Master's thesis

NTNU
Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

Thakshina Tharmapalan

Logistic Mixed Models for Demographic and Health Indicators

Master's thesis in Applied Physics and Mathematics
Supervisor: Geir-Arne Fuglstad
July 2021

NTNU
Norwegian University of Science and Technology

Thakshina Tharmapalan

Logistic Mixed Models for Demographic and Health Indicators

Master's thesis in Applied Physics and Mathematics
Supervisor: Geir-Arne Fuglstad
July 2021

Norwegian University of Science and Technology
Faculty of Information Technology and Electrical Engineering
Department of Mathematical Sciences

NTNU
Norwegian University of
Science and Technology

Summary

Neonatal mortality rate and vaccination coverage of measles-containing vaccine are important health indicators of a nation. The Demographic and Health Surveys (DHS) program is one major source for tracking health indicators to identify the need for interventions in low- and middle-income countries. The DHS data is transformed to a standardized format with the same structure across countries. The standardized format facilitates comparisons across surveys and countries. Samples in surveys are often collected from a population in a way that is time- and cost-effective. A complex design generally includes unequal inclusion probabilities of the units in the population, stratified sampling, and cluster sampling. DHS data is based on stratified multistage sampling. Mixed models to estimation do not typically account for these design features. However, a common way to acknowledge clustering is to include unstructured cluster-specific random effects. This adjusts the apparent sample size for the intra-cluster correlation. The main goal of this thesis is to make inferences about estimates of neonatal mortality rate and coverage of first dose measles-containing vaccine among 1-year-olds using logistic mixed models on DHS data collected in Nigeria and Kenya.

This study focuses on model estimates at the national level adjusting for explanatory variables such as maternal age, urban or rural residency, and cluster effects. The maternal age groups are incorporated as unstructured and structured random effects to evaluate the ability to borrow strengths from groups to improve the estimates, in addition, to reduce the variance in the estimates. The explanatory and predictive strengths of the models are investigated, and inference is conducted using the integrated nested Laplace approximation (INLA) with the package R-INLA in R. The methods are evaluated on a set of scoring rules through a simulation study and on real survey data.

This analysis concludes that mixed effects models applied on DHS data manage to incorporate clustering for small cluster effects. Unstructured and structured random effects succeed to limit variation in estimates between maternal age groups. The uncertainty in the estimates of the outermost age groups, which consist of only a small amount of data, is reduced. Mixed models should, however, be reconsidered when it is applied to complex survey data because they fail to correctly account for important aspects of survey methodology such as sampling weights.

Sammendrag

Nyfødtdødelighet og vaksinedekning mot meslinger er viktige helseindikatorer for en nasjon. Demographic and Health Surveys (DHS) er en pålitelig kilde for å spore indikatorer for å forbedre en nasjons helse i lav- og mellominntektsland. DHS dataene blir transformert til et standardisert format med samme struktur på tvers av land. Det standardiserte formatet gjør det mulig å sammenligne estimater på tvers av undersøkelser og land. Prøver fra en populasjon gjennom undersøkelser er ofte samlet på en tids- og kostnadseffektiv måte. Undersøkelsedesign er typisk kompleks fordi det generelt inkluderer ulike inkluderingssannsynligheter for observasjoner i befolkningen, stratifisert prøvetaking og klyngesampling. DHS data er basert på stratifisert flertrinnssampling. Blandede modeller til estimering tar vanligvis ikke hensyn til disse designfunksjonene. Imidlertid brukes ustrukturerte klyngespesifikke tilfeldige effekter generelt for å anerkjenne klyngingen. Dette justerer den tilsynelatende prøvestørrelsen for korrelasjonen innen klyngen. I denne masteroppgaven anvendes blandede modeller for å estimere nyfødtdødelighet og vaksinasjonsdekning mot mesling av første dose blant 1-åringer på DHS data samlet i Nigeria og Kenya.

Denne studien fokuserer på modellestimater på nasjonalt nivå, og justerer for forklarende variabler som mors alder, urbant- eller landlig bosted, og klyngeeffekt. Aldersgrupper for mødre er innlemmet i modellen som strukturerte og ustrukturerte tilfeldige effekter for å evaluere evnen til å låne styrker fra grupper og for å redusere avviket i estimatene. De forklarende og prediktive styrkene til de utviklede modellene undersøkes. Modellene er estimert ved hjelp av den integrerte nestede Laplace-tilnærmingen (INLA) med pakken R-INLA i R. Metodene blir evaluert ved hjelp av et sett av mål i et simuleringsstudie og på reelle undersøkelsesdata.

Denne analysen konkluderer med at blandede modeller brukt på DHS data klarer innlemme for klynging for små klyngeeffekter. Ustrukturerte og strukturerte tilfeldige effekter bidrar til å begrense variasjon i estimater mellom mors aldersgrupper. Blandede modeller bør imidlertid revurderes når de brukes på komplekse undersøkelsesdata fordi de ikke tar hensyn til viktige aspekter som inkluderingssannsynligheter.

Preface

This thesis was written as the final part of the five-year master program in Applied Physics and Mathematics at the Norwegian University of Science and Technology (NTNU), with specialization in Industrial Mathematics.

Working on this thesis has been an exciting and rewarding experience, both academically and personally. It has allowed me to utilize the knowledge I have gained during the last five years and develop a deeper understanding of several fields in statistics. Survey data from the Demographic and Health Surveys program is used for analysis. The simulations and the generation of the results is implemented in R, with the help of INLA packages.

I want to give a very special thanks to my supervisor Geir-Arne Fuglstad for his excellent support and guidance during the period of developing this thesis. My gratitude also goes to my friends for the wonderful times we have shared together the past five years at NTNU. Finally, my deep and sincere gratitude to my family for the love and support they have given me throughout the entire process.

Thakshina Tharmapalan,
Trondheim, July 2021

Table of Contents

Chapter 1

Introduction

The United Nation's 2030 development agenda is described by the Sustainable Development Goals (SDGs) agreed by world leaders. The SDGs consist of 17 goals. The targets include reducing extreme poverty, hunger and halting the spread of diseases such as HIV and AIDS (UN, 2021). Two important health indicators of binary responses are neonatal mortality rate (NMR: the proportion of children who die within the first month after live birth) and coverage of the first dose of measles-containing vaccine (MCV1) among 1-year-olds. Target 3.2 in SDGs about NMR reads, "by 2030, end preventable deaths of newborns with all countries aiming to reduce neonatal mortality ... to at least as low as 12 per 1,000 live births". Measles is on the other hand, a highly contagious disease, and where unvaccinated children, particularly under the age of five, are at the highest risk of measles and its complications, including death (WHO, 2019). A major proportion of child mortality is today preventable and treatable by cost-effective interventions (Sharrow et al., 2020). The tracking and coverage of these indicators are essential at a local, regional and global level to see improvements in child and maternal health. Thereby, scarcity of resources can point to where they are needed the most and is crucial in saving lives (Hancioglu and Arnold, 2013).

Estimating mortality rates, for instance, is complicated. Vital registration systems are necessary to obtain exact estimates. Typically, low- and middle-income countries do not have this system in place. If complete birth- and death data or vaccination status in a country cannot be provided, NMR and MCV1 must be estimated based on data arising from surveys and censuses. A primary source for estimates and forecasts of health indicators in low- and middle-income countries is surveys provided by the Demographic and Health Surveys (DHS) program. Survey programs collect answers from questionnaires through physical visits in selected households, and which include a wide range of topics related to children and maternal health. The surveys are gathered from face-to-face interviews in households and are often known as household survey data. The DHS program provides data that is collected from household surveys in more than 90 countries. In this thesis, DHS data from both Nigeria and Kenya is used to investigate if conclusions are consis-

tent between countries. All DHS data passes through a standardization process where the structure of the original raw data is transformed into a common format (ICF et al., 2018). Hence, with some adjustments, the methods may apply to other low- and middle-income countries that are covered by the DHS program, where measurements such as child mortality must be estimated from survey data.

Survey data has in general a complex correlation structure, because the data cannot be assumed to be collected by random sampling. The observations might have different probabilities of being selected in the sample, i.e. different inclusion probabilities. A complex sample consists of the individual observations together with the sample weights that are related to the inclusion probabilities, and the design descriptors (Lee and Forthofer, 2006). The complexity of a survey design is predefined and determined by the resource constraints for planning, conducting, and processing the survey. Complex survey data often include stratified multistage cluster sampling. Stratification ensures that the total sample represents all desirable subgroups, such as the inclusion of residents from each subnational area. Cluster sampling is sampling in restricted areas called clusters, and which yields to correlation among observations because of unobserved heterogeneity between clusters. This may be a result of the clustering of subjects within groups due to their similarities. A failure to address the complex sampling issues might result in biased estimates of the model parameters. Mixed effects models typically do not fully account for the complex survey design. However, unstructured cluster-specific random effects can generally be incorporated in regression models to account for the clustering. This adjusts the apparent sample size for the intra-cluster correlation.

The DHS data used for this analysis is designed to produce representative estimates for a majority of the survey indicators at the national level, for urban and rural areas separately, at the regional level, and for selected indicators at the county level. Urban and rural clusters can further be disaggregated into more specific groups, such as maternal age groups within each type of cluster. Maternal age is an important explanatory variable for binary outcomes related to child health indicators. Multiple groups like maternal age groups as a categorical covariate might, however, be challenging because the oldest and youngest groups generally represent a minority group of the observations. In such cases, methods that allow borrowing strength from remaining groups can be helpful. Random effects are one way to borrow information from groups in different manners. Independent and identically distributed random effects construct a shrinkage towards the common mean of the estimates to account for unstructured variability in the data. The shrinkage is dependent on the number of observations in each group. The estimates in groups of few observations are forced towards the mean, otherwise, they remain unaffected when there is enough information. Structured random effects help to reduce the uncertainty in the estimates, in addition to limit the variation between the estimates in the groups. The models are estimated using the integrated nested Laplace approximation (INLA).

This thesis will focus on the estimation of NMR and coverage of MCV1 among 1- year-olds while accounting for complex survey designs. NMR and MCV1 are two important

demographic and health indicators with either rare or very prevalent outcome. In particular, the associations between urban or rural areas and maternal age are explored as dependent variables as they are believed to play a crucial role in vaccination coverage and child mortality rates, respectively. Ignoring the survey design that involves clustering, assumes that observations from the same cluster are independent. Sampling effort in rural areas for complex sampling schemes might, for instance, be different than for urban areas. Mixed models are applied to account for clustering by including unstructured cluster-specific random effects. The goal of this thesis is to investigate methods that incorporate structured and unstructured random effects to explore the ability to improve the estimates by borrowing strength across groups for clusters and maternal age groups.

There is no simple way to account for survey design applied on mixed models. The sample weights, for instance, are difficult to adjust for. In real survey data, the truth is, however, not known, as that would require a census of the population. Simulation studies allow to produce finite populations within a known and controlled framework. In this thesis, surveys are independently and repeatedly drawn from three individual simulated populations of different characteristics. The populations have different predefined cluster effects, and the sampling frames consist of different inclusion probabilities of clusters. The purpose of the simulation study is to investigate the ability to account for clustering by comparing fixed effects and mixed effects, and the importance of accounting for inclusion probabilities.

The structure of this thesis is as follows. Chapter 2 provides a brief review of prerequisite material required for the analysis, while Chapter 3 describes the data sets utilized. The simulation study is presented in Chapter 4 to assess the performance of the methods. Chapter 5 presents the results of the methods applied to real surveys conducted in Nigeria in period 2014-2018 and Kenya in period 2010-2014. Finally, findings are discussed in Chapter 6.

Chapter 2

Background

In this section, the relevant background theory needed to understand the methods used in this thesis is presented.

2.1 Generalized linear mixed models

In analysis of data with multilevel structure, groups in the data are treated as a random sample from a population of groups. When applying a fixed effects model with categorical covariates, some groups may not have sufficient information to make inferences due to the small group size. Mixed models can borrow strengths between groups. Random effects are added for the groups instead of fixed effects. This can also be interpreted by introducing correlations among observations in the same group. For non-normal responses, the framework of the generalized linear models (GLMs) can be used. Generally, a set of coefficients and independent variables are incorporated in a linear function, called a linear predictor. The linear predictor can consist of both fixed effects and random effects. As a result, the generalized linear mixed models (GLMMs) is a synthesis of GLMs and linear mixed models.

Multilevel models may contain both nested and non-nested factors. Non-nested factors are individual-specific covariates that arise when individuals are characterized by overlapping categories of attributes (Gelman and Hill, 2006). For instance, consider a simple study of the association between NMR and maternal age groups. Then a non-nested factor is the age of the mother. The individuals may be further nested within geographical areas, such as urban or rural residency and clusters. A simple model allowing for between-cluster variation is

$$\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \gamma_{0i} = \beta_0 + \beta_1 x_{ij1} + \cdots + \beta_K x_{ijK} + \gamma_{0i}, \tag{2.1}$$

where i denotes the level-2-units (e.g. clusters) and j is the level-1-units (e.g. children),

and $i = 1, \ldots, M$, $j = 1, \ldots, n_i$, with n_i multiple measurements per individual within cluster i. The coefficient of the fixed effects are β_k, $k = 1, \ldots, K$, and which in this context may be the coefficients separating urban and rural clusters. The fixed effects are dependent variables of interest. The random effects in the model, γ_{0i}, are random deviations from the fixed intercept β_0. Responses, y_{ij}, are usually assumed to follow a normal distribution $\mathcal{N}(\eta_{ij}, \sigma^2_\gamma)$, where σ^2_γ indicates the variance, and therefore the degree of heterogeneity within clusters (Hedeker, 2005).

The model can be easily extended to include more random effects,

$$\eta_{ij} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{u}_{ij}^T\boldsymbol{\gamma}_i + \boldsymbol{z}_{ij}^T\boldsymbol{\delta}, \quad i = 1, \ldots, M, \quad j = 1, \ldots, n_i \tag{2.2}$$

where $\boldsymbol{u}_{ij}$ is typically a subvector of the covariates $\boldsymbol{x_{ij}}$, and $\boldsymbol{z}_{ij}^T$ in the third term denotes the non-nested factors of covariates. The vector of random effects is assumed to be independent and identically normal with $\boldsymbol{\gamma}_i \sim \mathcal{N}(\mathbf{0}, \boldsymbol{\Sigma})$, and $\boldsymbol{\Sigma}$ is a positive definite covariance matrix. An example is that $\boldsymbol{\gamma}_i$ accounts for random between-cluster variation. Further, $\boldsymbol{\delta} \sim \mathcal{N}(0, \boldsymbol{\Sigma\delta})$ may be coefficients of random effects on age groups across clusters. The observations are assumed to arise from a distribution in the exponential family such as in the framework of a GLM. The expected value of the response, $\mu_{ij} = \mathrm{E}[Y_{ij}|\boldsymbol{\delta}, \boldsymbol{\gamma}_i, \boldsymbol{x}_{ij}]$, is related to the linear predictor through a suitable link function

$$\mu_{ij} = h(\eta_{ij}) \quad \text{or} \quad \eta_{ij} = g(\mu_{ij}). \tag{2.3}$$

The response function, h, is the inverse of the link function, $g = h^{-1}$.

