University of Venda

MULTILEVEL MODELLING OF DETERMINANTS OF CONTRACEPTIVE METHOD CHOICE AMONG WOMEN IN SOUTH AFRICA

By

Nematswerani Phumudzo
(15015373)

submitted in fulfilment of the requirements for the Master of Science Degree in Statistics

in the

Department of Statistics
School of Mathematical and Natural Sciences
University of Venda, Thohoyandou, Limpopo
South Africa

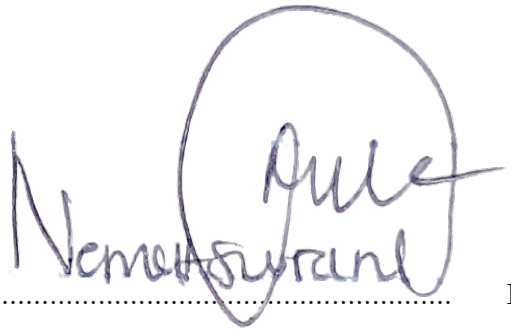
Supervisor: **Dr A. Bere**

Co-Supervisor: **Dr C. Sigauke**

April 2021

Declaration

I, **Nematswerani Phumudzo**, [student number: 15015373], hereby declare that the dissertation titled: "**Multilevel Modelling of Determinants of Contraceptive Method Choice Among Woman in South Africa**" for the Master of Science degree in Statistics at the University of Venda, hereby submitted by me, has not been submitted for any degree at this or any other university, that it is my own work in design and in execution, and that all reference material contained therein has been duly acknowledged.



Signature:..... Date: 28 April 2021

Abstract

Multilevel models take into account various degrees of aggregation in the data. This study aims to bring together multilevel models from both frequentist and Bayesian perspectives in identifying determinants of contraceptive choices. The study uses the data from the 2016 South African Demographic and Health Survey (SADHS). To analyse the dataset, a multinomial logistic regression model has been used, model parameters were estimated in SPSS for frequentist models. The Bayesian analyses with non informative priors were strengthened by the use of the state of the art Hamiltonian Monte Carlo algorithm (HMC), as implemented in the **RStan** package in the R statistical software. The Bayesian final model was selected based on Watanabe–Akaike information criterion (WAIC), which has been shown to outperform conventional information-criterion such as DIC. The results established that an individual woman's choice of contraception is a function of both individual characteristics and community effects. In bivariate analysis, injections showed a continued dominance as a preferred choice in SA. Community level education was the most useful determinant of contraceptive choices. Thus, this study recommends that Empowering woman through education, will have a positive effect on overall contraceptive prevalence.

Key words: *Multilevel modelling, Rstan, Bayesian, HMC, Multinomial regression, WAIC.*

Dedication

I dedicate this dissertation to my Mother
whose interest in this, as in all my ventures
was never less than my own.

Acknowledgements

This thesis is not the result of my effort alone. My Supervisor Dr Bere A provided immense intellectual guidance with seemingly infinite patience. I would also like to thank my co-supervisor, Dr Sigauke C.

Special thanks to DSI-NRF Centre of Excellence in Mathematical and Statistical Sciences (CoE-MaSS), South Africa, for financial support. "Opinions expressed and conclusions arrived at are those of the author and are not necessarily to be attributed to the CoE-MaSS."

Table of contents

Declaration i

Abstract ii

Dedication iv

Acknowledgement v

List of Figures viii

List of Tables x

1 Introduction 1
- 1.1 Background . . . 1
- 1.2 Statement of the Problem . . . 3
- 1.3 Study Aim and Objectives . . . 5
  - 1.3.1 Objectives of the Study . . . 5
- 1.4 Proposed Methodology . . . 5
  - 1.4.1 Data . . . 5
  - 1.4.2 Target Population and Statistical Methods . . . 6
- 1.5 Organization of the Study . . . 6

List of Abbreviations 1

2 Literature review 7
- 2.1 Introduction . . . 7
- 2.2 Approaches to Statistical Analysis . . . 8
  - 2.2.1 Frequentist Paradigm . . . 8
  - 2.2.2 Bayesian Paradigm . . . 9
- 2.3 Key Determinants of Contraceptive Choices . . . 11

2.4 Concluding Remarks 13

3 Study Methodology **14**
3.1 Introduction 14
3.2 Multinomial Logistic Regression 14
3.2.1 The Model 15
3.3 Multinomial Logit Model with Random Effects 16
3.4 Frequentist Estimation 18
3.4.1 Likelihood Function of a Multinomial Random Effects Model 18
3.4.2 Numerical Integration 20
3.4.3 Random Intercept Multinomial Logit Model 21
3.4.4 Random Intercept Variance Component Model 22
3.4.5 Odds Ratios with 95% Confidence Interval (CI) 22
3.5 The Bayesian Hierarchical Modelling 23
3.5.1 The Likelihood Function 23
3.5.2 The Prior Distribution 23
3.5.3 Posterior Distribution 25
3.5.4 Prior and Posterior 26
3.5.5 Markov Chain Monte Carlo (MCMC) 28
3.5.6 Markov Chain Monte Carlo Algorithm 28
3.5.7 Hamiltonial Monte Carlo (HMC) 29
3.5.8 Assessing and Improving Markov Chain Monte Carlo Convergence 31
3.5.9 Software for Bayesian Data Analysis 32
3.6 Criteria for Model Selection 33
3.6.1 Akaike Information Criterion (AIC) 33
3.6.2 Bayesian Information Criterion (BIC) 34
3.6.3 Deviance Information Criterion (DIC) 34
3.6.4 Widely Applicable Information Criterion (WAIC) . . . 35

