{realpsu}$	0.012000	0.012162	0.000000	0.000162	0.000162	-0.000162

Table 5: Comparisons of models estimates for simulated data with $N = 10000$

N = 10000						
AGHQ	true value	prediction	MSE	RMSE	MAE	BIAS
beta0	1.734601	0.835958	0.807560	0.898643	0.898643	0.898643
beta1	-1.854745	0.456093	5.339972	2.310838	2.310838	-2.310838
beta2	-1.115562	0.657365	3.143271	1.772927	1.772927	-1.772927
beta3	0.788394	0.396925	0.153248	0.391469	0.391469	0.391469
$\tau_{realpsu}$	0.012000	0.000000	0.000144	0.012000	0.012000	0.012000
wtd AGHQ						
beta0	1.734601	0.881944	0.727023	0.852657	0.852657	0.852657
beta1	-1.854745	0.454100	5.330766	2.308845	2.308845	-2.308845
beta2	-1.115562	0.697352	3.286657	1.812914	1.812914	-1.812914
beta3	0.788394	0.377764	0.168617	0.410630	0.410630	0.410630
$\tau_{realpsu}$	0.012000	1.214500	1.446006	1.202500	1.202500	-1.202500

MCMC						
beta0	1.734601	1.988201	0.064313	0.253600	0.253600	-0.253600
beta1	-1.854745	-2.227099	0.138647	0.372353	0.372353	0.372353
beta2	-1.115562	-1.163398	0.002288	0.047837	0.047837	0.047837
beta3	0.788394	0.880978	0.008572	0.092585	0.092585	-0.092585
$\tau_{realpsu}$	0.012000	0.040822	0.000831	0.028822	0.028822	-0.028822
INLA						
beta0	1.734601	1.630116	0.010917	0.104486	0.104486	0.104486
beta1	-1.854745	-1.806320	0.002345	0.048426	0.048426	-0.048426
beta2	-1.115562	-0.978050	0.018910	0.137512	0.137512	-0.137512
beta3	0.788394	0.735635	0.002783	0.052758	0.052758	0.052758
$\tau_{realpsu}$	0.012000	0.011993	0.000000	0.000007	0.000007	0.000007

From Tables 1 to 5 we observe that as the sample size increased, our model with INLA gave the best estimates of the intercept and random effect parameters in all sample sizes considered as we see the model had the least values of all accuracy measures. The next to follow was the AGHQ approximation, then the MCMC model, while the *wtd* AGHQ did not perform well in the random effect estimation for the simulated dataset. For the interaction parameter, wtd AGHQ had the best estimate for N = 150, MCMC had the best estimate for N = 500 and 5000, INLA had the best for N = 2000 and 10000. For both main effect parameters INLA had the best estimate for N = 150, for N = 500 both MCMC and INLA had the best estimates for beta2 and beta1 respectively, while for N = 2000, 5000, and 10000, INLA had the best estimates for both main effect parameters.

The summary of the simulation studies showed that the Bayesian approach of MCMC and INLA, overall gave better or close parameters' estimates for the simulated data, while the INLA approach with design information was mostly preferred over the other methods, meaning that other methods were able to also fit GLMM without design information, the INLA approach presented in this study utilizes the design information to get better parameter estimates, as reported in the simulation output tables. The preference of the INLA GLMM over other MCMC (Bayesian approaches performed better) stems from the ease for which the posterior (marginal) is evaluated as against the joint posterior in MCMC.

Comparing all the models together to see how the simulated dataset favoured each of the models we use the Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), Deviance information criterion (DIC), deviance, log-likelihood and Widely Applicable Information Criterion (WAIC). AIC compares the quality of different statistical models for a given dataset, balancing how well a model fits the data against its complexity (number of parameters), a lower AIC value indicates a better model fit relative to its complexity. BIC selects the best model among a set of models by considering both how well it fits the likelihood and its number of parameters, with lower BIC values indicating a better model fit while penalizing for excessive complexity. DIC is a Bayesian equivalent of AIC that compares different statistical models by balancing their goodness of fit against their complexity, and choosing the model that best explains the data without being overly complex; it is and is particularly useful when analysing models with hierarchical structures. Deviance is a measure of error, and a lower deviance indicates a better fit. Log-likelihood represents the probability of observing the data given the model's parameters. A higher log-likelihood value indicates a better fit between the model and the observed data. WAIC is a generalization of the AIC and it's an estimate of out-of-sample relative K-L divergence (KLD), while a smaller WAIC value indicates a better model.