2.1.1 Mixed Logit Models

For binary response variables and multilevel data, the mixed effects logistic regression is the most popular GLMM. The logit link function is utilized so that for individual j in cluster i,

$$\eta_{ij} = \mathrm{logit}(\mu_{ij}) = \log(\frac{\mu_{ij}}{1 - \mu_{ij}}), \tag{2.4}$$

where μ_{ij} is the conditional expectation of a random variable, Y_{ij}, and equals the conditional probability of a successful outcome, $y_{ij} = 1$, given the random effects, $\mu_{ij} = \mathrm{E}(Y_{ij}|\boldsymbol{\delta}, \boldsymbol{\gamma}_i, \boldsymbol{x}_{ij}) = \mathrm{P}(Y_{ij} = 1|\boldsymbol{\delta}, \boldsymbol{\gamma}_i, \boldsymbol{x}_{ij})$. We assume $Y_{ij}|\mu_{ij} \sim \mathrm{Bernoulli}(\mu_{ij})$. The linear predictor in (2.2) becomes

$$\eta_{ij} = \log \frac{\mathrm{P}(Y_{ij} = 1|\boldsymbol{\delta}, \boldsymbol{\gamma}_i, \boldsymbol{x}_{ij})}{\mathrm{P}(Y_{ij} = 0|\boldsymbol{\delta}, \boldsymbol{\gamma}_i, \boldsymbol{x}_{ij})} = \boldsymbol{x}_{ij}^T\boldsymbol{\beta} + \boldsymbol{u}_{ij}^T\boldsymbol{\gamma}_i + \boldsymbol{z}_{ij}^T\boldsymbol{\delta},$$

and we attain

$$\mu_{ij} = \text{expit}(\eta_{ij}) = \frac{1}{1 + \exp(-\eta_{ij})}, \tag{2.5}$$

where $\text{expit}(x) = \text{logit}^{-1}(x)$.

2.2 Bayesian hierarchical modelling

The Bayesian approach along with the frequentist approach are two major paradigms of statistical inference. "In frequentist statistics, it is assumed that the data is distributed according to a model with fixed and unknown parameters" Bolstad (2007). The parameters are considered as unobserved and unknown but fixed constants. Inference about the parameters is based on likelihoods from the sampling distribution. Likelihoods are probabilistic descriptions of how the data arises conditionally on parameters. They are not conditional on the sample collected. Hence, prior belief about the parameters is not taken into account. An alternative approach is the Bayesian approach. In the Bayesian framework, the parameters of interest are treated as random variables. To help understand Bayesian hierarchical modeling, a brief introduction to Bayesian statistics is presented prior to the theory of Bayesian hierarchical models. The theory here is primarily based on the book by Gelman et al. (2014).

Bayesian inference is based on the posterior distribution containing all information about the unknown parameter after having observed the data. Inference about the parameters of interest, θ, is based on the posterior distribution of the parameters given the observed data, y. To obtain the posterior distribution, the joint probability distribution is required. The function of the joint probability distribution is the product of the prior distribution, $\pi(\theta)$, and the sampling distribution, $\pi(y|\theta)$,

$$\pi(\theta, y) = \pi(\theta)\pi(y|\theta). \tag{2.6}$$

The prior belief of the parameter is incorporated in the model by assigning a prior distribution to the parameters. Bayesian inference apply the Bayes' theorem to update the knowledge about θ conditioning on the outcome, y,

$$\pi(\theta|y) = \frac{\pi(\theta)\pi(y|\theta)}{\pi(y)}. \tag{2.7}$$

$\pi(y)$ is a normalization constant, independent of θ which means that

$$\pi(\theta|y) \propto \pi(\theta)\pi(y|\theta). \tag{2.8}$$

A statistical model may involve several parameters that are related in some way. Correlation within groups of the observations, may be desired to be incorporated in the model. Then it is appropriate to consider hierarchical models, which allow the parameters of the prior distribution can themselves be estimated from data. Hence, problems of overfitting can be reduced (Gelman et al., 2014). The parameters in hierarchical models are given a probabilistic specification in terms of further parameters, known as hyperparameters.

A Bayesian hierarchical model is a Bayesian model with one or more layers of latent structures and can be split into stages. Let y_i be a measured quantity in observation i for $i = 1, \dots n$. Further, assume that the distribution of y_i is conditional on a set of latent variables, and the distribution of the latent variables is conditional on a set of hyperparameters with prior distributions. Let $\boldsymbol{x}$ denote the unobserved latent field and let $\boldsymbol{\theta}$ be the parameters. Three-stage hierarchical models can be defined by the following stages.

$$\text{Stage 1: } \boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta} \sim \pi(\boldsymbol{y}|\boldsymbol{x}, \boldsymbol{\theta})$$

$$\text{Stage 2: } \boldsymbol{x}|\boldsymbol{\theta} \sim \pi(\boldsymbol{x}|\boldsymbol{\theta})$$

$$\text{Stage 3: } \boldsymbol{\theta} \sim \pi(\boldsymbol{\theta})$$

The models used in this thesis belong to Latent Gaussian Models (LGMs), a subclass of Bayesian hierarchical models. In a LGM, the prior distribution on a latent field must be Gaussian conditioned on the parameters. Structure additive regression models, such as the GLMMs are members of LGMs. The unified framework of structure additive models assumes the distribution of the response variable, y_i, to belong to an exponential family conditional on the mean, μ_i, as described in (2.3) and likelihood parameters. The linear predictor from the GLMM is now referred to as the structured predictor η_i, which is connected with the mean μ_i through a link function, $g(\mu_i) = \eta_i$. The structured predictor accounts for the effects of covariates in an additive way,

$$\eta_i = \beta_0 + \sum_{k=1}^{n_\beta} \beta_k x_{ki} + \sum_{j=1}^{n_f} f^{(j)}(u_{ji}) + \epsilon_i,$$

where β_0 is the intercept, the $\{\beta_k\}$'s represent the linear effects of covariates $\{x_{ki}\}$ on the response. $\{f^{(j)}(\cdot)\}$'s are random functions of the covariates $\{u_{ji}\}$, and $\{\epsilon_i\}$'s are unstructured random terms (Rue et al., 2009). In this analysis, maternal age groups are considered as fixed effects or random effects, while cluster information are considered as random effects. The maternal age groups are categorical covariates with seven distinct groups, $u = \{1, 2, \dots, 7\}$. Then, the random effects are, $f(u) = \gamma_u$, for $u = 1, \dots, 7$, where $(\gamma_1, \dots, \gamma_7) \sim \mathcal{N}(0, \boldsymbol{Q}^{-1})$. The cluster effect would be $f(\text{“cluster ” i}) = \gamma_i$ where $\gamma_i, \dots, \gamma_M \overset{iid}{\sim} \mathcal{N}(0, \sigma_\gamma^2)$ for M clusters.

2.3 Structured and unstructured random effects

This thesis focus on categorical predictor variables such as maternal age groups and urban or rural type of residency. In this context, observations are divided into groups according to the categorical predictor variables. In a population study, the samples collected from different geographical regions are often disaggregated by so-called clusters in which responses tend to be correlated. However, by using fixed effects for each cluster, the estimates may reproduce the exact observed values. To account for correlation among clusters, we need to introduce some form of shrinkage to model the variation between the clusters and from the main level of this multilevel data structure of limited data. Accordingly, random effects are proposed to specify how some model parameters vary randomly across groups, and to capture variability in response that is not explained by fixed effects and response distribution.

There are several types of random effects, and which are implemented in INLA. The types of random effects utilized in INLA for this analysis are presented primarily using Gómez-Rubio (2020). The vector of random effects, $\boldsymbol{u}$, in INLA is multivariate Gaussian distribution with

$$\boldsymbol{u} \sim \mathcal{N}(0, \boldsymbol{\Sigma}), \tag{2.9}$$

where $\boldsymbol{\Sigma} = (\boldsymbol{Q}/\tau)^{-1}$ is a covariance matrix, τ^2 is a generic variance parameter, and $\boldsymbol{Q}$ is a precision matrix that defines the dependence structure of the random effects and that may depend on further parameters (Gómez-Rubio, 2020). The structure of $\boldsymbol{Q}$ is quite sparse when random effects are a latent Gaussian Markov random field and which prepares the ground for fast computations (Rue and Held, 2005). In this analysis, independent and identically distributed (iid) random effects and random effects of random walks of order one (rw1), and random walks of order two (rw2) are utilized.

The simplest way to account for unstructured variability in the data is by assuming the random effects to be iid. The precision matrix, $\boldsymbol{Q}/\tau$, of iid random effects is an identity matrix, with $\boldsymbol{Q} = \text{diag}\{s_1, \dots s_n\}$ denoting scaling factors of a diagonal matrix. "Scaling factors are defined by means of parameter scale within the $f()$ function used to define the latent effect in the formula" Gómez-Rubio (2020). Now, let $\boldsymbol{u} = (u_1, \dots u_n)$ be a vector of Gaussian observations. Then, rw1 assumes that increments $\Delta u_i = u_i - u_{i-1} \sim \mathcal{N}(0, \tau_{\text{rw1}})$. This is equivalent to assuming that the distribution of vector $\boldsymbol{u}$ is Gaussian with zero mean and precision matrix $\boldsymbol{Q}/\tau_{rw1}$, where $\boldsymbol{Q}$ contains the neighborhood structure of the model (Gómez-Rubio, 2020). Similarly, the rw2 for Gaussian vector $\boldsymbol{u} = (u_1, \dots u_n)$ is defined by assuming independent second order increments $\Delta^2 u_i = u_i - 2u_{i-1} + u_{i-2} \sim \mathcal{N}(0, \tau_{\text{rw2}})$.

Random effects on categorical covariates aim to limit the variation between groups in form of shrinkage, sometimes called partial-pooling. The shrinkage is weighted by overall balance between groups in the random effects structures depending on the group size. Mixed

effects models are very useful when there is a number of groups, but only very few data in each groups. The parameters are estimated by maximizing a penalized log-likelihood. "The log-likelihood of the model is a measure of the fit (or lack there of), while the penalty helps us avoid fitting overly complex smooths" Simpson (2021).

The DHS data collected in Nigeria between 2014-2018 is used to illustrate the behavior of the different random effects in model fits. A line between the expit of the mean estimates of the NMR associated with maternal age groups as random effects of iid, rw1 and rw2 in urban area is shown in Figure 2.1. Data in urban is further divided in seven age groups, which means the data in each groups is limited. Moreover, the youngest and oldest age groups do typically have fewer observations. The plots from left to right shows that the smoothness of the line between the mean of the estimates increases in the order of iid, rw1 and rw2.

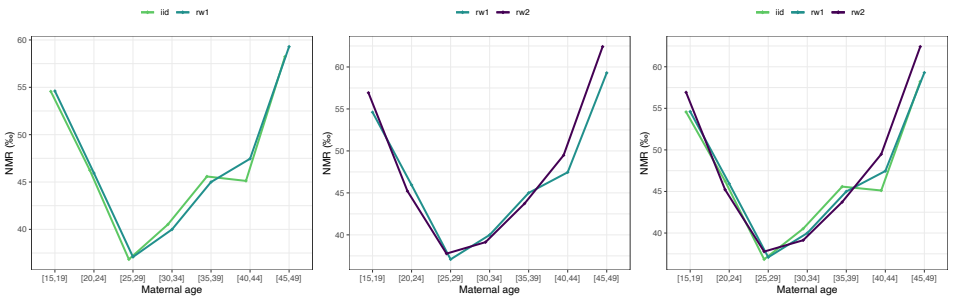


Figure 2.1: Illustration of the behaviour of random effects of iid, rw1 and rw2 on DHS data from Nigeria collected in period 2014-2018 for estimating NMR (per 1 000 live births) associated with maternal age groups in urban area.

The estimates in groups with few observations are expected to shrink toward the common mean for unstructured iid random effects. For limited data, rw1 and rw2 collect information from their neighboring groups. In a rw1 and rw2 model, the variation between the values of its neighbors has a penalty that increases with increased change. The rw1 model has a first order penalty to limit the variation between the estimates of the groups. Similarly, rw2 has a penalty, but of second order. Thus, for a model with rw2, it is not only expected a decrease in the deviation between the group, but also a reduction in the change between its neighbors. This aims to describe the change to be secondly derivable as illustrated in Figure 2.1. These penalties of orders 1 and 2 results in a smoother variation in models rw1 and rw2 compared to the iid random effects model. Moreover, the smoothing is stronger for rw2, which is not surprising as we expect realizations of higher order.

2.4 Inference

The posterior distribution of the parameters of a Bayesian model can be used for inference about the model. The posteriors cannot be computed analytically for most Bayesian mod-

els. Hence, there exist several approximations. The most common sampling method is the Markov chain Monte Carlo (MCMC). The theory of the MCMC can be found in, for example, in Chapter 11 of Gelman et al. (2014). In MCMC, the values of the parameter, θ, are drawn from proposal distributions, and the draws are updates according to some acceptance rule to improve the approximation of the target posterior distribution $\pi(\theta|y)$. The sampling is done sequentially and each sample is drawn based on the previous value drawn. This forms a Markov chain, which is a sequence of random variables, $\theta_1, \theta_2, \ldots,$ where the distribution of any θ_t for any arbitrary t, and given all θ's only depends on the latest value, θ_{t-1}. By increasing the number of steps in a Markov chain, the approximation of the desired posterior distribution improves.

Highly accurate estimates with MCMC algorithms requires generating large samples. Moreover, MCMC methods perform poorly on LGMs as discussed in Section 1.4 in Rue et al. (2009). The components of the latent field $\boldsymbol{x}$ are strongly dependent on each other, and $\boldsymbol{x}$ and $\boldsymbol{\theta}$ are dependent, which requires modifications of the algorithm. An alternative, deterministic approach for estimating the posterior of LGMs called Integrated Nested Laplace approximation (INLA) is introduced in the paper Rue et al. (2009). It outperforms MCMC for LGMs for the computational cost. Rue et al. (2017) present three assumptions that is required for accurate with high degree of uncertain city, and computationally feasible approximations. The number of hyperparameters, $\boldsymbol{\theta}$ is small and not more than around 20. Furthermore, the latent field, $\boldsymbol{x}|\boldsymbol{\theta}$ of a Gaussian distribution, and Gaussian Markov random field (GRMF) for high dimension of $\boldsymbol{x}$. The definition of GMRFs is provided in Section 2.2.1 of Rue and Held (2005). Finally, the observations, $\boldsymbol{y}$, are mutually conditionally independent given $\boldsymbol{x}$ and $\boldsymbol{y}$. To understand the theoretical aspect of INLA, a brief presentation is given, for further reading, see e.g. Rue et al. (2009) and Martins et al. (2009).

INLA is based on the Laplace approximation, which aims to approximate an integral

$$I = \int_{-\infty}^{\infty} \exp(f(x))dx, \tag{2.10}$$

where $f(x)$ is the density of a random variable X. Let $x^\star$ denote the location of its maximum, $x^\star = \text{argmax}_x f(x)$. Then the second order Taylor approximation for $f(x)$ around $x = x^\star$ is

$$f(x) \approx f(x^\star) + (x - x^\star)f'(x^\star) + \frac{1}{2}(x - x^\star)^2 f''(x^\star).$$

$f'(x^\star) = 0$ eliminates the second term, and the approximation of the integral in (2.10) can be expressed as

$$I \approx C \int_{-\infty}^{\infty} \exp\left(- \frac{(x - x^\star)^2}{2\sigma^{2\star}} \right) dx \tag{2.11}$$

where the constant $C = \exp(f(x^\star))$ and $\sigma^{2\star} = -1/[f''(x^\star)]^{-1}$. Hence we get the integrand in a form of the density of a Gaussian distribution.