4 Data Analysis **36**
4.1 Data collection 36
4.1.1 Sources of data 36
4.1.2 Target population 37
4.2 Exploratory data analysis (EDA) 38
4.2.1 Response Variable 38
4.2.2 Explanatory Variables 39
4.3 Bivariate Analyses 41
4.3.1 Chi-squre(χ^2) Test of Association 44

4.4 Classical Multivariate Analysis 45
4.4.1 Classical Multilevel Multinomial Empty Model 45
4.4.2 Results of Random Intercept Multinomial Models . . . 47
4.4.3 Model Comparison and Selection 50
4.5 Results of Bayesian Modelling Framework 56
4.5.1 Bayesian Empty Model 56
4.5.2 Model Assessment and Comparison 59
4.5.3 Result of Final Bayesian Multilevel Multinomial Regression Model 59
4.5.4 Convergence Diagnostics 61
4.5.5 Statistical Computing 67

5 Discussion and Conclusions 69
5.1 Introduction 69
5.2 Discussion 69
5.3 Conclusion 71
5.4 Future research 71

References 72

List of Figures

4.1 Structure of the data considered for analysis. 37
4.2 Distributions and Correlations of the aggregated level 2 covariates. 42
4.3 Bar plot of Contraceptive choices by marital status. 43
4.4 Kernel/ density plots of Markov Chain Monte Carlo for the first 9 parameters. 61
4.5 Trace plots for the first 9 parameters. 62
4.6 Trace plots of Markov Chain Monte Carlo for the first 9 covariates 63
4.7 Geweke plots for the first nine coefficients from the posterior distribution 64

List of Tables

4.1 Frequency and percentage distribution of sampled woman by contraceptive method currently used. 38
4.2 Frequency and percentage distribution of respondents' age groups. 39
4.3 Frequency and percentage distributions of respondents' marital status. 40
4.4 Frequency and percentage(%) distribution of women by number of children they have. 40
4.5 Summary statistics for aggregated cluster level covariates. . . 41
4.6 Cross tabulation for marital status by number of children. . . 43
4.7 Cross tabulation depicting relationships between contraceptive choices and some demographic and maternal characteristics. 44
4.8 Detailed description of Model constructing strategy. 45
4.9 Classical Multilevel multinomial random intercept Model A. . 46
4.10 Fixed effects of Three fitted classical multilevel multinomial logit random intercept Models. 49
4.11 Results of model comparison statistics 50
4.12 Fixed effects of the frequentist multinomial logistics regression. 51
4.13 Odds ratios and 95% confidence interval of the odds of parameters from a classical multilevel-multinomial model, estimating contraceptive method choice among women in South Africa. 53
4.14 Bayesian Multilevel multinomial epty model (Model I). 57
4.15 Summary of the Posterior Distributions for Multilevel-multinomial Models estimating the determinants of contraceptive method choice among women in South Africa. 58
4.16 WAIC of Bayesian Multilevel models 59
4.17 Summary of the Posterior Distribution for a Multilevel-multinomial Model Parameters estimating the determinants of contraceptive method choice among women in South Africa 60

4.18 Odds ratios of the parameters of a final multilevel Bayesian model. 65

Chapter 1

Introduction

1.1 Background

Worldwide, around three quarters of all pregnancies are deemed either undesirable or unplanned, yet account for almost three hundred thousand pregnancies that happen daily (World Health Organization and Partners in Health, 2009). It is evident that South Africa is no exception to these alarming figures, because 330 out of 1000 teenage pregnancies in South Africa end in abortions (Hodes, 2016). Abortion clinics are booked to capacity and the queue can stretch over to four months, because of that reason woman opt for illegal abortions, which is one of leading causes of maternal deaths worldwide. The vast majority of these deaths could be prevented, not only by offering immediate medical care, but also by providing family planning counselling and services, which could counteract future unintended pregnancies and unsafe induced abortions (Luvai, 2017).

The use of contraceptives is an important element of family planning. In

South Africa, proper family planning is a national concern and a necessity in protecting women's health and rights, impacting upon fertility and population growth. Family planning prevents unwanted pregnancies and associated consequences. It in this manner promotes economic development of a country (Ferede, 2013). In addition to functioning as a barrier to conception, the use of family planning techniques such as condoms, a modest kind of contraception, provides protection against STIs including HIV. (Magadi and Curtis, 2003).

Modelling determinants of contraceptive method choice constitutes a vital part of the health policy of a country, especially for a developing country like South Africa, which has been particularly vulnerable to unplanned pregnancies and sexually transmitted infections(STI's) including HIV (Cowan and Pettifor, 2009). Substantial proof is found in existing writing that widening the choice of contraceptive methods results in an overall increase of contraceptive prevalence rate (Ross and Stover (2013); Kulczycki (2004); Do and Kurimoto (2012)).