The comparison of the models was not very satisfying owing to the fact that different model comparison measures are used for models under classical and Bayesian approaches. Even for Bayesian approaches such as MCMC and INLA, the model comparison measures are not comparable except for DIC. MCMC outputs deviance measures as do classical GLMM, while INLA outputs log-likelihood as do classical GLMM. We present the results of these measures for the simulated dataset in the tables that follow.

Table 6: Comparisons of models fit for simulated dataset (N = 150, 500, 2000)

	N = 150			
	AGHQ	*wtd AGHQ*	*MCMC*	*INLA*
AIC	114.610	747.103	-	-
BIC	127.109	759.602	-	-
logLik	-52.305	-368.552	-	-64.7212
Deviance	104.610	673.240	97.251	NA
DIC	-	-	109.992	112.3667
WAIC	-	-	-	112.7762
	N = 500			
AIC	397.240	2896.932	-	-
BIC	416.385	2916.076	-	-
logLik	-193.620	-1443.466	-	-208.669
Deviance	387.240	2689.856	334.376	-
DIC	-	-	395.880	395.1731
WAIC	-	-	-	395.2918
N = 2000				
AIC	1460.486	9986.509	-	-
BIC	1486.060	10012.083	-	-
logLik	-725.243	-4988.254	-	-744.436
Deviance	1375.119	9104.577	1237.850	-
DIC	-	-	1448.474	1461.383
WAIC	-	-	-	1461.462

Table 7: Comparisons of models fit for simulated dataset (N = 5000, 10000)

N = 5000				
	AGHQ	*wtd AGHQ*	*MCMC*	*INLA*
AIC	3507.801	25065.660	-	-

BIC	3537.698	25095.560	-	-
logLik	-1748.901	-12527.830	-	-1768.39
Deviance	3497.801	23145.370	2772.385	-
DIC	-	-	3461.351	3505.714
WAIC	-	-	-	3505.81
N = 10000				
AIC	7299.027	52661.430	-	-
BIC	7332.496	52694.900	-	-
logLik	-3644.514	-26325.710	-	-3665.51
Deviance	7289.027	48787.940	6148.140	-
DIC	-	-	7245.793	7296.893
WAIC	-	-	-	7297.05

7From Table 6 and Table 7, we see that the log-likelihood for INLA was near the values of those of AGHQ and wtd AGHQ for all the *N* values simulated, while the DIC values of INLA and MCMC were also close for the *N* values simulated, with MCMC being slightly lower in values than INLA. Comparing the deviance values of the classical approaches to MCMC showed that MCMC had lower deviance values. Then for the classical model-based approaches, the AGHQ was preferred to the wtd AGHQ for all values of *N* simulated AGHQ had far lower values than wtd AGHQ.

Hence from the simulation study, we saw that our INLA GLMM performed satisfactorily in estimating the model true parameters for all sample sizes, and there were close estimates with MCMC and model-based AGHQ in certain instances of coefficients, but overall in estimating the main effects and interaction and intercept terms given the INLA GLMM performed better with lowest error term values. However, due to the differences in classical and Bayesian estimations, comparing both methods are not easy due to no common diagnostics for both, even for Bayesian methods DIC was the only common diagnostics for comparison while WAIC requires complicated computation to obtain from MCMC algorithms. The overlapping measures of log-likelihood and deviance for AGHQ with INLA and MCMC respectively, meant that we could see how both Bayesian methods compared with AGHQ.

GLMMs on 2021 NMIS with INLA inference.

We here perform the Bayesian generalised linear mixed modelling on the two-stage survey design with the binary outcome on the NMIS 2021 data using the integrated Laplace approximation with the response variable being “if children had a fever in last two weeks” and predictors such as the number of children in household, if children slept under a mosquito net last night, children’s age in months, education level of the mother, wealth index of household and type of place of residence. The implementation here considers the design in section 3.1, for the evaluation of the model, making it fall into the category of design-based approach. The stratification variable is the place of residence with two values, “0” and “1” with level labels “urban” and “rural” respectively. The levels of clustering are defined by the enumeration area (EA) and the Household (HH).