The posterior marginals of interest may be written as

$$\pi(x_i|\boldsymbol{y}) = \int \pi(x_i|\boldsymbol{\theta}, \boldsymbol{y})\pi(x_i|\boldsymbol{y})d\boldsymbol{\theta}, \tag{2.12}$$

$$\pi(\theta_i|\boldsymbol{y}) = \int \pi(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}_{-j} \tag{2.13}$$

where $\boldsymbol{\theta}_{-j}$ is all θ's but θ_j. The INLA approach aims to construct nested approximations,

$$\tilde{\pi}(x_i|\boldsymbol{y}) = \int \tilde{\pi}(x_i|\boldsymbol{\theta}, \boldsymbol{y})\tilde{\pi}(x_i|\boldsymbol{y})d\boldsymbol{\theta},$$

$$\tilde{\pi}(\theta_i|\boldsymbol{y}) = \int \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})d\boldsymbol{\theta}_{-j}.$$

Let $\tilde{\pi}$ denote the approximation of the posterior $\pi(\boldsymbol{\theta}|\boldsymbol{y})$ and

$$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y}) \propto \frac{\pi(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})}{\pi_G(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{y})}\bigg|_{\boldsymbol{x}=\boldsymbol{x}^\star(\boldsymbol{\theta})} \tag{2.14}$$

and where $\tilde{\pi}_G(\boldsymbol{x}, \boldsymbol{\theta}, \boldsymbol{y})$ is the Gaussian approximation to the full conditional of $\boldsymbol{x}$, and $\boldsymbol{x}^\star(\boldsymbol{\theta})$ is the mode of the full conditional for $\boldsymbol{x}$, for a given $\boldsymbol{\theta}$ as stated in Rue et al. (2009). According to Rue et al. (2009) the approximation in (2.14) the marginal distribution of x_i can be computed by using numerical integration:

$$\tilde{\pi}(x_i|\boldsymbol{y}) = \sum_k \tilde{\pi}(x_i|\theta_k, \boldsymbol{y})\tilde{\pi}(\theta_k|\boldsymbol{y})\Delta_k, \tag{2.15}$$

where Δ_k is the weight with corresponding value θ_k.

$\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is primarily used to integrate out the uncertainty with respect to $\boldsymbol{\theta}$ when approximating the posterior marginals of x_i in (2.12). The INLA method can be divided into three main steps as proposed in Section 3.1 in Rue et al. (2009). The first step is to optimize $\log\{\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})\}$ with respect to $\boldsymbol{\theta}$ to locate the mode of $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$. For that, the quasi- Newton method can be used to approximate the second derivatives of $\log\{\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})\}$. The second step is to compute the negative Hessian matrix $\boldsymbol{H} > 0$, of $\log\{\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})\}$ at mode $\boldsymbol{\theta}^\star$, by using finite differences. Let $\boldsymbol{\Sigma} = \boldsymbol{H^{-1}}$ which is the covariance matrix of $\boldsymbol{\theta}$ for Gaussian density. Further, let $\boldsymbol{\Sigma} = \boldsymbol{V\Lambda V^T}$ be the eigen-decomposition of $\boldsymbol{\Sigma}$. Define

$$\boldsymbol{\theta}(\boldsymbol{z}) = \boldsymbol{\theta}^\star + \boldsymbol{V\Lambda^{1/2}z} \tag{2.16}$$

where $\boldsymbol{z} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I})$ if $\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is a Gaussian distribution. The third step is to use $\boldsymbol{z}$-parametrization to explore $\log\{\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})\}$ to locate the majority of the probability mass.

A grid with all point of $\log \tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{y})$ is constructed. Each point is considered as significant, and which is used in the numerical integration in (2.15). Finally, an interpolant to $\log(\tilde{\pi}(\boldsymbol{\theta}|\boldsymbol{u}))$ is created from the points estimated in the grid, and to compute marginals $\tilde{\pi}(\theta_j|\boldsymbol{y})$. A more detailed explanation of the algorithms is found in Rue et al. (2009). Now the set of weighted points θ_k is attained. We want further to obtain accurate approximations for the posterior marginal for x_i's, with $x_i|\boldsymbol{\theta}$. Rue et al. (2009) propose three approximation methods; Gaussian-, Laplace- and simplified Laplace approximation. In general, the Laplace approximation is preferred. However, a much smaller cost with the simplified Laplace approximation compensates for the slight loss in accuracy.

INLA method is computationally beneficial because the approach can provide accurate approximations in seconds or minutes. Moreover, an extensive variety of different LGMs can easily be applied with the same general implementation with minor adjustments such as changing the likelihood, model components, and priors. INLA is implemented in the R-package `INLA`[1]. It can produce estimates in a matter of seconds. Hence simulation studies may be performed with several different models. To assess the performance of models and determine the most appropriate model for estimation and forecast, some assessment criteria and scoring rules must be established. The next section outlines some common choices.

2.5 Scoring predictions

There are several statistical scores for validating the accuracy of regression methods. Model performance can be evaluated by conduction simulation studies with M repeated simulations, and to estimate the ability to reproduce parameters of interest. Based on inference for the M simulations, properties such as mean squared error, mean bias error, coverage, and mean signed deviation can be measured. This section presents these scoring rules for a single simulation. In addition, the continuous ranked probability score, which is used to evaluate the predictive performance of the methods applied on real data set by DHS is presented.

The mean bias error (MBE) is the first scoring rule considered. Let $y_i, i = 1, 2, \ldots, n$ denote the simulation rates, and let the predicted rates denote $\hat{y}_i, i = 1, 2, \ldots, n$ for n predicted data points for which the true simulated points are known. Then the MBE of an estimator, $\hat{y}$, is simply the average bias in the prediction,

$$\text{MBE} = \frac{1}{n}\sum_{i=1}^{n}(y_i - \hat{y}_i). \tag{2.17}$$

[1] `https://www.r-inla.org`

The bias of a predicted rate measures the tendency of a model to overestimate or underestimate the rate. The MBE is frequently used to determine whether any steps need to be taken to correct the bias in the model. A negative MBE indicates that the predictions are smaller in value than observations. The MBE is not commonly used as a measure of the model error, because high individual errors in prediction may still produce low MBE.

The mean squared error (MSE) imply the average of the errors squared

$$\text{MSE} = \frac{1}{n}\sum_{i=1}^{n}(\hat{y}_i - y)^2, \tag{2.18}$$

where y is the true rate and $\hat{y}$ is the predicted rate. MSE of low values indicates a better performing model.

For $j = 1, 2, \ldots M$ simulations, a total score are evaluated by the average of the scoring rules, SR, for every singe simulations performed,

$$\text{SR} = \frac{1}{M}\sum_{j=1}^{M}\text{SR}_j. \tag{2.19}$$

The coverage of a $100(1-\alpha)\%$ confidence interval satisfies

$$\text{Coverage} = \frac{1}{M}\sum_{j=1}^{M}\mathbb{1}\{\hat{\pi}_{\alpha/2} < \pi < \hat{\pi}_{1-\alpha/2}\}, \tag{2.20}$$

where the notation $\mathbb{1}\{\hat{\pi}_{\alpha/2} < \pi < \hat{\pi}_{1-\alpha/2}\}$ is the indicator function, which takes on the value 1 if $\{\hat{\pi}_{\alpha/2} < \pi < \hat{\pi}_{1-\alpha/2}\}$, and takes on the value 0 otherwise. $\hat{\pi}$ is the estimates of the lower and upper bounds in expit-scale. π is the observed point estimate of the simulated data. In context of NMR and MCV1 as binary responses, π is the observed proportion of deaths or vaccinated children, respectively.

The mean signed deviation (MSD) of the estimates of the cluster effects, σ_c, for M simulations is

$$\text{MSD}(\hat{\sigma}_c) = \frac{1}{M}\sum_{j=1}^{M}\hat{\sigma}_{ci} - \sigma_c. \tag{2.21}$$

The continuous ranked probability score (CRPS) is a generalized version of the mean absolute error that applies on probabilistic forecasts. Good probabilistic forecasts aim to maximize the sharpness of the predictive distributions subject to calibration. Calibration refers to the statistical consistency between the distributional forecasts and the observations, and sharpness is about the concentration of the predictive distributions (Gneiting

and Raftery, 2007). CRPS measures the compatibility between the predictive distribution and the observations. Thus, the CRPS not only accounts for the point predictions, but also the associated uncertainty.

Given a probabilistic forecast with cumulative distribution function $F(y)$, the crps may be expressed as

$$\text{crps}(F, x) = \int_{-\infty}^{\infty} (F(y) - \mathbb{1}\{x \leq y\})^2 dy, \tag{2.22}$$

where x is the observation and $\mathbb{1}$ is the Heaviside step function that is assigned to 1 if $\{x \leq y\}$ is true and 0 otherwise. From (2.22), the crps is always positive, and the closer the CRPS is to zero, the closer the predictive distribution is to the true value. The crps captures the uncertainty in the predictions as well as the uncertainty in the observations. Hence, the crps favors predicted value with a higher bias but much lower variance than a predicted value with low bias and high variance.

For a predictive distribution that is Gaussian with mean μ and variance σ^2, the crps is

$$\text{crps}(N(\mu, \sigma^2), x) = \sigma \left[\frac{1}{\sqrt{\pi}} - 2\phi(\frac{x-\mu}{\sigma}) - \frac{x-\mu}{\sigma}\big(2\Phi(\frac{x-\mu}{\sigma}) - 1\big) \right], \tag{2.23}$$

where ϕ is the probability density function and Φ is the cumulative probability function of a standard normal distribution. The crps is expressed in the same unit as the observed variable and is in practice the average over n individual crps values. Let CRPS denote the average of the crps values defined as

$$\text{CRPS} = \frac{1}{n} \sum_{i=1}^{n} \text{crps}(F_i, y_i), \tag{2.24}$$

where y is an observed quantity, and F is the prediction of the corresponding y.

2.6 Complex survey designs

Many complex surveys are based on stratified multistage clustering design. Namely, clusters are grouped into strata, often with several levels of stratification and several stages of clustering. In particular, each sample from DHS data that is used for this analysis is based on a stratified two-stage cluster design. In this section, aspects of survey methodology relevant for understanding the concepts of stratification and one- and two-stage clustering designs are briefly outlined. This section is primarily based on the methodology detailed in the book "Sampling: Design and Analysis" by Lohr (2010).

Household survey data contains information that allows us to study characteristics in a population such as child mortality rates. This information is retrieved from individual

respondents in a survey that is representative of the general population. In a survey, the observation units are the individuals, and the individuals that are included in a sample are randomly chosen. These units can either be a specific number of people or the whole unit in a given area of interest, like country, county, city, village, or similar. However, collecting samples completely at random from a population is impractical and expensive. Thus, the sample is not independently drawn. The inclusion probability, π_i, for each individual i is determined by the predefined survey design. The inclusion probability might be unequal for each individual and generally lead to the construction of sampling weights for individuals, which ensure unbiased estimation of descriptive parameters when incorporated into an analysis (Heering et al., 2010). The sampling weight, w_i for individual i, is the inverse of the inclusion probability, $w_i = 1/\pi_i$.

In survey data, the population of interest is real and fixed with N individuals or units. The particular sample that is included is a subset of the population. The sample included consists of the indices of the units of the population containing n number of units. It is not possible to compute the true population since information is only accessible from the units in the sample. A population can further be divided into strata when additional information is available, so it can be included in the sample to increase precision. A stratified sample is a sample from subgroups from the population. The subgroups are called strata and are mutually exclusive, where the estimates can be obtained from each stratum. In a stratified sample, the population of N is divided into H strata, where each stratum, h, consists of N_h sampling units. In stratified sampling, different inclusion probabilities of the units in the different strata are considered.

Cluster sampling has the same concept as a stratified sample, but the subgroups are smaller and often based on geographical location. By that, a sample from clusters limits the geographical spread of sampled observations. A unit in a population must belong to a cluster that is selected to be included in the sample. In cluster sampling, each subgroup is called a primary sampling unit (psu). Each cluster is further divided into secondary population units, a such unit that is included in the sample is called secondary sample unit (ssu).

A simple random sample (SRS) is the most basic selection process of sampling from a population and provides the theoretical basis for the more complicated forms such as one- and two-stage cluster sampling. A SRS unit is randomly and independently selected from the population, with an equal probability of being chosen. Hence, each unit i is equally weighted, $w_i = N/n$. In one-stage cluster sampling, we denote the number of clusters in the population as N and as n psus in the sample. In the cluster sampling of the simplest form, a SRS, S, of n psus is taken such that the inclusion probability is equal for each unit in the population. Let M_i and m_i denote the number of secondary population units and the number of ssus in psus i in respective order. Then, in one-stage cluster sampling, all units within the psus from an SRS are included in the sample. In other words, $M_i = m_i$ if a psu i is sampled, and $m_i = 0$ otherwise. For the two-stage clustering, let N be the number of clusters and M be the number of secondary population units from the population. For the simplest form, an SRS, S, of n psus are taken from the population, and thereafter SRSs of

m ssus from each of the n psus are conducted. These will be denoted by S_i for psus i. In two-stage clustering, the number of population units, M_i and the number of sampled units m_i in psus i can differ.

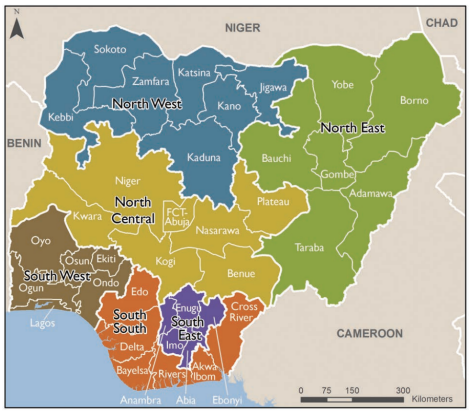
Chapter 3

Descriptive analysis of the data

To access the DHS data, an application must be submitted through the DHS website[1]. The DHS program processes the raw data into a recode data file. All variables in the recode data file are in a standardized format. The data frame must be interpreted according to the information from the associated recode data file, and subsequently extract the information that is relevant for the analysis. The standard recode manual has a format with the same structure across countries. The standardized format facilitates comparisons across surveys. Specifically, a sampling manual called "Standard Recode Manual for DHS-7" is used for the datasets collected in Nigeria and Kenya and is described in ICF et al. (2018). The models implemented for this analysis focus on mortality rates and vaccination rates at the national-, residence- and county/provincial level as the surveys provide enough data to produce direct estimates at these levels.

The DHS program collects data from responses based on questionnaires gathered during physical household visits. Data used in this analysis is designed to produce representative estimates for a majority of the survey indicators at the national level, for urban and rural residency separately, at the regional level, and for selected indicators at the county level (Kenya National Bureau of Statistics et al., 2015b). Figure 3.1a and 3.1b provides a map of Nigeria and Kenya in state and county level in respective order. Nigeria consists of total 36 states and 1 Federal Capital Territory, while Kenya has 47 counties. The states or counties were further stratified into urban and rural groups. In Nigeria, each of the 36 states and the Federal Capital Territory are separated into urban and rural residency. Thereby, 74 sampling strata in total were identified by DHS. In Kenya, each of the 47 counties were stratified, totaling 92 strata because Nairobi and Mombasa are fully urban. The data we will use can be considered as point measurements of trials and successes.