Demographic characteristics, social and religious convictions, economic status and education levels of the female populace can also influence the choice of a contraceptive method. This population-based study is an attempt to determine the determinants of choice of contraception in South Africa based on the 2016 South African Demographic and Health Surveys (SADHS) data.

1.2 Statement of the Problem

Most existing recent South African contraception analysis studies have mainly focussed on using binary logistic regression to identify those factors that determine use/non-use of contraceptives by individuals in the population (for example, see Seutlwadi et al. (2012); Stephenson et al. (2008b); Kaida et al. (2010); among others). One of the major inspirations for the current study which focusses on determinants of contraceptive method choice is the fact that, the challenge for policymakers does not end in urging more couples to use contraceptives, but it rather stretches out to the choice of an appropriate contraceptive method that suits the customer's needs.

The data to be used in this dissertation is secondary level data from the 2016 South African Demographic and Health Survey (SADHS). The DHS datasets have a multilevel nature in the sense that woman are nested within clusters, thus leading to observations in the same cluster exhibiting some similarities. It would be reasonable to believe that woman from the same cluster will have contraceptive choices that are more highly correlated with one another than they are with contraceptive choices of individuals chosen at random from the population. We are interested in measuring the degree of association among women in the cluster that is still present after controlling for the observed covariates. The association is a result of unobserved factors. This within-cluster correlation would be due, for example, to similar cultural beliefs, the same health care facilities and other factors. To account for this heterogeneity the current study will resort to multilevel modelling.

Few studies have applied multilevel modelling for contraceptive method

choice in a South African context. There is work that has been done by Stephenson et al. (2008a), where they involved the use of multilevel-multinomial regression models to model determinants of contraceptive method choice in Eastern Cape. The fact that its a small-scale study focused in one province limits the applicability of the results on a larger scale, particularly considering a multi-regional and multi-cultural setting of South Africa. The authors used multilevel modelling using a frequentist method of estimation, and on an outdated data from 1998 SADHS. The current study will apply multilevel modelling on a more recent dataset from 2016 SADHS and the study will cover the whole country.

The majority of publications in the literature focus on the frequentist technique of estimation. One significant disadvantage of the frequentist method is that it does not account for uncertainty in parameter estimation, Bayesian analysis is one technique to overcome this. Bayarri and Berger (2004) indicated that there are several areas of frequentist technique that may be substituted with Bayesian methodology that offer prevailing results. In this study we will use Bayesian multilevel multinomial logistic regression to models the determinants of contraceptive method choice in South Africa.

Bayesian multilevel multinomial logistic regression models are not widely used. Very few studies have attempted to compare frequentist and Bayesian approaches in estimating parameters of a multilevel multinomial logistic regression model, and those that have, have yielded mixed findings. Therefore more research and testing using the same methodology is required, to gain a better understanding of an estimation technique which yields better results as far as the multilevel multinomial logistic regression model is concerned.

1.3 Study Aim and Objectives

This study aims to bring together frequentist and Bayesian estimation together with multilevel analysis in determining the factors which explain the determinants of contraceptive choice among woman in South Africa.

1.3.1 Objectives of the Study

The specific objectives of this research are to:

- build a multilevel multinomial logistic regression model in order to identify the determinants of contraceptive choice using classical/frequentist approach,
- build a multilevel multinomial logistic regression model using Bayesian statistical approach, and compare the results of the two approaches
- make provision of relevant recommendations for policy makers and suggest directions for future research

1.4 Proposed Methodology

1.4.1 Data

In order to explore the whole country, this study will use the secondary level data from the 2016 South African Demographic and Health Survey (SADHS). The data was collected from May to November 2016, and that is the latest South African national dataset with information on contraceptive use.

1.4.2 Target Population and Statistical Methods

Since this research study focuses on the determinants of contraceptive method choice among women, the sample will be limited to Women who were at risk of conception at the time of the survey. For analysis we fit multilevel multinomial logistic regression models to cater for possible clustering of responses at regional level. This study will use both frequentist and Bayesian approaches to estimate parameters.

1.5 Organization of the Study

The rest of the dissertation is organised as follows. In Chapter 2 a literature review provides a background of statistical modelling framework linking the multilevel modelling and contraceptive choices. It is in this chapter where we review several recent studies that attempt to link the Bayesian analysis and multinomial logistic regression models. Chapter 3 focuses on the theory behind multilevel or hierarchical models built from frequentist and Bayesian perspectives. Chapter 4 gives a detailed explanation of the data, some basic exploratory analyses to explore elementary relationships between the variables, lastly the theory explained in Chapter 3 is used to fit the models. Chapter 5 provides discussion, conclusion, recommendations and directions for future research.

Chapter 2

Literature review

2.1 Introduction

Across many fields, the determinants of contraceptive choice is not a new scientific problem. The problem has been studied by many researchers from medical science, epidemiology, sociology and other disciplines. In this chapter the main focus will be on the review of the related contraceptive choice literatures. Special attention is paid to methodological approaches used for analysis.