The specification is such that the likelihood used for the observations is the binomial distribution and we prefer to specify the priors on the hyperparamters in the likelihood as logit priors and the penalised complexity prior for precision parameter. The summary of the resulting model specification showing the posterior estimates of the model parameters is shown in Table 8.

Table 8: Posterior Estimate of GLMM using INLA

Fixed effect	mean	sd	2.5% CI	97.5% CI	mode
(Intercept)	-0.5000	0.1370	-0.7680	-0.2320	-0.4990
sleep under mosq. Net_ALL	0.0060	0.0710	-0.1330	0.1450	0.0060
sleep under mosq. Net_Some	-0.1190	0.0890	-0.2920	0.0550	-0.1190
sleep under mosq. Net_No net	-0.2420	0.0710	-0.3810	-0.1020	-0.2420
number of children in HH	-0.0510	0.0200	-0.0900	-0.0120	-0.0510
child_age_month	-0.0030	0.0010	-0.0050	0.0000	-0.0030
Education level_Primary	0.2250	0.0760	0.0750	0.3750	0.2250
Education level_Secondary	0.0820	0.0770	-0.0680	0.2330	0.0820
Education level_Higher	-0.2390	0.1120	-0.4590	-0.0180	-0.2390
Wealth index_Poorer	-0.0330	0.0780	-0.1850	0.1200	-0.0330
Wealth index_Middle	-0.1230	0.0880	-0.2960	0.0500	-0.1230
Wealth index_Richer	-0.2440	0.1020	-0.4450	-0.0440	-0.2440
Wealth index_Richest	-0.3320	0.1200	-0.5670	-0.0970	-0.3320
Type of place of residence_Rural	0.2570	0.0980	0.0650	0.4500	0.2570
Random effect					
Precision for realpsu	1.3700	0.1290	1.1400	1.6400	1.3500

The results from Table 8, showed that sleeping under mosquito net, education level of mothers, wealth index and place of residence were all significant indicators of malaria incidence among children. With higher incidence in children who live around rural areas compared to those in urban areas as reported by the marginal posterior estimates value of 0.26 for factor type of place of residence. Also, the incidence of malaria fever weakened as wealth index of family increased, similar was the case for level of education of mothers of children.

The model selection criteria obtained for the INLA model, for which we represent as *INLAglmm*, were Watanabe-Akaike information criterion (WAIC) of 12374.65, DIC of 12383.52, marginal log-Likelihood of -6425.06 and deviance of 12266.0. The marginal posterior mean effect of number of children under 5 years in household, age of child in months and place of residence, either urban or rural area and the posterior marginal are seen to Gaussian in distribution.

The trace and density plots of the model parameters as well as the summary plot of the variance component associated with primary sampling unit and the residual precision component fitted and linear predictor in Figure 1 shows that the residual precision far from zero and the parameter estimates are in line with specifications.