[1] https://www.dhsprogram.com



(a) Map of the 36 states of Nigeria and the Federal Capital Territory, taken from National Population Commission (NPC) (2019a), with permission from the DHS program.

(b) Map of the 47 counties of Kenya, taken from Kenya National Bureau of Statistics et al. (2015a), with permission from the DHS program.

The sampling design of DHS data is stratified followed by applying a two-stage proportional-to size sampling cluster design. “Implicit stratifications were achieved at each of the lower administrative levels by sorting the sampling frame before sample selection according to administrative order and by using a probability proportional to size selection during the first sampling stage” National Population Commission (NPC) (2019b). The strata are divided into small geographical areas called enumeration areas (EAs). An EA is the same as a psu. The DHS program selects a sufficient number of psus in each stratum. Then a fixed number of households or ssus are selected from an up-to-date list of the households for every psus. In the first stage of clustering, all ssus within the psus are included in the sample. In a two-stage cluster sample, only a subsample of the units in psus are included in the sample. The probability of sampling a psu is proportional to the estimated size of the psu. Thereby, the inclusion probability of a household is approximately proportional to the size of the EA and is approximately equal for all psus.

To ensure reliable estimates for all strata, the DHS program can oversample population groups or geographical areas when needed. Oversampling some strata (such as urban areas or selected regions or provinces) can increase the precision of estimates, besides ensuring that the sample includes the units from every stratum. Hence, the final dataset includes enough cases to produce reliable results. “In these cases, sampling weights need to be applied to account for varying design weights and non-response levels”, Hancioglu and Arnold (2013). Stratification requires well- defined and known strata. Accurate registers of inhabitants, however, might be difficult to obtain for low- and medium-income countries.

In this study, we focus on two binary responses; NMR and MCV1. We investigate how these responses separately are associated with maternal age groups, urban or rural type

of residency, and cluster effects in the countries of Nigeria and Kenya. Each row in each dataset represents a child, and the columns are the reported values of the variables from the survey questionnaire. NMR (per 1 000 live births) is defined within this analysis as the proportion of children who die within the first month after live birth. For estimating NMR, data collected over a five-year period is utilized. The DHS dataset collected from Nigeria in the period 2014-2018 is henceforth referred to as "NDHS14-18". On the other hand, the dataset collected from Kenya during 2010-2014 is referred to as "KDHS10-14". The relevant variables from the dataset for estimating NMR are presented in Table 3.1. In order to estimate mortality rates, the function `getBirths()` implemented by Li et al. (2015) is utilized. The function reformats full birth records into person-month format. Since NMR estimation acquires information of newborns within 30 days after birth, a subset is retrieved accordingly.

Table 3.1: Relevant variables from the DHS recode manuals for NMR.

v001	Cluster number
v013	Maternal age in 5-year groups from 15-49 years
v025	Type of place of residence (urban or rural)
b5	Child is alive
b7	Age at death of the child (months)

In this thesis, we study estimates of MCV1 (per 100 children) as the binary response variable for coverage among 1-year-olds (12-23 months) who have known status about whether the child has received the first dose of measles-containing vaccine or not. In MCV1 analysis, the DHS data from the countries is read through the `read_dta()` function to study the vaccination status in 2018 for Nigeria and in 2014 for Kenya. We henceforth refer to the dataset for Nigeria as "NDHS18", and the dataset for Kenya as "KDHS14". Table 3.2 shows the relevant variables that are used for estimating MCV1. We filter out the relevant information by collecting the subset of all one-year-old children that are alive.

Table 3.2: Relevant variables from the DHS recode manuals for MCV1.

v001	Primary sampling unit (cluster id)
v013	Maternal age in 5-year groups from 15-49 years
v025	Type of place of residence (urban or rural)
b8	Current age of child (years)
h9	Received MEASLES

All women aged 15-49 in a sampled household are interviewed. The maternal age groups are given as v013 $\in \{[15\text{-}19], [20\text{-}24], [25\text{-}29], [30\text{-}34], [35\text{-}39], [40\text{-}44], [45\text{-}49]\}$. In the context of NMR and MCV1, we consider one trial for each observation. A column is added to count for one trial for each observation.

3.0.1 Datasets for NMR

The NDHS14-18 has an overall representative sample of 40 427 households nationwide, but the dataset reduces to 21 555 mothers and 32 982 newborns. Among newborns within the first month of age, there are approximately 4.1% non-survivors observed in the data. The plots in Figure 3.2 from left to right illustrate the distributions of the observations, average observations per cluster, and the observed NMR between maternal age groups and within the urban and rural areas in Nigeria. The left plot of the figure shows that the majority of the observations belong to the age group of 25-29 years, and the number decreases further away from this age group. The two outermost age groups, that is age group 15-19 years and 45-49 years, have few observations compared to the observations in the remaining groups. They represent around 7% of the distribution. Moreover, observations from rural areas dominate the distribution of the observations and represent approximately 66% of the total observations.

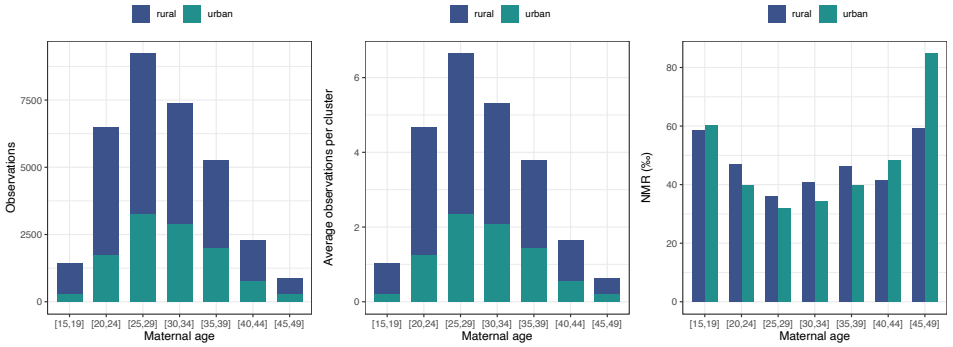


Figure 3.2: Distribution of newborns (observations) of mothers interviewed in Nigeria during 2014-2018 by maternal age groups and between urban or rural residency. NMRs is the observed values.

The center plot in Figure 3.2 shows that the distribution of average observations per cluster is similar to the distribution between the observations but on a considerably smaller scale. The NDHS14-18 dataset contains a total of 1 389 clusters. As shown in the figure, there is on average very few children from each cluster. The right plot shows that the observed NMR is relatively higher in the two outermost age groups and where urban areas have higher NMR than rural areas. However, towards the center age groups, the rural areas have slightly higher observed NMR compared to urban areas.

KDHS10-14 contains a data frame with a sample size of 40 300 households. The dataset reduces to information from 14 399 mothers for estimating NMR. The binary response data contains 19 509 births of which approximately 2.4% are non-survivors. For the extracted dataset, the sampling is split into 1 593 clusters. The plots in Figure 3.3 from left to right show the distributions of the observations, average observations per cluster, and the observed NMR between maternal age groups and within the urban and rural areas in Kenya.

As for the NDHS18 data, most of the observations belong to the age group 25-29 years, and the number of observations decreases further from this age group. The distribution of the average observations per cluster is similar to the distribution of the observations. In general, there are few observations per cluster, with a highest of just below 4 children per cluster in age group 25-29 years. In the age group, 45-49 there are approximately 0.25 children per cluster. The right plot shows that the late age groups 40-44 and 45-49 have higher observed NMRs compared to the NMRs in the remaining age groups.

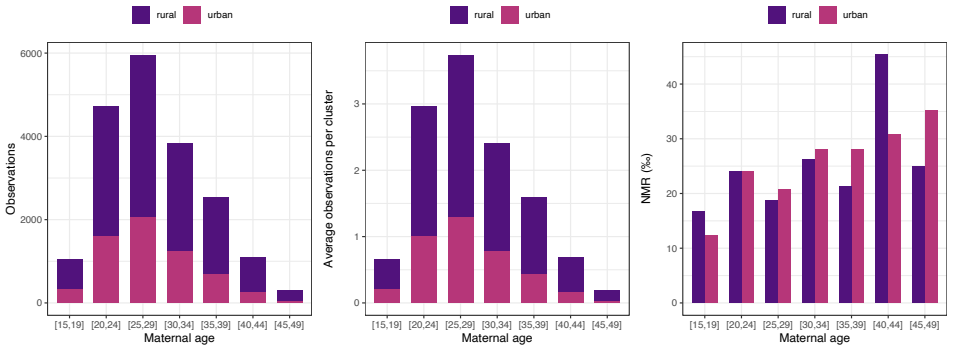


Figure 3.3: Distribution of newborns (observations) of mothers interviewed in Kenya during 2010-2014 by maternal age groups and between urban or rural residency. NMR is the observed values.

3.0.2 Datasets for MCV1

The NDHS18 data contains of total 6 059 one-year-old children where 53% of the children have received measles vaccine and the remaining children have not received measles vaccine. The plots in Figure 3.4 from left to right illustrate the distributions of the observations, average observations per cluster, and the observed MCV1 between maternal age groups and within the urban and rural areas in Nigeria.

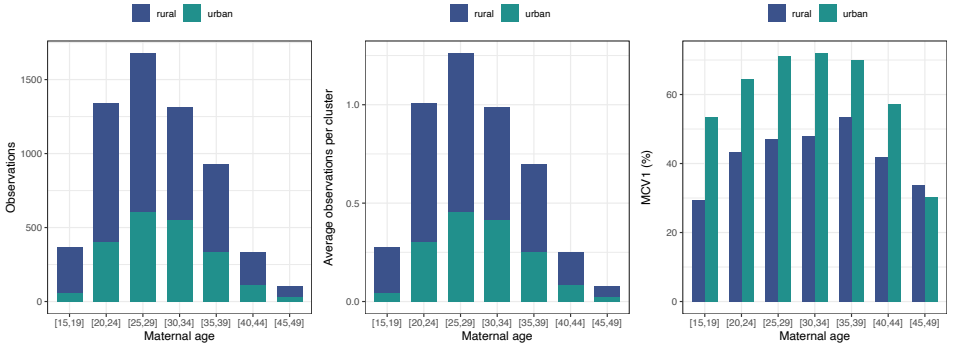


Figure 3.4: Distribution of vaccination coverage of measles among 1-year-olds (observations) of mothers interviewed in Nigeria during 2014-2018 by maternal age groups and between urban or rural residency. MCV1s are the observed values.

Observations from rural areas represent 65% of the dataset. The outermost age groups have a lower number of observations and average observations per cluster compared to the observations and in the remaining age groups. The dataset contains $1\,300$ clusters. Moreover, the observed average MCV1 is generally higher in the urban areas than for the rural areas.

The KDHS14 dataset contains of total $4\,047$ one-year-old children where approximately 85% have received the first dose of measles-containing vaccine and the remaining 15% have not received measles vaccine. The plots in Figure 3.5 from left to right illustrate the distributions of the observations, average observations per cluster, and the observed MCV1 between maternal age groups and within the urban and rural areas in Kenya. The left plot shows that the majority of roughly 70% of the observations for each age group are from rural areas as shown in the left plot of Figure 3.5. Likewise, the average observations per cluster of a total of $1\,406$ clusters have a similar distribution with the outermost age groups containing few observations compared to the remaining age groups. The observed MCV1 is generally higher for the urban areas compared to the rural areas.

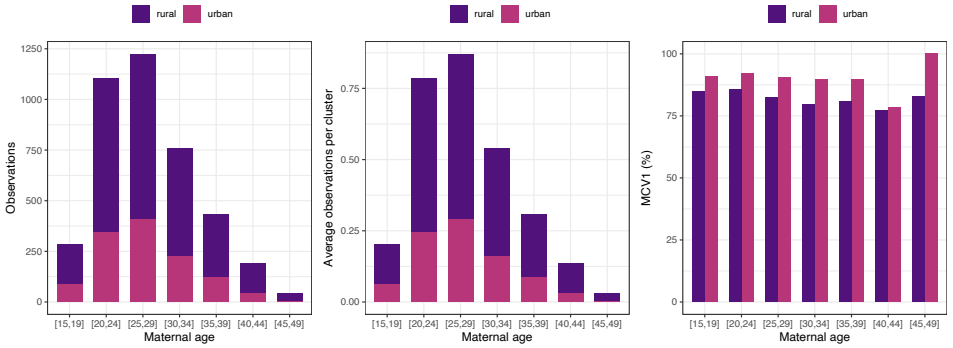


Figure 3.5: Distribution of vaccination coverage of measles among 1-year-olds (observations) of mothers interviewed in Nigeria during 2014-2018 by maternal age groups and between urban or rural residency. MCV1s are the observed values.

Chapter 4

Simulation study

The goal of this simulation study is to investigate the importance of accounting for clustering and informative inclusion probabilities in the survey design.

4.1 Main objectives

In survey data, only a small proportion of the full population is observed. Without knowing the full population, it is challenging to evaluate methods. Simulated populations allow us to compare and evaluate estimated rates obtained from the methods to the 'true' simulated rates. The answers collected through survey questionnaires from residents in the same cluster might be more similar than answers from residents in different clusters. It can therefore be assumed that the NMRs, for instance within the same cluster are more similar than rates between different clusters. Moreover, observations in the total sample might have different inclusion probabilities. In this simulation study, multiple surveys are drawn from simulated populations. The clusters in each fixed population have different predefined cluster effects. For every survey drawn from a population, scoring rules from mixed effects which accounts for correlation among clusters are compared with scoring rules from fixed effects where this assumption is ignored. The average of the scoring rules of all surveys drawn for every single population is calculated. The goal of this simulation study is to evaluate the importance of accounting for clustering and inclusion probabilities by comparing the average scoring rules to assess model performances. Simulations are constructed based on the NDHS14-18 data with NMR as a response, however, the same approach could be adopted to the other responses used for this analysis.

4.2 Simulated populations

In this study, three populations are constructed. Each population consists of $10\,000$ clusters, where $7\,500$ clusters are assigned rural and the remaining $2\,500$ are urban clusters.