The chapter is organized as follows: Section 2.1 deals mainly with the review of relevant literature on frequentist and Bayesian inference. Section 2.2 reviews articles on determinants of contraceptive choices. It starts by presenting the context of familly planning in other parts of the world and narrows down to the case of South Africa .

2.2 Approaches to Statistical Analysis

In light of the objectives, the dissertation focuses on the modelling of determinants of contraceptive method choice using a frequentist and Bayesian analysis. First and foremost, it is noteworthy that there is an important distinction in philosophy between Bayesian and frequentist estimation. (Bayarri and Berger, 2004). Basically the essential contrast between the two paradigms lies on how they define what probability expresses (Samaniego, 2010).

2.2.1 Frequentist Paradigm

The frequentist approach, as the name suggests, is the paradigm that interpret probabilities as a long-run frequency of a "repeatable" event. As per Bayarri and Berger (2004), a frequentist paradigm can be loosely refereed to as "classical". It is in this paradigm where parameters of interest are treated as fixed. The paradigm is divided into two inferential techniques namely: model−based and design−based. As argued by Särndal et al. (1978), the difference between the two techniques lies in the sources of random variation that is capable of giving the stochastic structure in the data. Särndal et al. (1978) and Cochran et al. (1977) are two classic references with an excellent comparison of model−based and design−based inferences.

Over the past three decades, a frequentist modelling framework to multinomial logistic regression models has been used extensively to model discrete choices (see, Park and Kerr (1990); McFadden and Train (2000); Starkweather and Moske (2011); among others). These models can range from single level to multilevel depending on sampling methods used in data collection. Single level classical multinomial logit models are now commonplace

tools for modelling the determinants of contraceptive choices and work very well when the data is not nested. Compared with multilevel models, single level models are usually straightforward to implement and produce relatively robust solutions.

Despite the hierarchical structure of DHS data sets, Indongo (2007) and Aragaw (2015) also used single level regression models, to model the determinants of contraceptive choices among woman. Unfortunately without properly accounting for violation of the independence assumptions the results of such studies can be biased. Fortunately multilevel modelling can be used to to adequately adjust for hierarchies that exists within the DHS data sets (Finch et al., 2016).

Although, hierarchical Frequentist models have been applied to contraceptive choice, the literature is "very small and scattered". In the case of South Africa, only one small contraceptive choice study was conducted using the methodology (Stephenson et al., 2008a). Another example of the use of a hierarchical analysis for polychotomous data is Magadi and Curtis (2003). In this article the authors used a two-level hierarchical model for contraceptive method choice which was estimated using a frequentist approach. We consider frequetist and Bayesian estimation of the same model.

2.2.2 Bayesian Paradigm

In the Bayesian approach which is an alternative statistical paradigm to the Frequentist inference, probabilities are perceived as subjective and are interpreted conditional to the availability of data (McElreath, 2020). The Bayesian approach took a sizable step forward in 1763, when Richard Price edited and published Bayes work posthumously. This paradigm uses prob-

ability to quantify uncertainty, or degree of belief, hence probability distributions called priors are used to represent what is believed before data are observed. As highlighted in Congdon (2010) Bayesian statistics combines a prior knowledge with the likelihood of the data to generate a posterior distribution.

Choosing of prior in the Bayesian paradigm is the pivotal point, with a choice of deciding between informative and non-informative (Lesaffre and Lawson, 2012). using hirachical Bayesian multinomial models, For computation ease many authors have adopted the Dirichlet priors, but it was noted by that the conjugate prior does not always provide (**?**) . In expense of simple conjugate the current study will adopt a hierarchical approach of specifying a prior distribution. Referring to Ebenezer and Lougue (2019), Multivariate normal priors will be used.

Since the exact computation of the posterior distribution is practically infeasible, a Markov chain Monte Carlo (MCMC) technique is required to sample from the target distribution (Lesaffre and Lawson, 2012). Metropolis Hastings and the Gibbs samplers are the most frequently used MCMC algorithms Wagenmakers et al. (2008). These can be computed from **MCMCpack** by Martin et al. (2020). Departing from the commonly used **MCMCpack**, this dissertation will utilize **Rstan** package by Carpenter et al. (2017), which uses Hamiltonian Monte Carlo(HMC) algorithm to obtain draws from the posterior distribution. While traditional MCMC chains conceivably require millions of iterations to arrive at stationarity, HMC often need a few thousand iterations to attain a reasonably mixed posterior distribution Monnahan et al. (2017).

Several authors have adopted hierarchical Bayesian multinomial models

to analyse categorical data, however relative little contraceptive related research has adopted the methodology. More recently, Ebenezer and Lougue (2019) employed a Bayesian generalized linear mixed model to analyse multinomial data for the problem of breast cancer. Their model is very comparable to the one considered here; we describe it further in the next chapter.

2.3 Key Determinants of Contraceptive Choices

According to Hermalin (1983), a woman's decision to use any contraceptive method to avoid pregnancy is dictated by the relative costs and benefits of a pregnancy. Tanfer et al. (1992) assumed that individual woman define these relative costs and benefits based on their socio-economic status, demographic and maternal characteristics. This study utilizes the same theoretical framework adopted by (Tanfer et al., 1992).