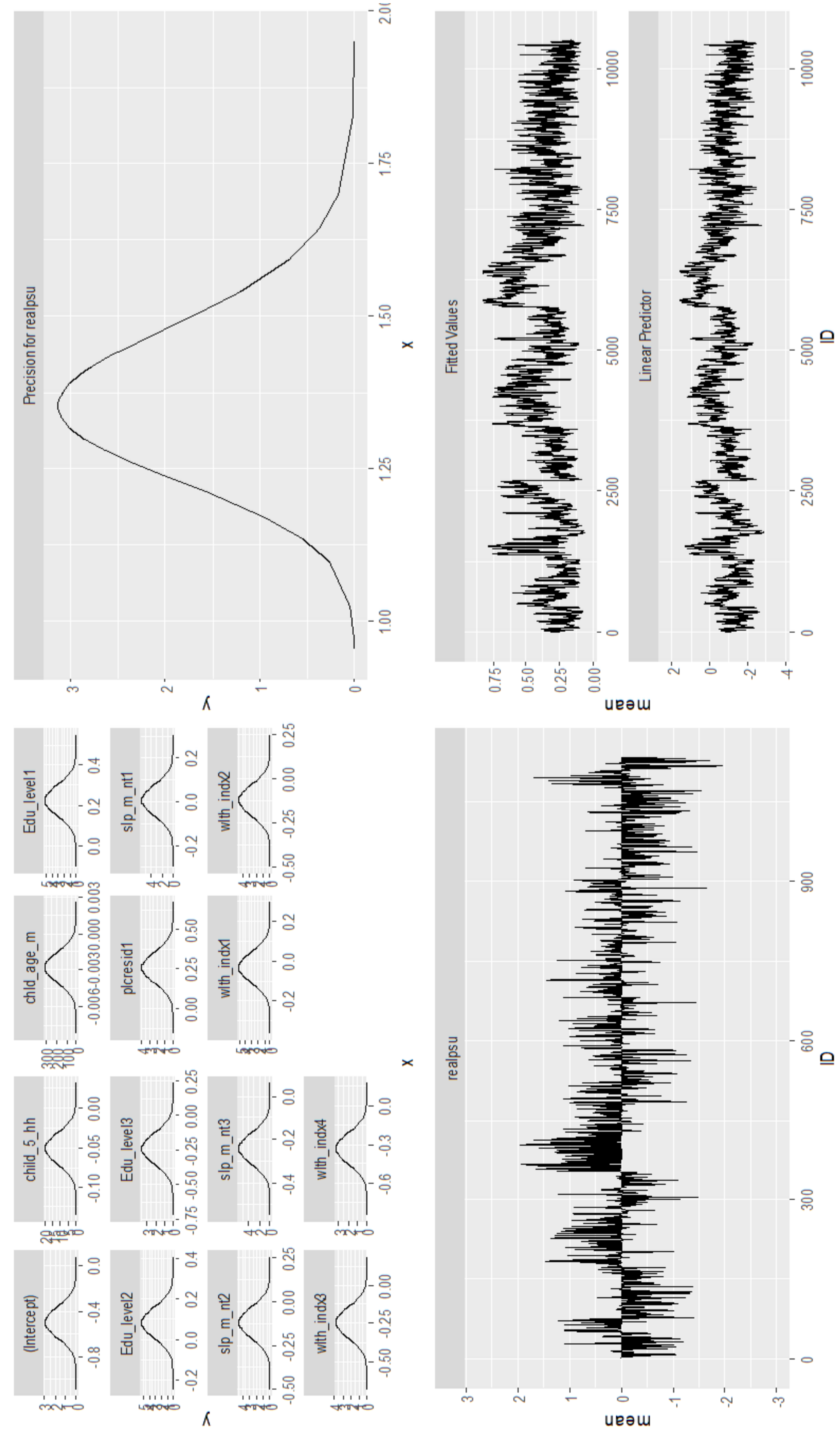


Figure 1: Summary plot for the coefficients from INLA GLMM

Comparing the INLAglmm with other models for the 2021 NMIS data.

Comparing the models from the classical and Bayesian approaches is not exactly straightforward since different model diagnostics are produced by the different models and also the specifications of the models are different. We shall without ignoring the flaws, attempt to put together the model comparison measures to see how the different models fit the Nigeria Malaria Indicator Survey 2021. The model comparison measures are given as produced by the model fitting process and shown in Table 9.

Table 9: Model Comparison Measures

	AIC	BIC	logLik	deviance	DIC	WAIC
AGHQ	12684	12793	-6327.10	12654.00		
wtdAGHQ	12296	12405	-6333.00	12266.00		
INLAglmm			-6425.06		12383.76	12374.65
MCMCglmm				12091.12	12388.94	

From Table 9, we can see for the classical case that the weighted approach, wtdAGHQ was better than the model-based approach AGHQ, since its AIC, BIC, log-likelihood, and deviance are lower than those of the latter. Comparing the log-likelihoods of the modelling approaches shows that *INLAglmm* had the lowest value. We also noted that the time taken to run the models was the least for the INLA model and up to minutes for the classical approaches.