Each urban or rural cluster contains 20 children divided into maternal age groups according to an approximate distribution as for the real data set. Accordingly, a data frame of 200 000 rows or children are created. A column is added to count for one trial for each observation. Let $x \in \{[15\text{-}19], [20\text{-}24], [25\text{-}29], [30\text{-}34], [35\text{-}39], [40\text{-}44], [45\text{-}49]\}$ denote the seven age groups and $n \in \{2, 4, 5, 4, 3, 1, 1\}$ be the number of children. Then, the number of children in each age group in every urban cluster and every rural cluster is described by the Cartesian product $X \times N$, that is, the set of the ordered pairs (x, n) with $x \in X, n \in N$. The clusters are individually simulated with a cluster-specific effect $\gamma_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\gamma^2)$ for $i = 1, \ldots, 10\,000$. Let the linear predictor $\tilde{\eta}_{ij} = \alpha_{\text{age}[i,j]} + \beta_{\text{area}[i]}$ describes the coefficients of a fixed effects model estimated from the real data set using `INLA` in R. Here age$[i, j]$ denotes the maternal age group in cluster i for child j, while area$[i]$ denotes either urban or rural area in cluster i. Assuming $\tilde{\mu}_{ij} = \text{expit}(\tilde{\eta}_{ij}) \sim \text{Bernoulli}(\tilde{\mu}_{ij})$, the coefficients are

$$[\alpha_1 + \beta_u, \alpha_2 + \beta_u, \ldots, \alpha_7 + \beta_u] = [\text{-}2.85, \text{-}3.13, \text{-}3.40, \text{-}3.29, \text{-}3.14, \text{-}3.16, \text{-}2.68]$$
and
$$[\alpha_1 + \beta_r, \alpha_2 + \beta_r, \ldots, \alpha_7 + \beta_r] = [\text{-}2.75, \text{-}3.03, \text{-}3.30, \text{-}3.19, \text{-}3.04, \text{-}3.06, \text{-}2.59],$$

denoting the $7 \cdot 2$ categories for age groups, $(\alpha_1, \alpha_2, \ldots, \alpha_7)$, in urban area for β_u or rural area for β_r in cluster i. For simulating the binary outcome for child j in cluster i, the cluster effects, γ_i is included in the linear predictor,

$$\eta_{ij} = \alpha_{\text{age}[i,j]} + \beta_{\text{area}[i]} + \gamma_i. \tag{4.1}$$

The expected value of the response variable $y_{ij} = \text{E}[Y_{ij}|\mu_{ij}] \sim \text{Bernoulli}(\mu_{ij})$ is simulated as $\mu_{ij} = (1 + \exp(-\eta_{ij}))^{-1}$ from (2.5). The three fixed populations differ with the characteristics of the distribution of the cluster-specific effect, drawn $\gamma_i \overset{iid}{\sim} \mathcal{N}(0, \sigma_\gamma^2)$. The populations, $p = \{p_1, p_2, p_3\}$, have 'true' standard deviation of the cluster effects, $\sigma_\gamma = \{0, 0.5, 1\}$, respectively.

4.3 Survey designs

Surveys are drawn from the simulated populations each containing 300 clusters from rural areas and 100 clusters from urban areas, corresponding to a total of 8 000 children. The simulations consist of every child from the clusters that are included in the survey. The inclusion probability of clusters with positive cluster effects or high NMR estimates is proportional to the inclusion probability of the remaining clusters with a proportional constant k, such that in each simulation $j = 1, 2, \ldots, N$ from a population

$$k \cdot P(i \text{ included} | \gamma_i > 0) = P(i \text{ included} | \gamma_i \leq 0). \tag{4.2}$$

In this analysis, two different sampling designs are considered. That is sampling design with both equal- or unequal inclusion probabilities of clusters. Simulations with $k = \{1, 2, 3, 4\}$ are explored. For $k = 1$, the inclusion probability is the same for all children, while for $k > 1$, different inclusion probabilities of clusters are considered. In this thesis, surveys are drawn from all three populations, p_1, p_2 and p_3 for equal inclusion probabilities. In addition, simulations on sampling design with unequal inclusion probabilities of clusters with $k = \{2, 3, 4\}$ are considered for population p_3. All six scenarios are given in Table 4.1. The highlighted cells in the table show the surveys drawn from an informative sampling design. The number of surveys N for every scenario is predetermined to a number large enough. For this analysis $N = 400$ surveys are performed for each scenario on the fixed populations.

Table 4.1: Surveys are drawn from the following scenarios for populations defined in Section 4.2 and k from (4.2). The highlighted cells shows the surveys drawn from sampling design of unequal inclusion probabilities of clusters.

Scenario	Population	σ_γ	k
1	p_1	0	1
2	p_2	0.5	1
3	p_3	1	1
4	p_3	1	2
5	p_3	1	3
6	p_3	1	4

4.4 Models and scoring rules

The simulation study aims to evaluate predictive strengths of the methods based on scoring rules of a GLM model and a GLMM model. The models are fitted with the package `INLA` in R, and which estimates posteriors with integrated nested Laplace approximations as described in Section 2.4.

Let $i = 1, \dots M$ denote the clusters for child $j = 1, \dots n_i$ and let age[i, j] represent the maternal age group and let area$[i]$ represent the type of residency (urban or rural) in cluster i for child j. The GLM model which is referred to as **M1** is

$$\textbf{M1}\text{: } y_{ij}|\mu_{ij} \sim \text{Bernouilli}(\mu_{ij}), \quad \mu_{ij} = \text{expit}(\alpha_{\text{age[i, j]}} + \beta_{\text{area}[i]}),$$

for $\alpha_1, \dots \alpha_7$. The GLMM model has additionally the cluster effects γ_i for cluster i as random effects in the model. The model is referred to as **M2** and is given following the distribution of the intercept and the priors of the hyperparameters

$$\textbf{M2}: y_{ij}|\mu_{ij} \sim \text{Bernouilli}(\mu_{ij}), \quad \mu_{ij} = \text{expit}(\alpha_{\text{age[i, j]}} + \beta_{\text{area[}i\text{]}} + \gamma_i)$$
$$\gamma_i|\sigma^2_\gamma \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_\gamma)$$
$$\frac{1}{\sigma^2_\gamma} \sim \text{Gamma}(1, 5 \cdot 10^{-5})$$

where the coefficients $(\alpha_1, \ldots, \alpha_7, \beta_u, \beta_r) \overset{iid}{\sim} \mathcal{N}(0, 1000)$, and $\gamma_1, \ldots \gamma_M|\sigma^2_\gamma \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_\gamma)$ are the cluster effects with variance σ^2_γ. INLA assigns by default the priors of Gamma$(1, 5 \cdot 10^{-5})$ for the precision of the random effects γ.

The true simulation rates are determined by the empirical proportion of non-survived children within the maternal age group and urban or rural areas. The prediction rates and true simulation rates for are used to calculate the scoring rules of MBE, MSE, and coverage presented in Section 2.5 for **M1** and **M2** from every 400 survey drawn from each population. The averages of the scoring rules obtained from the surveys are determined using (2.19). Moreover, (2.21) is used to evaluate the MSD value in **M2** for each survey.

4.5 The importance of accounting for clustering

Simulations from sampling design of equal inclusion probabilities for all three populations are studied. That is, scenarios $1, 2$, and 3 from Table 4.1. Scoring rules for **M1** and **M2** are compared. Notations MBE_{M1}, MSE_{M1} and $\text{Coverage}_{\text{M1}}$ are used for the average scoring rules of **M1**, while MBE_{M2}, MSE_{M2} and $\text{Coverage}_{\text{M2}}$ are the average scoring rules of **M2**. The average MSD for all surveys obtained from **M2** are denoted as $\text{MSD}_{\text{M2}}(\hat{\sigma}_\gamma)$.

The most prominent difference between the scoring rules of **M1** and **M2** are from population $p3$ and is presented in Table 4.2. The table shows the average scoring rules for every seven age groups and within urban or rural areas. MBE and MSE measures are relatively similar under both models. In particular, the assessment of the performance of the models cannot be differentiated based on the MSE values. Although the MBE values of the models are somehow different, it is of a minimal scale that may be neglected. The most interesting difference, however, is the coverage for the models. In general, **M1** has lower coverage in each age group within urban and rural areas compared to the coverage of **M2**. A lower coverage indicates that the observed values are covered fewer times relative to higher coverage, because of narrower credible intervals. Since **M1** does not incorporate the cluster effects in the model, the standard deviation of the estimates is in general narrower. On the other hand, high coverage in **M2** is a result of wider credible intervals, as the effective sample size decreases by assuming correlation within the clusters. In general, **M2** gives better scoring rules and indicates that it is an advantage to incorporate the cluster effects in the model. The MSD for **M2** for the estimated cluster effects is $\text{MSD}_{\textbf{M2}}(\hat{\sigma}_\gamma) = -0.052$ or in other words, **M2** underestimate the cluster effects by approximately 5%.

Table 4.2: Scoring rules over 400 simulations with $k = 1$ in (4.2) and p_3 ($\sigma_\gamma = 1$) for estimating NMR associated with maternal age groups and urban or rural area in Nigeria with **M1** as as fixed effects model, while **M2** also includes cluster effects as iid random effects. The coverage are calculated from a 95% credible interval. $\text{MSD}_{\mathbf{M2}}(\hat{\sigma}_\gamma) = -0.052$.

Maternal age	Residency	MBE_{M1} (10^{-4})	MBE_{M2} (10^{-4})	MSE_{M1} (10^{-5})	MSE_{M2} (10^{-5})	$\text{Coverage}_{\text{M1}}$	$\text{Coverage}_{\text{M2}}$
15-19 years	urban	0	-10	19	18	0.865	0.932
	rural	-1	-10	11	10	0.840	0.920
20-24 years	urban	5	-3	6	6	0.858	0.938
	rural	-25	-34	9	9	0.820	0.912
25-29 years	urban	-5	-14	10	10	0.877	0.945
	rural	3	-7	17	17	0.935	0.940
30-34 years	urban	-44	-54	32	32	0.920	0.943
	rural	-2	-11	12	12	0.935	0.950
35-39 years	urban	4	-4	5	5	0.917	0.953
	rural	-4	-11	3	3	0.927	0.963
40-44 years	urban	9	2	4	4	0.925	0.963
	rural	2	-6	6	6	0.950	0.968
45-49 years	urban	27	16	17	17	0.938	0.943
	rural	19	9	24	23	0.940	0.935

It is difficult to differentiate the models for populations p_1 and p_2. The similar tables of MBE, MSE and coverage of **M1** and **M2**, including the $\text{MSD}_{\mathbf{M2}}(\hat{\sigma}_\gamma)$ from simulations of populations p_1 and p_2 are therefore given in Table A.1 and A.2, respectively, in the Appendix.

4.6 The impact of informative sampling

The framework of the real data consists of clusters with different inclusion probabilities. Thus, scoring rules from simulations with different inclusion probabilities are considered. By oversampling clusters with positive cluster effects, we manipulate to overestimate the mortality rate. That is the scoring rules for scenarios $4, 5$, and 6 from Table 4.1. The scoring rules with the most extreme case of $k = 4$ are presented in Table 4.3. Notations MBE_{M1}, MSE_{M1} and $\text{Coverage}_{\text{M1}}$ are used for the average scoring rules of **M1**, while MBE_{M2}, MSE_{M2} and $\text{Coverage}_{\text{M2}}$ are the average scoring rules of **M2**. The average MSD for all surveys obtained from **M2** are denoted as $\text{MSD}_{\mathbf{M2}}(\hat{\sigma}_\gamma)$.

The most interesting difference between the scoring rules of **M1** and **M2** are from population $p3$ and is presented in Table 4.3. The table shows the average scoring rules for every seven age groups and within urban or rural area. The results shows that by oversampling clusters with higher mortality rates increases the MBE considerably in both models relatively to the respective MBE values in Table 4.2. The error rates of the MBE and MSE measures are indeed relatively similar for **M1** and **M2**. This implies that mixed models should be avoided if we suspect informative sampling design. According to the coverage, however, **M2** performs better, but underestimate the cluster effects with roughly 5% as $\text{MSD}_{\mathbf{M2}}(\hat{\sigma}_\gamma) = -0.048$. The results from Table 4.3 shows that the model with cluster

effects performs better for the overall scoring rules. Moreover, the cluster effects does not result in any negative impact.

Table 4.3: Scoring rules over 400 simulations with $k = 4$ in (4.2) and $\sigma_\gamma = 1$ for estimating NMR associated with maternal age groups and urban or rural area in Nigeria with **M1** as as fixed effects model, while **M2** also includes cluster effects as iid random effects. The coverage are calculated from a 95% credible interval. $\text{MSD}_{\textbf{M2}}(\hat{\sigma}_\gamma) = -0.048$.

Maternal age	Residency	MBE_{M1} (10^{-4})	MBE_{M2} (10^{-4})	MSE_{M1} (10^{-5})	MSE_{M2} (10^{-5})	$\text{Coverage}_{\text{M1}}$	$\text{Coverage}_{\text{M2}}$
15-19 years	urban	27	23	17	16	0.890	0.938
	rural	16	11	9	9	0.897	0.960
20-24 years	urban	5	1	6	6	0.870	0.940
	rural	-5	-10	7	7	0.920	0.958
25-29 years	urban	9	5	11	10	0.892	0.948
	rural	18	12	20	19	0.907	0.920
30-34 years	urban	-116	-121	42	42	0.848	0.875
	rural	-2	-12	12	12	0.922	0.935
35-39 years	urban	-7	-17	5	5	0.943	0.960
	rural	-3	-11	3	3	0.917	0.940
40-44 years	urban	0	-8	5	4	0.907	0.948
	rural	-10	-19	7	7	0.927	0.935
45-49 years	urban	-11	-20	16	16	0.950	0.950
	rural	43	31	24	23	0.958	0.955

The difference in the scoring rules for **M1** and **M2** is too minimal to draw any clear conclusion about which model is preferable according to the scoring rules for scenarios 4 and 5. The scoring rules for the models are given in Table A.3 and A.4, respectively, in the Appendix.

Chapter 5

Data analysis

In this section, the results of the probability estimates of NMR and MCV1 at the national level associated with maternal age, place of residency (urban or rural), and cluster effects are studied. Different models are used, of which either fixed effects or random effects for maternal age groups are compared. The cluster effects are included or excluded as iid random effects. The urban or rural covariate is eventually added as fixed effects. We introduce the models used for evaluating the estimates. Then the results of the estimates of NMR in Nigeria and Kenya in respective order are presented. Further, the results of estimates of MCV1 for the countries in the same order are given.

5.1 Model descriptions

For binary outcomes, we assume

$$y_{ij}|\mu_{ij} \sim \text{Bernouilli}(\mu_{ij}), \quad \mu_{ij} = \text{expit}(\eta_{ij}),$$

from (2.5) and where $i = 1, \ldots, M$ denotes clusters for child $j = 1, \ldots, n_i$. The cluster effects are $\gamma_i|\sigma^2_\gamma \overset{iid}{\sim} \mathcal{N}(0, \sigma^2_\gamma)$, with priors, $1/\sigma^2_\gamma \sim \text{Gamma}(1, 5 \cdot 10^{-5})$. Further, let age$[i, j]$ and area$[i]$ represent the maternal age groups and the residency category, respectively, for individual j in cluster i. Let $\alpha_{\text{age}[i,j]}$ and $\beta_{\text{area}[i]}$ be the coefficients in the fixed effects models with Gaussian prior on $(\alpha_1, \ldots, \alpha_7, \beta_u, \beta_r) \sim \mathcal{N}(0, 1000)$ unless otherwise specified.

The models with maternal age groups and cluster information as categorical covariates are

$$\begin{aligned} m1_\alpha \text{ is } \eta_{ij} &= \alpha_{\text{age}[i,j]} \\ m1_{\alpha+\gamma} \text{ is } \eta_{ij} &= \alpha_{\text{age}[i,j]} + \gamma_i \\ m1_{r(\alpha)+\gamma} \text{ is } \eta_{ij} &= \alpha_{\text{age}[i,j]} + \gamma_i, \quad (\alpha_1, ..., \alpha_7)|\mu_{ij} \sim \mathcal{N}(0, \mathbf{\Sigma}) \end{aligned} \tag{5.1}$$

where $\boldsymbol{\Sigma} = (\boldsymbol{Q}/\tau)^{-1}$ denotes the covariance matrix, τ^2 i a generic variance parameter, and $\boldsymbol{Q}$ is a precision matrix of iid random effects that defines the dependence structure as in (2.9). $m1_{\alpha}$ and $m1_{\alpha+\gamma}$ are compared to observe the difference in the estimates when cluster-specific effects are excluded or included.