Social and economic determinant factors associated with contraceptive choices include education attainment. The more educated a woman gets, the more likely she is to be aware of her risk of pregnancy and knowledgeable about contraceptive options (Larsson and Stanfors, 2014). In USA, better educated woman constantly reported a low use of Sterilization which is a long term method, and much higher likelihood of using a pill(a short term method) Mosher and Jones (2010). Similar to the results of Mosher and Jones (2010), a Kenyan study by Kungu et al. (2020), demonstrated that young woman with at least secondary education and coming from families with high wealth status are more likely to use short term contraceptive methods, and the pattern is reversed for poor and uneducated woman. In the case of South Africa, contrary to Kungu et al. (2020), woman coming

from less educated communities are negatively associated with long term methods.

Evidence of associations between contraceptive choice and demographic or residential characteristics was noted in (Chigbu et al., 2008). Radovich et al. (2018) analysed the determinants of contraceptive choices among women across 33 sub-saharan countries. Women aged $15-19$ and 20–24 were significantly less likely to use IUD/implant (long−term method). A South African study reveals that, compared with their younger $(15-24)$ counterparts, older woman (25+) had a considerably more frequent use of sterilization (Stephenson et al., 2008a). Regarding parity, women with five or more children were positively associated with the use of a more permanent method (sterilization) instead of injection, as compared to women without children (Scott and Glasier, 2006). The marital status of a women is important in contraceptive research because it reflects how sexually active that person is. This is especially important considering the controversy about sexual contact. The study highlighted that marital status, ethnicity and residence type are also closely related to the choice of method. Since the provision of contraceptives is free and highly accessible in both rural and urban areas of South Africa this study will not use wealth and type of residence as factors associated with contraceptive choices.

2.4 Concluding Remarks

Departing from previous studies, this study applies the Bayesian and classical multinomial logistic regression methodology to contraception data. The review reveals several gaps in the literature, including little or no research on the topics of contraceptive choices in South Africa. It is evident from the reviewed studies that very few population based studies on contraceptive choices have been carried out in South Africa and some of them were based on an outdated data, and some where conducted with inadequate methodologies. Also, no study has been done to compare frequentist and Bayesian approach in estimating parameters of a Multi-level Multinomial logistic regression model. Hence this study seeks to address the identified research gaps by combining multilevel modelling with both frequentist and Bayesian methods of parameter estimation.

Chapter 3

Study Methodology

3.1 Introduction

This chapter provides an in-depth review of methodological approaches for clustered multinomial response models and multilevel models. Both frequentist and Bayesian methods will be addressed. Special attention is paid to methodological aspects of the random effects multinomial logit model from a frequentist and a Bayesian viewpoint. After introducing methods a brief exploration of techniques used in obtaining parameters and precision estimates will follow. Software to be used in implementing these approaches will also be highlighted.

3.2 Multinomial Logistic Regression

The multinomial logistic model is a generalization of the binary logistic model (McFadden and Train, 2000). The response variable takes three or more categories. We now consider models for the probabilities π_{ij}.

3.2.1 The Model

Assume Y_i is a nominal categorical outcome variable for the i^{th} observation, which can take M possible integer values denoted by $(1, 2, ..., M)$ with corresponding probability π_{im} when

$$\pi_{im} = \Pr(Y_i = m) \quad \text{for} \quad m = 1, \ldots, M. \tag{3.2.1}$$

Let $\mathbf{x_i}$ be a vector of p explanatory variables for the i^{th} observation, $\mathbf{x_i} = (x_1, x_2, ..., x_p)^T$. Under MNL (multinomial logistic regression) structure with $Y_i = M$ as the baseline (reference category), the model is:

$$\begin{aligned} \log\left(\frac{\pi_{im}}{\pi_{iM}}\right) &= \log\left(\frac{\pi_{im}}{1 - \sum_{h=1}^{M-1} \pi_{im}}\right) = \eta_{im} \\ &= \mathbf{x_i^T}\beta_m = \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p, \end{aligned} \tag{3.2.2}$$

where $i = 1, 2, ..., N$ and $m = 1, 2, ..., M-1$. For any $m \neq M$ and the coefficient vector $\beta_m = (\beta_{m1}, \beta_{m2}, \ldots, \beta_{(M-1)p})^T$ where $m = 1, 2, \ldots, M-1$ (Starkweather and Moske, 2011). To calculate the probabilities, we have:

$$\Pr(Y_i = m) = \pi_{im} = \frac{\exp\left(\mathbf{x_i^T}\beta_m\right)}{1 + \sum_{h=1}^{M-1} \exp\left(\mathbf{x_i^T}\beta_h\right)}, \tag{3.2.3}$$

for the non-baseline categories $h = 1, 2, \ldots, M-1$, while for the baseline category probability is

$$\Pr(Y_i = M) = \pi_{iM} = \frac{1}{1 + \sum_{h=1}^{M-1} \exp\left(\mathbf{x_i^T}\beta_\mathbf{h}\right)}. \tag{3.2.4}$$

The multinomial logistic regression model is a GLM (Generalizedlinear Model) with three main components Starkweather and Moske (2011).