In summary, the results of the simulation study and application to real-world data [38] emphasize the impact of Bayesian methodology on a wide range of application areas and under various real-life scenarios of differing populations and sampling sizes. The contrast made with existing approaches of model-based analysis and MCMC with survey weights and INLA shows that the methodology for GLMM presented here can be attractive in survey data analysis. The selling point of the INLA approach is its computational efficiency. The marginal posterior is evaluated from the sparse precision matrix of the latent field as discussed in section 2, which in turn leads to quicker computations and hence improves the computation time for complex models and large data as usually found in survey data. While several works have highlighted the merits of INLA as an alternative to MCMC, ([36]; [34]; and [39]), our work here brings to the forefront the methodology of GLMM for complex survey design as a latent Gaussian model for INLA estimation. The impact of complex survey information in data analysis has already been highlighted in the literature, where it was also shown that design-based analysis for GLMM is scarce, hence the significance of this study. The practical implication of this study is the application of the Bayesian approach for complex survey data and specifically the INLA approach for efficiency in the face of complex hierarchical models and large datasets.

Conclusion

While there are solutions for analyses of survey data with survey information for multiple regression and generalized linear model, there are scarce resources on GLMM and this study sought to fill this gap and also shed more light on the line of application of the approximate Bayesian inference of INLA. Our modelling approach was motivated by the Nigeria Malaria Indicator Survey 2021 dataset and the variable of interest was a binary outcome of the incidence of fever, hence we examined a binary two-stage survey GLMM. Our simulation study compared the performance of different GLMMs for different sample sizes and the effect of prior specifications under INLA method for binary two stage survey data. We saw that our INLA approach performed better in estimating model parameters and precision of random effect, compared to MCMC and AGHQ approximations, although there were settings where their estimates were close. We note the significant effect of survey design in the results from analysis of simulations.

Classical GLMM using Adaptive Gauss-Hermite Quadrature approximation (AGHQ) and weighted AGHQ approximation, and Bayesian MCMC and our design-based INLA were applied to the 2021 NMIS dataset. We saw from the study that for the classical case that the weighted approach, wtdAGHQ was better than the model-based approach AGHQ, since its AIC, BIC, log-likelihood and deviance are lower than those of the latter. The INLA approach was preferred when compared using log-likelihood over the classical approaches and over the MCMC approach comparing their DIC. We also noted that the time taken to run the models was least for the INLA model and up to minutes for the classical approaches, which is of great merit especially when analysing large datasets from survey studies. We note that the design-based approach for GLMM binary outcome and using INLA is therefore advocated especially in analyses of survey data.

Despite the potential of the INLA method for survey data analysis, there are yet some limitations to the approach. Its computational cost is exponential concerning the number of hyperparameters $\boldsymbol{\theta}$, and due to its reliance on large sparse matrix computations for the posterior approximation, with a large number of hyperparameters, INLA will require a large amount of memory for such large models. Majorly, INLA was developed only for the subclass of structured additive regression models that can be expressed as LGMs, so it cannot be applied to models that cannot be expressed as LGMs. Also, implementing new models with INLA requires a high level of programming and expertise due to the intricate numerical optimization [36]. The subjectivity of prior specification is also a challenge as it is in any Bayesian approach to data analysis. Another noticeable limitation of this study is that the scope was only on a two-stage sampling design involving binomial response variables. Hence, future research can extend the distribution of response variables to other members of the exponential family of distributions.

Reference

[1] Guan, Y. and Haran, M. A (2018). A Computationally Efficient Projection- Based Approach for Spatial Generalized Linear Mixed Models. *Journal of Computational and Graphical Statistics*, DOI:10.1080/10618600.2018.1425625

[2] Liu, L., and Xiang, L. (2019). Missing covariate data in generalized linear mixed models with distribution-free random effects. *Computational Statistics & Data Analysis*, 134, 1-16. doi:10.1016/j.csda.2018.10.011

[3] Declercq, L., Jamshidi, L., Fernandez-Castilla, B., Beretvas, S.N., Moeyaert, M., Ferron, J.M. and Van den Noortgate, W. (2019). Analysis of single-case experimental count data using the linear mixed effects model: A simulation study. *Behavior Research Methods* (2019) 51:2477–2497. https://doi.org/10.3758/s13428-018-1091-y

[4] Bandara, K., Abdel-Salam, A.G. and Birch, J.B. (2020). Model robust profile monitoring for the generalized linear mixed model for Phase I analysis. *Appl Stochastic Models Bus Ind*. 2020;36:1037–1059. DOI: 10.1002/asmb.2587.