Incorporating the cluster effects in the model results in a shift in the fixed effects that depends on the true value. For a true value, y, of small quantity, the expit function can be approximated as $\text{expit}(y) = \frac{e^y}{1+e^y} \approx e^y$. Let $\eta_{ij} = \alpha_{\text{age}[i,j]}, (\alpha_1, \ldots, \alpha_7)$ be the estimates of a fixed effects model with maternal age groups as covariate. When including the cluster information as random effects,

$$\text{E}[e^{\alpha_k+\gamma_i}|\alpha_k, \sigma^2_\gamma] = e^{\alpha_k+\sigma^2_\gamma/2} \tag{5.2}$$

for small values of α_k will decrease estimates of α_k when the cluster effects are included. This is a result of a non-linear link function, where intercept and variance will affect each other. Moreover, we expect a decrease in the effective sample size when accounting for correlation among children within the same cluster. For example, a sample size of, say, 1000 children will only represent the amount of information for around 500, or perhaps 200 children. This results in an increase in the uncertainly of estimates of $m1_{\alpha+\gamma}$, because the model accounts for correlation among clusters. Now, $m1_{\alpha+\gamma}$ and $m1_{r(\alpha)+\gamma}$ are compared to observe the difference in the estimates using fixed effects or random effects on maternal age groups as covariate.

To explore the impact of type of residency on NMR or MCV1 in addition to the maternal age groups and cluster effects, we model

$$\begin{aligned} m2_{\alpha+\beta+\gamma} \text{ is } \eta_{ij} &= \alpha_{\text{age}[i,j]} + \beta_{\text{area}[i]} + \gamma_i \\ m2_{r(\alpha)+\beta+\gamma} \text{ is } \eta_{ij} &= \alpha_{\text{age}[i,j]} + \beta_{\text{area}[i]} + \gamma_i, \quad (\alpha_1, ..., \alpha_7)|\mu_{ij} \sim \mathcal{N}(0, \boldsymbol{\Sigma}) \end{aligned} \tag{5.3}$$

where $\boldsymbol{\Sigma} = (\boldsymbol{Q}/\tau_{rw})^{-1}$ denotes the covariance matrix of rw1- or rw2 random effects described in Section 2.3. By further disaggregating the maternal age groups on urban and rural residency, the observations are distributed into more definitive groups, which results in a smaller number of observations in each group. As discussed in Section 3, the distributions of the datasets used for this analysis show that the outermost maternal age groups have fewer observations relative to the remaining groups. Random effects of rw1 and rw2 are therefore expected to give more accurate estimates, because they have properties that allow collecting information from neighboring groups when there are few observations in the respective groups.

In this analysis, the CRPS values are measured for assessing the predictive performance of the methods. In particular, each age group, α_k, for $k = 1, \ldots, 7$ is left out and the remaining data is used to make predictions. Then, the posteriors of $\text{E}[\text{expit}(\alpha_k + \beta_u + \gamma)|\alpha_k, \beta_u]$

and E[expit$(\alpha_k + \beta_r + \gamma)|\alpha_k, \beta_r]$ are computed, i.e, we marginalize out the cluster effects. The age groups $(\alpha_1, ..., \alpha_7)|\mu_{ij} \sim \mathcal{N}(0, \boldsymbol{\Sigma})$ where $\boldsymbol{\Sigma} = (\boldsymbol{Q}/\tau_{rt})^{-1}$ denotes the covariance matrix of iid-, rw1- or rw2 random effects described in Section 2.3. The empirical proportions for each age groups within urban or rural areas are then computed. Finally, the CRPS values are computed based on the predictive distributions and the observed values. The CRPS is calculated at expit-scale in per 1000, and where the lowest CRPS value is the most favorable model according to CRPS criteria.

5.2 Analysis of NMR coverage

5.2.1 Nigeria

Figure 5.1 shows the variation in NMR estimates between maternal age groups in Nigeria on an expit-scale along with 95% credible interval. The left plot shows the estimates of $m1_{\alpha}$ and $m1_{\alpha+\gamma}$, and the right plot shows estimates of $m1_{\alpha+\gamma}$ and $m1_{r(\alpha)+\gamma}$ defined in (5.1).

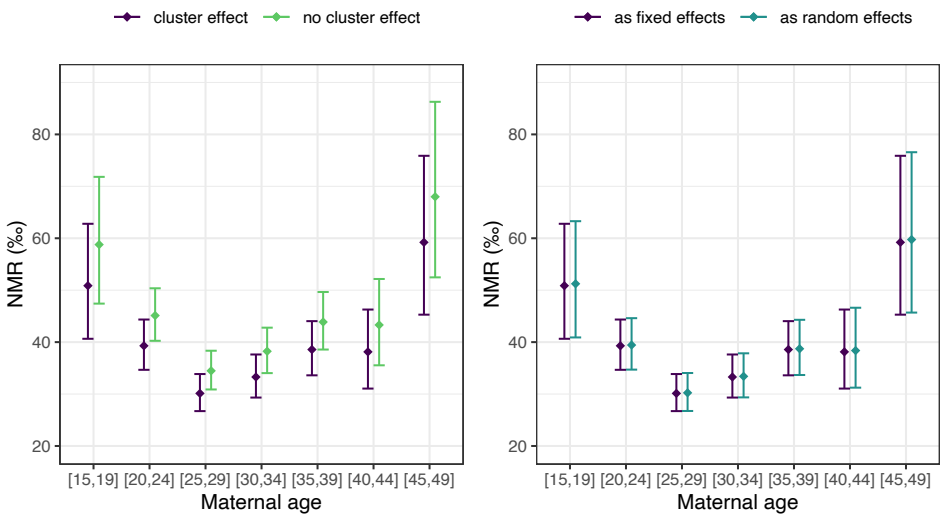


Figure 5.1: Variations in NMR estimates between maternal age groups in Nigeria. The left plot shows the estimates of expit(α_{age}) and expit$(\alpha_{\text{age}} + \gamma_{iid})$, and the right plot shows estimates of expit$(\alpha_{\text{age}} + \gamma_{iid})$ and expit$(\alpha_{\text{age}} + \gamma_{iid})$, $(\alpha_1, ..., \alpha_7)$ as iid random effects described in (5.1). A 95% credible interval is given.

The left plot in the figure shows that the probability estimates of NMR associated with maternal age groups and cluster effects have lower fixed effects in general compared to the model that has no cluster-specific effects, but with a similar range in variance relative to each other. Accordingly, there has been a shift between the cluster effects model and the model with no cluster effects as expressed in (5.2). The 95% credible interval of the

clustering effects is calculated to be $\hat{\sigma}_c = [0.42, 0.60]$ with median 0.50. The right plot shows that when cluster effects is included in both fixed effects and iid random effects models, the estimates NMR associated with maternal age groups are very similar.

The variation between maternal age groups within urban and rural areas in NMR for models $m2_{\alpha+\beta+\gamma}$ and $m2_{r(\alpha)+\beta+\gamma}$ are given in Figure 5.2. A 95% credible interval is given. The plot shows that the mortality rate is slightly higher for rural areas compared to urban areas. The figure illustrates the advantage of assigning maternal age groups as rw1 and rw2 random effects, as the outermost age groups with few data are shrunk towards the mean of the estimates of the neighborhood groups. Thereby, the variation between maternal age groups is less pronounced in the random effects models compared to the fixed effects model. Despite having relatively high mortality estimates in these age groups in the fixed effects model, there is not enough information to conclude that they are convincingly different from the other age groups. The figure shows that the estimates from rw1 and rw2 models stabilize more between the age groups compared to the fixed effects model. If a line were drawn between the medians for each model as in Figure 2.1, the smoothness of the lines would have increased in the order of fixed, rw1 and rw2.

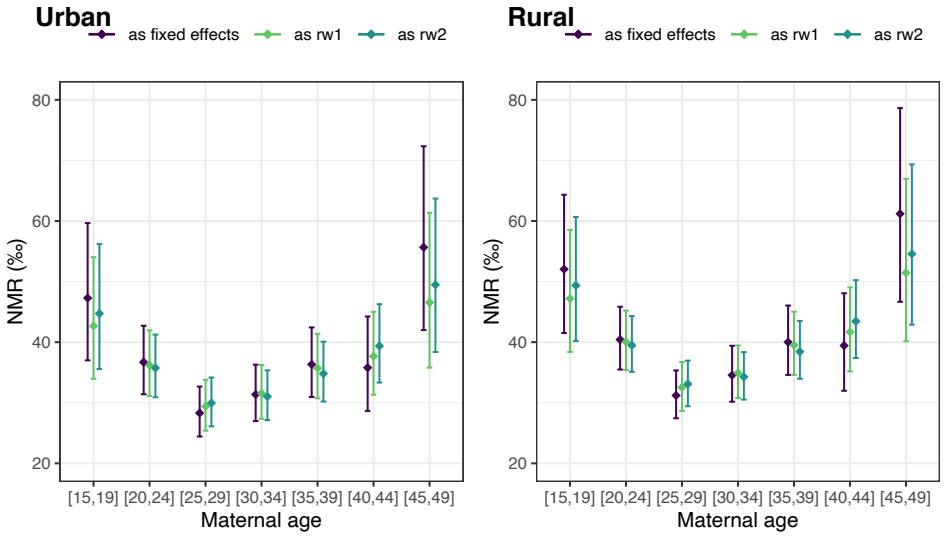


Figure 5.2: Variations in NMR estimates between maternal age groups within urban or rural areas in Nigeria. The figure shows estimates of expit$(\alpha_{\text{age}} + \beta_{\text{area}} + \gamma_{iid})$, with α_{age} as iid, rw1 or rw1 random effects described in (5.3). The estimates show a 95% credible interval.

The CRPS measures for predicting NMR in missing age groups in urban or rural area in Nigeria are given in Table 5.1 $\overline{\text{CRPS}}$ shows the average CRPS in each category. The highlighted cells in the table show the model type with the best CRPS value and best average CRPS in total for both urban and rural areas. According to the table, the model type of rw2 has the average best score of all groups and the best score for most age groups.

Although rw2 is best for most age groups, the other models are better for some of the age groups.

Table 5.1: CRPS values for predictions of NMR in missing age groups in urban or rural area in Nigeria in expit-scale in per 1000. $\overline{\text{CRPS}}$ are the average values. The highlighted cells are the model types with lowest (most favorable) CRPS.

Maternal age	Residency	iid	rw1	rw2	$\overline{\text{CRPS}}$
15-19 years	urban	18.71	18.19	14.12	17.01
	rural	14.57	14.05	9.44	12.69
20-24 years	urban	2.68	2.90	3.54	3.04
	rural	4.26	5.48	7.63	5.79
25-29 years	urban	3.02	1.79	1.20	2.00
	rural	2.98	1.78	1.02	1.93
30-34 years	urban	2.73	1.48	2.07	2.09
	rural	2.19	2.78	4.78	3.25
35-39 years	urban	2.64	3.65	4.94	3.74
	rural	3.98	6.14	8.64	6.25
40-44 years	urban	7.98	4.43	3.62	5.34
	rural	2.38	3.26	3.92	3.19
45-49 years	urban	43.97	43.03	38.05	41.68
	rural	15.52	14.17	11.06	13.58
$\overline{\text{CRPS}}$	urban	11.68	10.78	9.65	
	rural	6.53	6.81	6.64	

5.2.2 Kenya

In Kenya, the NMR estimates are not affected by the inclusion of cluster effects as shown in the left plot of Figure 5.3. Indeed, the 95% credible interval of the cluster effects is $\hat{\sigma}_c = [0, 0.04]$, with median 0.03. The estimates of $m1_{\alpha+\gamma}$ and $m1_{r(\alpha)+\gamma}$ are also very similar as shown in the right plot of the figure. The NMR estimates from $m2_{\alpha+\beta+\gamma}$ against $m2_{r(\alpha)+\beta+\gamma}$ in Figure 5.4 show that the variation between the groups is less pronounced in the random effects model compared to the fixed effects model. The maternal age groups in the fixed effects model are treated as distinct groups, which makes the variance of outermost age groups wider as they have very few observations. The random effects model handles these groups by collecting strength from neighboring groups. The variation between maternal age groups is less pronounced in the order of fixed effects, rw1 and rw2, for the same reason as discussed for Nigeria. There is a clear difference in the variance of the estimates for each model. The variances in the random effects models compared with the fixed effects model are smaller as expected.

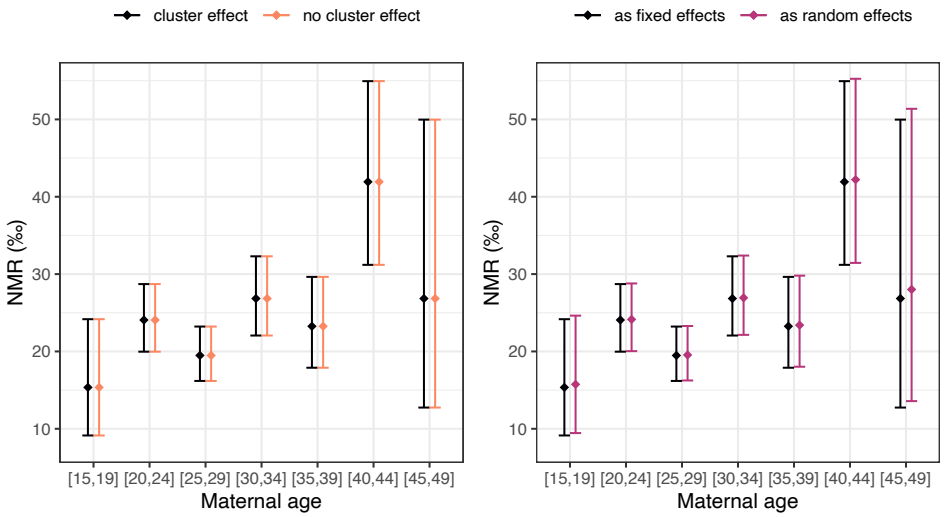


Figure 5.3: Variations in NMR estimates between maternal age groups in Kenya. The left plot shows the estimates of expit(α_{age}) and expit($\alpha_{\text{age}} + \gamma_{iid}$), and the right plot shows estimates of expit($\alpha_{\text{age}} + \gamma_{iid}$) and expit($\alpha_{\text{age}} + \gamma_{iid}$), ($\alpha_1, ..., \alpha_7$) as iid random effects described in (5.1). A 95% credible interval is given.

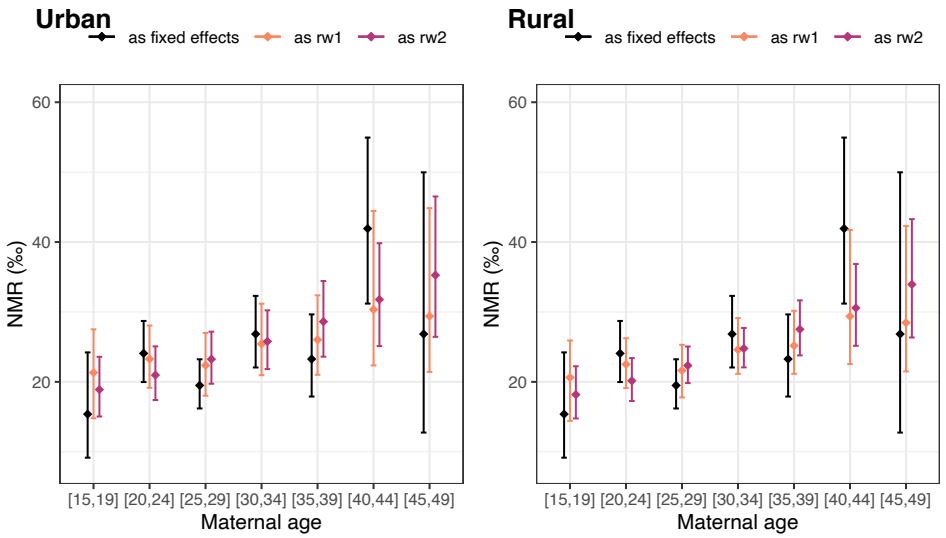


Figure 5.4: Variations in NMR estimates between maternal age groups within urban or rural areas in Kenya. The figure shows estimates of expit($\alpha_{\text{age}} + \beta_{\text{area}} + \gamma_{iid}$), with α_{age} as iid, rw1 or rw1 random effects described in (5.3). The estimates show a 95% credible interval.