• **Random component**

The the response f follows a Multinomial distribution:

$$y_i \sim \text{Multinomial}(y_i|\pi_{im})$$

while, $\pi_{im} = \Pr(Y_i = m)$ for $m = 1, \ldots, M$

• **Systematic Component**

This specifies explanatory variables $(X_1, X_2, ..., X_p)$ and their linear combination.

$$\eta = \mathbf{x_i^T}\beta_\mathbf{j}$$

• **Link function**

Generalized Logit function.

$$g(.) = \eta_i = \text{logit}(\pi_i) = \log\left(\frac{\pi_i}{1-\pi_i}\right) \tag{3.2.5}$$

3.3 Multinomial Logit Model with Random Effects

In this study, our nominal response is contraceptive choices taking values "Oral", "barrier", "sterilisation","other methods" and "injection", which we index $k = 1, 2, 3, 4$ and 5. We use the logit model to pair each of the first four response categories with "injection" (baseline category) and fit these models simultaneously. The general form of the baseline-category logit model with random effects is given by

$$\eta_{ijk} = \text{logit}(\pi_{ijk}) = \frac{P(Y_{ij} = k|\boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \boldsymbol{u}_i)}{P(Y_{ij} = M|\boldsymbol{x}_{ij}, \boldsymbol{z}_{ij}, \boldsymbol{u}_i)} \tag{3.3.1}$$

$$= \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_k + \boldsymbol{z}_{ij}^T\boldsymbol{u}_i \tag{3.3.2}$$

Where

$$\pi_{ijk} = h(\eta_{ijk}) = \frac{\exp(\eta_{ijk})}{1 + \sum_{k=1}^{M-1} \exp(\eta_{ijk})} = \frac{\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_k + \boldsymbol{z}_{ij}^T\boldsymbol{u}_i\right)}{1 + \sum_{k=1}^{M-1} \exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_k + \boldsymbol{z}_{ij}^T\boldsymbol{u}_i\right)}, \tag{3.3.3}$$

where $\boldsymbol{\beta_k} = (\beta_{0k}^T, \ldots, \beta_{pk})^T$ is a vector of fixed effects and $\boldsymbol{u}_i = (u_{i0}, \ldots, u_{is})^T$ is a $s+1$ dimensional cluster-specific random effect.

The models (3.3.2) and (3.3.3) can be presented in matrix form as multivariate generalized linear mixed models for categorical responses. We use the notation Y_{ij} for the j^{th} observation of cluster i. Y_{ij} takes values from contraceptive choices $\{1, \ldots, M\}$, or $y_{ij} = (y_{ij1}, \ldots, y_{ij(M-1)})^T$. The corresponding model for observation y_{ij} has the form:

$$\boldsymbol{\eta}_{ij} = \text{logit}(\boldsymbol{\pi}_{ij}) = \boldsymbol{X}_{ij}\boldsymbol{\beta}_k + \boldsymbol{z}_{ij}\boldsymbol{u}_i \tag{3.3.4}$$

$$\boldsymbol{\pi}_{ij} = h(\boldsymbol{\eta}_{ij}) \tag{3.3.5}$$

where vector $\boldsymbol{\beta}$ is the vector for fixed parameters, and u_i is the vector for the random effects; $M-1 \times q$-dimensional X_{ij} and the $(M-1) \times v$-dimensional Z_{ij} are the model matrices for the fixed and random effects respectively, all typically have the forms:

$$\boldsymbol{X}_{ij} = \begin{pmatrix} \boldsymbol{x}_{ij}^T & 0^T & \dots & 0^T \\ 0^T & \boldsymbol{x}_{ij}^T & \dots & 0^T \\ \vdots & \vdots & \ddots & \vdots \\ 0^T & 0^T & \dots & \boldsymbol{x}_{ij}^T \end{pmatrix} \quad \boldsymbol{z}_{ij} = \begin{pmatrix} \boldsymbol{x}_{ij}^T \\ \vdots \\ \boldsymbol{z}_{ij}^T \end{pmatrix} \quad \boldsymbol{\beta} = \begin{pmatrix} \beta_1 \\ \vdots \\ \beta_{M-1} \end{pmatrix}$$

u_i usually follows a multivariate normal distribution with mean 0 and variance-covariance matrix $\sum$.

3.4 Frequentist Estimation

In applications, assuming that the correct model is being used, questions arise as how to estimate the parameters β_{ij}. The solution to the question is related to the notion of the likelihood. In this section we briefly introduce a review of general concept behind the estimation approaches to be adopted in this dissertation. To obtain the parameter estimates (and their precision) in a frequentist way, we have to maximize the total marginal likelihood which results in the Maximum Likelihood Estimate (MLE).