[5] Wang, Y., Zhang, X.Y., Lu, H., Croft, J.B. and Greenlund, K.J. (2022). Constructing Statistical Intervals for Small Area Estimates Based on Generalized Linear Mixed Model in Health Surveys. *Open Journal of Statistics* , 12, 70-81. https://doi.org/10.4236/ojs.2022.121005

[6] Chauvet, J., Trottier, C. and Bry, X. (2019). Component-based regularisation of multivariate generalised linear mixed models. *Journal of Computational and Graphical Statistics*, In press, 10.1080/10618600.2019.1598870. hal-02064508

[7] Achana, F., Gallacher, D., Oppong , R., Kim, S., Petrou, S., Mason, J. and Crowther, M. (2021). Multivariate Generalized Linear Mixed-Effects Models for the Analysis of Clinical Trial–Based Cost-Effectiveness Data. *Medical Decision Making* 2021, Vol. 41(6) 667–684

[8] da Silva, G.P., Laureano, H.A., Petterle, R.R., Junior, P.J.R. and Bonat, W.H. (2022). Multivariate generalized linear mixed models for underdispersed count data. *arXiv:2205.10486v1* [stat.ME] 21 May 2022

[9] da Silva, G.P., Laureano, H.A., Petterle, R.R., Junior, P.J.R. and Bonat, W.H. (2023). Multivariate Generalized Linear Mixed Models for Count Data. arXiv:2301.00921v1 [stat.ME] 3 Jan 2023

[10] Adeniyi, I. A. and Yahya, W. B. (2020). Bayesian Generalized Linear Mixed Effects Models Using Normal-Independent Distributions: Formulation and Applications. Munich Personal RePEc Arhcive Paper No. 99165

[11] Liu, C. (1996). Bayesian robust multivariate linear regression with incomplete data. *Journal of the American Statistical Association*, 91(435), pp 1219-1227.

[12] Yimer, B. B. and Shkedy, Z. (2021) Bayesian inference for generalized linear mixed models: A comparison of different statistical software procedures, *RMS: Research in Mathematics & Statistics*, 8:1, 1896102, DOI: 10.1080/27658449.2021.1896102

[13] Plummer, M. (2003). *JAGS; A Program for Analysis of Bayesian Graphical Models using Gibbs Sampling*. Proceedings of the 3rd International Workshop on Distributed Statistical Computing (DSC 2003), Vienna, 20-22 March 2003, 1-10.

[14] Martinez-Minaya, J., Lindgren, F., Lopez-Quilez, A., Simpson, D. and Conesa, D. (2021). The Integrated nested Laplace approximation for fitting models with multivariate response. *arXiv:1907.04059v2* [stat.CO] 24 Feb 2021

[15] He, S. and Lee, W. (2022). Generalized linear mixed-effects models for studies using different sets of stimuli across conditions. *Front. Psychol.* 13:955722. doi: 10.3389/fpsyg.2022.955722

[16] Xu, W. and Zhou, H. (2012). Mixed effect regression analysis for a cluster-based two- stage outcome-auxiliary-dependent sampling design with a continuous outcome. *Biostatistics* (2012), 13, 4, *pp.* 650–664 doi:10.1093/biostatistics/kxs013

[17] Burgard, J. P. and Dörr, P. (2018). Survey-weighted generalized linear mixed models. *Research Papers in Economics*, No. 1/18, Universität Trier, Fachbereich IV – Volkswirtschaftslehre, Trier

[18] Avery, L., Rotondi, N., McKnight, C., Firestone, M., Smylie, J. and Rotondi, M. (2019). Unweighted regression models perform better than weighted regression techniques for respondent-driven sampling data: results from a simulation study. *BMC Medical Research Methodology* (2019) 19:202

[19] Florez, A. J., Molenberghs,G., Verbeke, G.Kenward, M.G. Mamouris, P. and Vaes, B. (2021). Fast two-stage estimator for clustered count data with overdispersion. I-BioStat, Universiteit Hasselt, B-3590 Diepenbeek, Belgium.