The CRPS measures for predicting NMR in missing age groups in urban or rural area for in Kenya are given in Table 5.2. The highlighted cells in the table show the model type with the lowest CRPS in total for each age group. $\overline{\text{CRPS}}$ shows the average CRPS for the age groups in urban or rural areas and each model type. The best CRPS on average and for most age groups are the model with rw2 random effects.

Table 5.2: CRPS values for predictions of NMR in missing age groups in urban or rural area in Kenya in expit-scale in per 1000. $\overline{\text{CRPS}}$ are the average values. The highlighted cells are the model types with lowest (most favorable) CRPS.

Maternal age	Residency	iid	rw1	rw2	$\overline{\text{CRPS}}$
15-19 years	urban	7.77	7.49	3.98	6.41
	rural	4.90	4.74	1.56	3.73
20-24 years	urban	1.87	2.41	3.79	2.69
	rural	1.64	3.16	5.32	3.37
25-29 years	urban	3.04	2.92	2.26	2.74
	rural	4.97	5.18	3.73	4.63
30-34 years	urban	2.53	3.00	1.93	2.49
	rural	2.18	1.99	1.15	1.77
35-39 years	urban	2.60	1.94	1.81	2.12
	rural	2.12	5.08	6.00	4.40
40-44 years	urban	4.57	4.32	3.03	3.97
	rural	20.26	20.17	15.84	18.76
45-49 years	urban	7.56	6.12	5.62	6.43
	rural	2.66	3.12	5.83	3.87
$\overline{\text{CRPS}}$	urban	4.28	4.03	3.20	
	rural	5.53	6.21	5.63	

5.3 Analysis of MCV1 coverage

5.3.1 Nigeria

The variation in estimates of the MCV1 coverage in Nigeria for $m1_{\alpha}$ are in general higher compared to the relative estimates in $m1_{\alpha+\gamma}$ as shown in the left plot of Figure 5.5. The uncertainty in the estimates of the cluster effects model from a 95% credible interval is generally wider, and especially for the outermost age groups. This is a result of the fact that correlation across children from the same clusters has been taken into account. The right plot in the figure shows the estimates of the vaccination rate from $m1_{\alpha+\gamma}$ and $m1_{r(\alpha)+\gamma}$. The model with maternal age groups as random effects have slightly narrower uncertainty in the estimates. Moreover, the outermost age groups do in particular have higher vaccination rate estimates relative to the rate of the fixed effects model. The random effects model collects information from the remaining groups to limit the variation. There is, however, no big difference between the estimates between the fixed effects and the random effects model. Nevertheless, the impact of the type of residency on MCV1 in Nigeria is remarkably strong as shown in Figure 5.6.

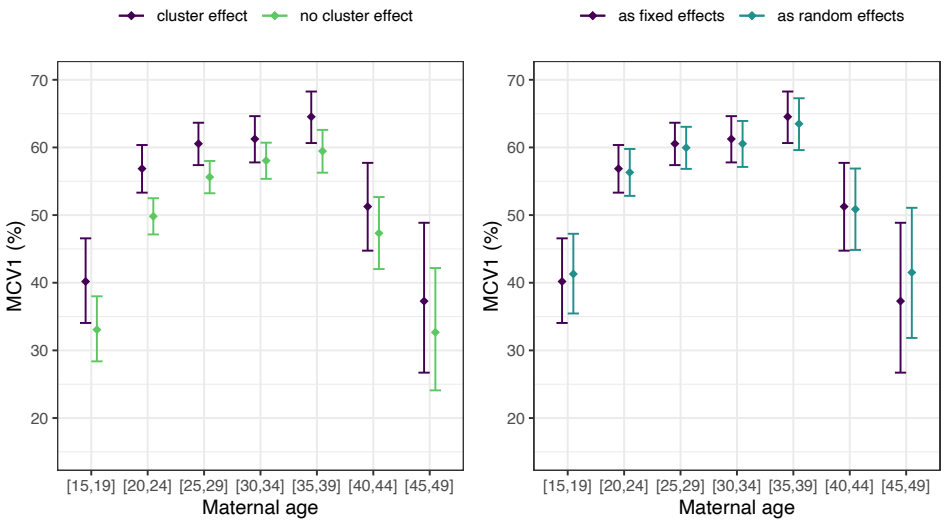


Figure 5.5: Variations in MCV1 estimates among 1-year-olds between maternal age groups in Nigeria. The left plot shows the estimates of expit(α_{age}) and expit$(\alpha_{\text{age}} + \gamma_{iid})$, and the right plot shows estimates of expit$(\alpha_{\text{age}} + \gamma_{iid})$ and expit$(\alpha_{\text{age}} + \gamma_{iid})$, $(\alpha_1, ..., \alpha_7)$ as iid random effects described in (5.1). A 95% credible interval is given.

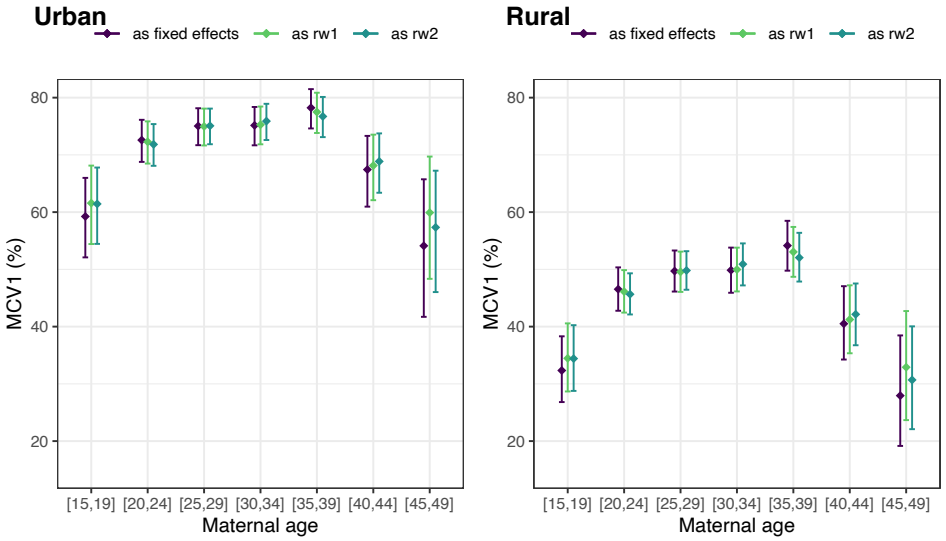


Figure 5.6: Variations in MCV1 estimates among 1-year-olds between maternal age groups within urban or rural areas in Nigeria. The figure shows estimates of expit$(\alpha_{\text{age}} + \beta_{\text{area}} + \gamma_{iid})$, with α_{age} as iid, rw1 or rw1 random effects described in (5.3). The estimates show a 95% credible interval.

The coverage of MCV1 in urban areas are higher compared to in rural areas. In urban areas, the vaccination rate gives estimates above 40% for children of all mothers for both models, while the estimates in rural areas show all under 60% from Figure 5.6. Indeed, the age groups 19-39 years in urban areas show median estimates around $70\% - 80\%$. On the other hand, the estimates for the very same age groups in rural areas show median estimates of around $45\% - 55\%$. The uncertainty in the fixed effects in the outermost age groups are wider because of the few data to collect enough information in these age groups.

The following table shows CRPS measures for predicting MCV1 coverage in missing age groups in urban or rural area in Nigeria. The rw1 and rw2 models are the preferable models based on CRPS. However, overall, the rw2 model performs better on average, and for most age groups.

Table 5.3: CRPS values for predictions of MCV1 coverage in missing age groups in urban or rural area in Nigeria in expit-scale in per 100. $\overline{\text{CRPS}}$ are the average values. The highlighted cells are the model types with lowest (most favorable) CRPS.

Maternal age	Residency	iid	rw1	rw2	$\overline{\text{CRPS}}$
15-19 years	urban	15.19	15.00	12.53	14.24
	rural	14.11	13.91	11.03	13.02
20-24 years	urban	3.37	2.13	2.38	2.63
	rural	2.57	2.33	1.57	2.16
25-29 years	urban	2.60	1.44	1.70	1.91
	rural	3.06	1.70	1.99	2.25
30-34 years	urban	2.90	1.86	3.40	2.72
	rural	3.59	2.18	4.09	3.29
35-39 years	urban	1.77	1.53	1.85	1.72
	rural	7.50	6.27	5.78	6.52
40-44 years	urban	8.80	8.19	10.25	9.08
	rural	2.56	1.88	1.64	2.03
45-49 years	urban	35.68	32.83	27.91	32.14
	rural	7.60	5.89	3.15	5.55
$\overline{\text{CRPS}}$	urban	10.04	9.00	8.57	
	rural	5.86	4.88	4.18	

5.3.2 Kenya

The left plot in Figure 5.7 shows the estimates of MCV1 coverage in Kenya associated with maternal age groups and when the cluster effects are included or excluded. The uncertainty in the estimates in groups 40-44 and 45-49 years are relatively wider under the models because of few data. The right plot in Figure 5.7 shows no strong difference in the estimates between models $m1_{f(\alpha)+\gamma}$ and $m1_{r(\alpha)+\gamma}$. The MCV1 coverage is generally higher in urban area, but the association with type of residency is not as strong as in Nigeria.

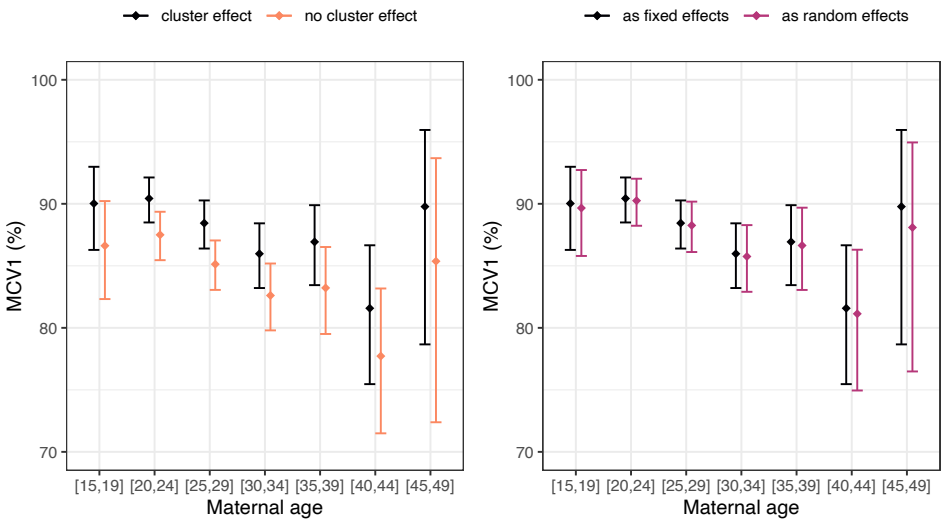


Figure 5.7: Variations in MCV1 estimates among 1-year-olds between maternal age groups in Kenya. The left plot shows the estimates of expit(α_{age}) and expit($\alpha_{\text{age}} + \gamma_{iid}$), and the right plot shows estimates of expit($\alpha_{\text{age}} + \gamma_{iid}$) and expit($\alpha_{\text{age}} + \gamma_{iid}$), ($\alpha_1, ..., \alpha_7$) as iid random effects described in (5.1). A 95% credible interval is given.

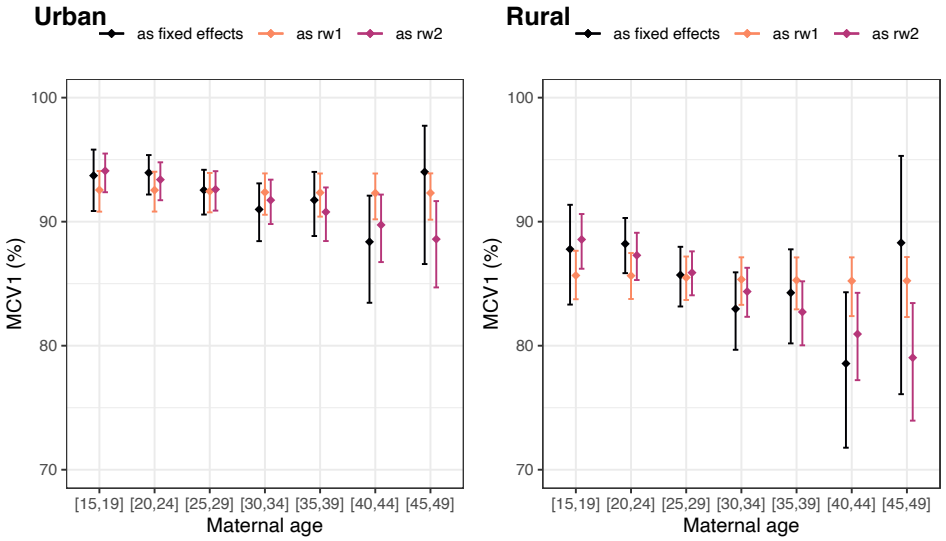


Figure 5.8: Variations in MCV1 estimates among 1-year-olds between maternal age groups within urban or rural areas in Kenya. The figure shows estimates of expit($\alpha_{\text{age}} + \beta_{\text{area}} + \gamma_{iid}$), with α_{age} as iid, rw1 or rw1 random effects described in (5.3). The estimates show a 95% credible interval.

The variation in MCV1 between maternal age groups as fixed effects or random effects of rw1 or rw2, within urban or rural areas in Kenya, are given in Figure 5.8. The models compared are $m2_{\alpha+\beta+\gamma}$ $m2_{r(\alpha)+\beta+\gamma}$. The standard deviation of the estimates from a 95% credible interval shows that rw1 and rw2 models are smaller than for the fixed effects model as expected. If a line were drawn between the median estimates for each model, the smoothness of the lines would have increased in the order of effects of fixed, rw1 and rw2. In general, the MCV1 estimates are higher in urban areas than in rural areas.

The CRPS measures for predicting MCV1 coverage the different maternal age groups in the urban or rural areas for estimating MCV1 in Kenya are given in Table 5.2. The best CRPS on average is with the rw2 model. However, iid and rw2 dominates the best CRPS values for the age groups.

Table 5.4: CRPS values for predictions of MCV1 coverage in missing age groups in urban or rural area in Kenya in expit-scale in per 100. $\overline{\text{CRPS}}$ are the average values. The highlighted cells are the model types with lowest (most favorable) CRPS.