3.4.1 Likelihood Function of a Multinomial Random Effects Model

Let $\boldsymbol{y}_{ij}^T|\boldsymbol{u}_i = (y_{ij1}, \dots, y_{ij(M-1)}) \sim MN(n_{ij}, \boldsymbol{\pi}),\ i = 1, \dots, n_i$, denote the multinomial distribution with M categories. The multinomial distribution has the form of a multivariate exponential family. The conditional density of $\boldsymbol{y}_{ij}$, given the explanatory variables, $\boldsymbol{X}_{ij}$ and $\boldsymbol{Z}_{ij}$ in equation (3.3.5), and the v- dimensional random effect $\boldsymbol{u}_i$, $\quad f(\boldsymbol{y}_{ij}|\boldsymbol{u}_i)$ belong to the multivariate exponential family with

$$\boldsymbol{\mu}_{ij} = E(\boldsymbol{y}_{ij}|\boldsymbol{u}_i)] = h(\boldsymbol{\eta}_{ij}), \quad \boldsymbol{\eta}_{ij} = \boldsymbol{X}_{ij}\boldsymbol{\beta} + \boldsymbol{Z}_{\boldsymbol{ij}}\boldsymbol{u}_{\boldsymbol{i}} \tag{3.4.1}$$

The general multinomial model is defined by equation in terms of the response vector $\boldsymbol{y}_{ij}$ or scaled multinomials/proportions $\boldsymbol{p}_{ij} = \frac{1}{n_{ij}\boldsymbol{y}_{ij}}$. For example, the baseline-category logit random effects model has

$$\pi_{ijk} = h_j(\boldsymbol{\eta}_{\boldsymbol{ij}}) = \frac{\exp(\eta_{ijk})}{1 + \sum_{k=1}^{M-1}\exp(\eta_{ijk})}, k = 1, \ldots, M-1 \quad \text{and} \quad \pi_{ijM} = \frac{1}{1 + \sum_{k=1}^{M-1}\exp(\eta_{ijk})}$$

Then the conditional probability function is

$$\begin{aligned} f(y_{ij}|u_i) &= \frac{n_{ij}!}{\prod_{k=1}^{M} y_{ijk}!}\prod_{k=1}^{M}\pi_{ijk}^{y_{ijk}} \\ &= \frac{n_{ij}!}{y_{ij1}!\ldots y_{ij(M-1)}!\left(n_{ti} - \sum_{j=1}^{M-1} y_{ijk}\right)!}\pi_{ij1}^{y_{ij1}}\ldots\pi_{ij(M-1)}^{y_{ij(M-1)}}\left(1 - \sum_{j=1}^{M-1}\pi_{ijk}\right)^{\left(n_{ti} - \sum_{j=1}^{M-1} y_{ijk}\right)} \\ &= \exp\left\{\boldsymbol{y}_{ij}^T + n_{ij}\log(\pi_{ijM}) + \log(M_{ij})\right\} \\ &= \exp\left\{n_{ij}\boldsymbol{p}_{ij}^T\boldsymbol{\phi}_{ij} + n_{ij}\log(\pi_{ijM}) + \log(M_{ij})\right\}, \end{aligned}$$

where the canonical parameter $\boldsymbol{\phi}_{ij} = (\phi_{ij1}\ldots\phi_{ijg})^T$, $\phi_k = \log\left(\frac{\pi_{ijk}}{i}\pi_{ijM}\right)$, $\pi_{ijk} = \left(1 - \sum_{k=1}^{M-1}\pi_{ijk}\right)$, and the dispersion parameter is $\frac{1}{n_{ij}}, M_{ij} = \frac{n_{ij}!}{\prod_{j=1}^{M} y_{ijk}!}$. Averaging out the continuous random effect through integretion, the marginal distribution has mean (using Adams law).

$$E(y_{ij}) = E[E(\boldsymbol{y}_{ij}|\boldsymbol{u}_i)] = E[\boldsymbol{h}(\boldsymbol{\eta}_{\boldsymbol{ij}}],$$

and variance-covariance matrix (law of total varience).

$$V(\boldsymbol{y}_{ij}) = E[V(\boldsymbol{y}_{ij}|\boldsymbol{u}_i)] + V[E(\boldsymbol{y}_{ij}|\boldsymbol{u}_i)].$$

The distribution of the total response for i^{th} cluster $n_i \times 1 -$ vector $y_i = (\boldsymbol{y}_{i1}^T, \ldots, \boldsymbol{y}_{in_i1}^T, \ldots,)^T = (y_{i11}, \ldots, y_{i1g}, \ldots, y_{in_i1}, \ldots, y_{in_ig})^T$ is obtained by assuming the conditional independence of $\boldsymbol{y}_{i1}, \ldots, \boldsymbol{y}_{in_i}$ given $\boldsymbol{u}_i$. The marginal probability function of $\boldsymbol{y}_i$ is

$$\begin{aligned} f(\boldsymbol{y}_i) = \int f(\boldsymbol{y}_i, \boldsymbol{u}_i) d\boldsymbol{u_i} &= \int f(\boldsymbol{y}_i|\boldsymbol{u}_i)\phi(\boldsymbol{u}_i, \sum) d\boldsymbol{u_i} \\ &= \int \left[\prod_{t=1}^{n_i} f(\boldsymbol{y}_{ij}|\boldsymbol{u}_i)\right] \phi(\boldsymbol{u}_i, \sum) d\boldsymbol{u_i}, \end{aligned}$$