[20] Dlangamandla, O. (2021). A Modelling Approach to the Analysis of Complex Survey Data. A thesis submitted to Department of Statistics Rhodes University.

[21] Rueda, M.d.M., Cobo, B. and Arcos, A. (2021). Regression Models in Complex Survey Sampling for Sensitive Quantitative Variables. *Mathematics* 2021, 9, 609. https://doi.org/10.3390/math9060609

[22] Lumley, T (2004). Analysis of complex survey samples. *Journal of Statistical Software* 2004, vol. 009, issue i08

[23] Lumley, T. (2024). *survey*: analysis of complex survey samples. R package version 4.4.

[24] Rue, H. and Held, L. (2005). Gaussian Markov Random Fields: Theory and Applications, Chapman & Hall/CRC Press, London.

[25] Held, L., & Sabanés Bové, D. (2014). *Applied Statistical Inference*. Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-37887-4

[26] Fitzmaurice, G., M., D., Verbeke, G. and Molenberghs, G. (2008). Longitudinal Data Analysis, Chapman & Hall.

[27] Stiratelli, R., Laird, N. and Ware, J. H. (1984). Random-effects models for serial observations with binary response, *Biometrics* 40(4): pp. 961–971. Available from: http://www.jstor.org/stable/2531147.

[28] Breslow, N. E. and Clayton, D. G. (1993). Approximate inference in generalized linear mixed models, *Journal of the American Statistical Association* 88(421): 9–25. Available from: http: //www.jstor.org/stable/2290687.

[29] Sauter, R. (2015). Generalised linear mixed models: likelihood and Bayesian computations with applications in epidemiology. 2015, University of Zurich, Faculty of Science. https://doi.org/10.5167/uzh-152199

[30] Simpson, D., Illian, J. B., Lindgren, F., Sorbye, S. H., & Rue, H. (2017). Going off grid: Computationally efficient inference for log-Gaussian Cox processes. *Biometrika*, *103 (1), 4*.

[31] Ordonez, J.A., Prates, M.O., Bazan, J.L. and Lachos, V.H. (2024). Penalized complexity priors for the skewness parameter of power links. *Can J Statistics*, 52:98-117. https://doi.org/10.1002/cjs.11769

[32] Handayani, D., Notodiputro, K. A., Sadik, K. and Kurnia, A. (2017). A comparative study of approximation methods for maximum likelihood estimation in generalized linear mixed models (GLMM). *AIP Conf. Proc.* 1827, 020033 (2017). https://doi.org/10.1063/1.4979449

[33] Rustand, D., van Niekerk, J., Krainski, E. T., Rue, H., & Proust-Lima, C. (2024). Fast and flexible inference for joint models of multivariate longitudinal and survival data using integrated nested Laplace approximations. *Biostatistics*, 2024, 25, 2, pp. 429–448. https://doi.org/10.1093/biostatistics/kxad019

[34] Blangiardo M. and Cameletti M. (2015). *Spatial and Spatio-temporal Bayesian Models with R –INLA*. John Wiley & Sons, Chichester

[35] Opitz T (2017). Latent Gaussian modeling and INLA: A review with focus on space-time applications. *Journal of the French Statistical Society* 158(3):62-85

[36] Rue, H., Martino, S. and Chopin, N. (2009). Approximate Bayesian Inference for Latent Gaussian Models Using Integrated Nested Laplace Approximations (with discussion). *Journal of the Royal Statistical Society*, Series B, 71, 319-392.

[37] Lumley, T. and Huang, X. (2023). Linear mixed models for complex survey data: implementing and evaluating pairwise likelihood. arXiv:2307.04944v1 [stat.ME] 11 Jul 2023

[38] Nigeria Malaria Indicator Survey 2021 Final Report. Abuja, Nigeria, and Rockville, Maryland, USA: NMEP, NPC, and ICF.

[39] van Niekerk, J., Bakka, H., Rue, H., and Schenk, O. (2021). New Frontiers in Bayesian Modeling Using the INLA Package in R. Journal of Statistical Software, 100(2). https://doi.org/10.18637/jss.v100.i02