Maternal age	Residency	iid	rw1	rw2	$\overline{\text{CRPS}}$
15-19 years	urban	1.19	1.20	2.38	1.59
	rural	0.72	0.68	2.86	1.42
20-24 years	urban	0.43	0.44	0.65	0.51
	rural	0.64	0.60	0.70	0.65
25-29 years	urban	0.42	0.43	0.41	0.42
	rural	0.37	0.42	0.47	0.42
30-34 years	urban	0.65	0.63	0.52	0.60
	rural	2.64	2.58	1.43	2.22
35-39 years	urban	1.44	0.95	0.67	1.02
	rural	3.26	1.80	0.73	1.93
40-44 years	urban	12.28	12.25	9.59	11.37
	rural	7.23	7.16	3.33	5.91
45-49 years	urban	2.74	2.87	6.07	3.89
	rural	1.97	1.90	3.31	2.39
$\overline{\text{CRPS}}$	urban	2.74	2.68	2.90	
	rural	2.40	2.16	1.83	

Chapter 6

Discussion

Surveys and censuses are the primary data source for estimating health indicators such as neonatal mortality rate and coverage of measles-containing vaccine in low- and middle-income countries. Survey data is complex due to its sampling design of stratifications, weights, and clustering. The observations might have different probabilities of being selected, and each observation is weighted depending on the inclusion probability of the associated observation. Moreover, correlation among clusters should be incorporated in regression analysis because children in the same cluster tend to be similar. A failure to account for the complex survey design is theoretically expected to give estimates and corresponding variances that are biased.

This thesis aimed to evaluate the importance of incorporating clustering by including cluster-specific random effects in model estimation. Probability estimates for neonatal mortality rate and first-dose measles-containing vaccine among 1-year-olds are investigated from data provided by Demographic and Health Surveys collected in Nigeria and Kenya. The associations between maternal age groups, urban or rural areas, and cluster effects are explored as explanatory variables. Logistic mixed models for binary responses are applied to the datasets to include unstructured cluster-specific random effects. The results of the estimates evaluated in the data analysis indicate that the cluster-specific effects should be included in model estimation. Indeed, including the cluster effect generally does not affect negatively the predictive strength of the methods for this particular analysis.

Incorporating the cluster effects as independent and identically distributed random effects results in a shift in the fixed effects of the mixed model. The shift is, for instance, clear for the results from the analysis of the probability estimation of NMR in Nigeria. The standard deviation of the cluster effects for this particular dataset was calculated to be $\hat{\sigma}_c = 0.5$. Although there is a shift between the estimates, the simulation study from population p_2 with the same cluster-specific effects of $\sigma_c = 0.5$, shows that the predicted and the true, empirical proportion rates are almost identical. However, a higher standard deviation such as $\sigma_c = 1$ for the cluster-specific random effects results in a larger impact for the scor-

ing rules under the models. The results from population p_3, i.e, with $\sigma_c = 1$, shows that that including cluster effects is a better choice. Mixed models do not yet eliminate the mean bias error values and mean squared errors for any of the methods achieved from informative sampling design in the simulation study. This indicates that different inclusion probabilities in the sample should not be ignored, and mixed models for regression analysis should be reconsidered when dealing with complex survey design.

The continuous ranked probability scores were calculated for assessing the predictive strength of the methods for maternal age groups as structured and unstructured random effects. Mixed models appear to be a good choice for predicting estimates of missing age groups. The results in the data analysis show that random effects succeed to limit the variation of the estimates between the maternal age groups. In addition, the uncertainties in the estimates of the outermost groups which consists of insufficient information, are reduced in the random effects compared to the fixed effects. For this particular analysis, random effects of random walk of order 2 give the best continuous ranked probability score for the most age groups. It can be discussed that this might be a result of the distribution of the observations and the observed rates for the age groups. Hence, a better performance with higher order structured random effects may be related to its strength to limit variation based on neighboring groups rather than arbitrarily groups such as for independently and identically distributed random effects.

Several aspects of the survey methodology can be discussed. For example, if the inclusion probability of the individuals in a specific area is twice as large as in another area, then the weights of the individuals in the different areas are not the same. Models which account for such interactions with various stratification variables can be developed. Stratification at the urban and rural level, and where associations of maternal age groups are separately studied for urban or rural type of residency. Additional work can include an assumption about interactions between maternal age groups and urban or rural areas as well as model estimation at county level.

Bibliography

Bolstad, W.M., 2007. Introduction to Bayesian Statistics , Second Edition. John Wiley Sons, Inc., Hoboken, New Jersey.

Gelman, A., Carlin, J.B., Stern, H.S., Dunson, D.B., Vehtari, A., Rubin, D.B., 2014. Bayesian Data Analysis, Third Edition. Chapman & Hall, CRC.

Gelman, A., Hill, J., 2006. Data Analysis Using Regression and Multilevel/ Hierarchical Models. Cambridge University Press.

Gneiting, T., Raftery, A.E., 2007. Strictly Proper Scoring Rules, Prediction. Journal of the American Statistical Association 102 , (477) 359–378.

Gómez-Rubio, V., 2020. Bayesian inference with INLA - 3.3 Types of mixed-effects models. https://becarioprecario.bitbucket.io/inla-gitbook/ch-mixed.html#sec:mixed-types. [Accessed: 05.05.21].

Hancioglu, A., Arnold, F., 2013. Measuring Coverage in MNCH: Tracking Progress in Health for Women and Children Using DHS and MICS Household Surveys. https://journals.plos.org/plosmedicine/article?id=10.1371/journal.pmed.1001391#pmed-1001391-t001. [Accessed: 31.03.21].

Hedeker, D., 2005. Generalized Linear Mixed Models. John Wiley Son.

Heering, S.G., West, B.T., Berglund, P.A., 2010. Applied Survey Data Analysis.

ICF, Demographic and Health Surveys Program, Rockville, Maryland, U.S.A.: ICF, 2018. Demographic and Health Surveys Standard Recode Manual for DHS7. URL: https://www.dhsprogram.com/pubs/pdf/DHSG4/Recode7_DHS_10Sep2018_DHSG4.pdf.

Lee, E.S., Forthofer, R.N., 2006. Analyzing complex survey data. URL: http://www.sagepub.com/sites/default/files/upm-binaries/6428_Chapter_6_Lee_(Analyzing)_I_PDF_7.pdf.

Li, Z.R., Martin, B., Mercer, L., 2015. getbirths: Reformat full birth records into person-month format. https://rdrr.io/github/bryandmartin/SUMMER/man/getBirths.html. [Accessed: 03.04.21].

Lohr, S.L., 2010. Sampling: Design and Analysis, Second Edition. Richard Stratton.

Martins, T.G., Simpson, D., Lindgren, F., Rue, H., 2009. Bayesian computing with inla: new features. Computational Statistics Data Analysis 67 , 68–83.

National Population Commission (NPC), National Malaria Elimination Programme (NMEP) of the Federal Ministry of Health, N.T.D.P.I.I., 2019a. Nigeria 2018 Demographic and Health Survey Key Findings. URL: https://www.dhsprogram.com/pubs/pdf/SR264/SR264.pdf.

National Population Commission (NPC), National Malaria Elimination Programme (NMEP) of the Federal Ministry of Health, N.T.D.P.I.I., 2019b. Nigeria Demographic and Health Survey 2018. URL: https://dhsprogram.com/pubs/pdf/FR359/FR359.pdf.

Rue, H., Held, L., 2005. Gaussian Markov Random Fields - Theory and Applications.

Rue, H., Martino, S., Chopin, N., 2009. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. Journal of the Royal Statistical Society J. R. Statist. Soc. B 71, Part 2 319–392.

Rue, H., Riebler, A., Sørbye, S.H., Illian, J.B., Simpson, D.P., , Lindgren, F.K., 2017. Bayesian Computing witn INLA: A Review. Annual Review of Statistics and Its Applications 4, 395–421.

Sharrow, D., Hug, L., Liu, Y., You, D., 2020. Levels and trends in child mortality: Report 2020. https://www.un.org/development/desa/pd/news/levels-and-trends-child-mortality-2020-report. [Accessed: 21.04.21].

Simpson, G., 2021. Using random effects in gams with mgcv. https://fromthebottomoftheheap.net/2021/02/02/random-effects-in-gams/. [Accessed: 04.07.21].

Kenya National Bureau of Statistics, M.o.H., National AIDS Control Council, Kenya Medical Research Institute, N.C.f.P., Development, The DHS Program, I.I., 2015a. Kenya 2015 Demographic and Health Survey - atlas of county-level health indicators. URL: https://www.dhsprogram.com/pubs/pdf/ATR16/ATR16.pdf.

Kenya National Bureau of Statistics, M.o.H., National AIDS Control Council, Kenya Medical Research Institute, N.C.f.P., Development, The DHS Program, I.I., 2015b. Kenya Demographic and Health Survey 2014. URL: https://dhsprogram.com/pubs/pdf/fr308/fr308.pdf.

UN, 2021. Millennium Development Goals and beyond 2015. https://www.un.org/millenniumgoals/. [Accessed: 15.02.21].

WHO, 2019. Measles. https://www.who.int/news-room/fact-sheets/detail/measles. [Accessed: 22.04.21].

Appendix A

Additional Simulation Results

In this section, results from surveys drawn from additional simulated populations referred to in Chapter 4 are displayed. The tables show the average scoring rules of MBE, MSE and coverage for **M1** and **M2** obtained from multiple surveys drawn. **M1** do not include the cluster-specific random effects, while **M2** incorporates for clustering in the mixed model. The scoring rules are defined in Section 2.5, and the models are described in Section 4.4. The results are shown in the appendix as the scoring rules for both models are very similar to each other.

Table A.1: Scoring rules over 400 surveys with $k = 1$ in (4.2) from population p_1 ($\sigma_\gamma = 0$) for estimating NMR associated with maternal age groups and urban or rural area in Nigeria with **M1** as as fixed effects model, while **M2** also includes cluster effects as iid random effects. The coverage scores are calculated from a 95% credible interval. $\text{MSD}_{\mathbf{M2}}(\hat{\sigma}_\gamma) = 0.009$.

Maternal age	Residency	MBE_{M1} (10^{-4})	MBE_{M2} (10^{-4})	MSE_{M1} (10^{-5})	MSE_{M2} (10^{-5})	$\text{Coverage}_{\text{M1}}$	$\text{Coverage}_{\text{M2}}$
15-19 years	urban	-44	-44	10	11	0.930	0.927
	rural	6	6	3	3	0.965	0.965
20-24 years	urban	16	16	3	3	0.930	0.930
	rural	-14	-14	4	4	0.953	0.943
25-29 years	urban	7	7	5	5	0.958	0.958
	rural	35	35	12	12	0.948	0.948
30-34 years	urban	-26	-27	20	20	0.940	0.943
	rural	19	19	7	7	0.960	0.958
35-39 years	urban	-6	-6	2	2	0.968	0.963
	rural	-6	-6	2	2	0.963	0.958
40-44 years	urban	7	7	3	3	0.955	0.950
	rural	7	7	4	4	0.935	0.940
45-49 years	urban	2	2	10	10	0.968	0.958
	rural	4	4	16	16	0.953	0.953

Table A.2: Scoring rules over 400 surveys with $k = 1$ in (4.2) from population p_2 ($\sigma_\gamma = 0.5$) for estimating NMR associated with maternal age groups and urban or rural area in Nigeria with **M1** as as fixed effects model, while **M2** also includes cluster effects as iid random effects. The coverage scores are calculated from a 95% credible interval. $\text{MSD}_{\mathbf{M2}}(\hat{\sigma}_\gamma) = -0.418$.

Maternal age	Residency	MBE_{M1} (10^{-4})	MBE_{M2} (10^{-4})	MSE_{M1} (10^{-5})	MSE_{M2} (10^{-5})	$\text{Coverage}_{\text{M1}}$	$\text{Coverage}_{\text{M2}}$
15-19 years	urban	-5	-5	9	9	0.955	0.955
	rural	-2	-2	5	5	0.925	0.922
20-24 years	urban	-7	-7	3	3	0.932	0.935
	rural	5	4	4	4	0.955	0.960
25-29 years	urban	-1	-1	5	5	0.968	0.963
	rural	13	13	13	13	0.955	0.955
30-34 years	urban	-1	-1	20	20	0.943	0.945
	rural	-4	-4	9	9	0.943	0.943
35-39 years	urban	0	0	3	3	0.960	0.960
	rural	3	3	2	2	0.950	0.945
40-44 years	urban	2	2	3	3	0.963	0.963
	rural	6	6	4	4	0.963	0.963
45-49 years	urban	3	3	12	12	0.955	0.955
	rural	1	1	19	19	0.950	0.953

Table A.3: Scoring rules over 400 surveys with $k = 2$ in (4.2) from population p_3 ($\sigma_\gamma = 1$) for estimating NMR associated with maternal age groups and urban or rural area in Nigeria with **M1** as as fixed effects model, while **M2** also includes cluster effects as iid random effects. The coverage scores are calculated from a 95% credible interval. $\text{MSD}_{\mathbf{M2}}(\hat{\sigma}_\gamma) = -0.069$.

Maternal age	Residency	MBE_{M1} (10^{-4})	MBE_{M2} (10^{-4})	MSE_{M1} (10^{-5})	MSE_{M2} (10^{-5})	$\text{Coverage}_{\text{M1}}$	$\text{Coverage}_{\text{M2}}$
15-19 years	urban	20	15	16	16	0.927	0.968
	rural	6	1	8	8	0.900	0.945
20-24 years	urban	9	5	5	5	0.910	0.958
	rural	8	4	7	7	0.902	0.955
25-29 years	urban	-18	-22	8	9	0.915	0.960
	rural	8	2	17	17	0.935	0.963
30-34 years	urban	-11	-16	26	26	0.927	0.945
	rural	12	1	12	11	0.943	0.963
35-39 years	urban	5	-3	5	5	0.940	0.960
	rural	2	-5	4	4	0.890	0.938
40-44 years	urban	-5	-12	5	5	0.900	0.927
	rural	9	1	6	6	0.927	0.948
45-49 years	urban	8	0	16	16	0.955	0.955
	rural	-7	-18	20	20	0.955	0.945

Table A.4: Scoring rules over 400 surveys with $k = 3$ in (4.2) from population p_3 ($\sigma_\gamma = 1$) for estimating NMR associated with maternal age groups and urban or rural area in Nigeria with **M1** as as fixed effects model, while **M2** also includes cluster effects as iid random effects. The coverage scores are calculated from a 95% credible interval. $\text{MSD}_{\mathbf{M2}}(\hat{\sigma}_\gamma) = -0.073$.

Maternal age	Residency	MBE_{M1} (10^{-4})	MBE_{M2} (10^{-4})	MSE_{M1} (10^{-5})	MSE_{M2} (10^{-5})	$\text{Coverage}_{\text{M1}}$	$\text{Coverage}_{\text{M2}}$
15-19 years	urban	3	-2	15	14	0.935	0.955
	rural	-8	-6	8	8	0.917	0.960
20-24 years	urban	9	4	6	6	0.905	0.958
	rural	7	2	7	7	0.892	0.948
25-29 years	urban	-22	-27	10	10	0.890	0.932
	rural	5	-1	15	14	0.953	0.970
30-34 years	urban	-6	-12	30	29	0.925	0.945
	rural	1	-9	10	10	0.955	0.973
35-39 years	urban	5	-3	5	5	0.932	0.950
	rural	6	-1	4	4	0.890	0.930
40-44 years	urban	-2	-9	5	4	0.940	0.955
	rural	10	2	7	6	0.927	0.960
45-49 years	urban	11	4	15	16	0.963	0.963
	rural	6	-5	24	23	0.950	0.948

NTNU

Norwegian University of
Science and Technology