Where $\phi(\boldsymbol{u}_i, \sum)$ denotes the density of the random effects, which are assumed to have no relations with the fixed effects. The generalised linear mixed model (GLMM) likelihood function is the marginal mass function of the observed multinomial data, $\boldsymbol{y_i}$, we consider it as a function of the parameters of interest, with the form:

$$\begin{aligned} L(\boldsymbol{\beta}, \sum) &= \prod_{i=1}^{m} f(\boldsymbol{y}_i) = \prod_{i=1}^{m} \int \left[\prod_{t=1}^{n_i} f(\boldsymbol{y}_{ij}|\boldsymbol{u}_i)\right] \phi(\boldsymbol{u}_i, \sum) d\boldsymbol{u_i} \\ &= \prod_{i=1}^{m} \int \left[\prod_{t=1}^{n_i} \exp\left\{\boldsymbol{y}_{ij}^T + n_{ij}\log(\pi_{ijM}) + \log(M_ij)\right\}\right] \phi(\boldsymbol{u}_i, \sum) d\boldsymbol{u_i}, \end{aligned}$$

where $\boldsymbol{\beta}$ and covariance matrix $\sum$ are the parameters of interest, that needs to be estimated, where the covariance matrix $\sum$ of the random effects $\boldsymbol{u-i}$ depends on an unknown parameter vector $\boldsymbol{\sigma}$, which represents the variance components.

3.4.2 Numerical Integration

Integration over the distribution of random effects is needed to solve the above likelihood expression. In the literature, some approximations (nu-

merical integration techniques) for estimating the integral over the random-effects distribution have been suggested. Here we use a deterministic method (Gauss-Hermite quadrature) to approximate the integration.

3.4.3 Random Intercept Multinomial Logit Model

In order to address cluster heterogeneity and intra-cluster correlation we use the random intercept multinomial model. We will look at the simple model containing only one random intercept. Here we consider a random intercept model, where the linear predictor, η_{ijk}, of a woman coming from the i^{th} cluster and choosing the k^{th} (contraceptive method choice) is given by:

$$\begin{aligned}\eta_{ijk} = g(\pi_{ijk}) = \text{logit}(\pi_{ijk}) = \frac{P(Y_{ij} = k|\boldsymbol{x}_{ij}, u_i)}{P(Y_{ij} = M|\boldsymbol{x}_{ij}, u_i)} \\ = \boldsymbol{x}_{ij}^T\boldsymbol{\beta}_k + u_i\end{aligned}$$

$$\begin{aligned}\pi_{ijk} = h(\eta_{ijk}) = \frac{\exp(\eta_{ijk})}{1 + \sum_{k=1}^{M-1}\exp(\eta_{ijk})} \\ = \frac{\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_k + u_i\right)}{1 + \sum_{k=1}^{M-1}\exp\left(\boldsymbol{x}_{ij}^T\boldsymbol{\beta}_k + u_i\right)},\end{aligned}$$

where u_i is the cluster-specific intercept for all categories. The fixed effects determine the effects of the covariates but the response strength may vary across different clusters. This model is obtained by specifying $z_{it} = 1$ from the general random effects multinomial logit model **??**. When our data is sparse, i.e., when the number of observations per cluster is small, we may use conditional likelihood (Demidenko, 2013).

3.4.4 Random Intercept Variance Component Model

We first employ a simple model with no explanatory variables i.e. an intercept-only model (usually called the empty model) that predicts the probability of contraceptive Choices. The functional form of the model is

$$\log\left(\frac{\pi_{ijk}}{\pi_{ijM}}\right) = \eta_{ijk} = \beta_{0j} + \xi_{ijk}. \tag{3.4.2}$$

Allowing the intercept to differ across clusters, leads to the random intercept that we express as

$$\beta_{0j} = \gamma_{00} + U_{0j}, \tag{3.4.3}$$

where γ_{00} is an average or overall intercept value that holds across groups (clusters), whereas U_{0j} refers to a group-specific deviation from the intercept. Hence the unified random intercept model can be expressed as

$$\log\left(\frac{\pi_{ijk}}{\pi_{ijM}}\right) = \gamma_{00} + U_{0j} + \xi_{ijk}. \tag{3.4.4}$$

3.4.5 Odds Ratios with 95% Confidence Interval (CI)

Since point estimates can be misleading, odds ratios should be interpreted in terms of their 95 percent confidence interval. A wide confidence interval means that the OR has a poor level of precision, while a small confidence interval indicates that the OR has a higher level of precision. The population log odds ratio has a 95% confidence interval of approximately:

$$CI = \ln(\text{OR}) \pm 1.96 \times \{\text{St.Er}\ln(\text{OR})\}, \tag{3.4.5}$$

where ln(OR) is the log odds and St.Er ln(OR) is the standard error of estimate of the log odds ratio. We can deduce the 95% confidence interval of OR as:

$$e^{\text{CI}} = \exp\{\log(\text{OR}) \pm 1.96 \times \{SE\log(\text{OR})\}\}$$

3.5 The Bayesian Hierarchical Modelling

This section introduces the Bayesian methods and the computational methodologies on which parameter estimates in this study are obtained. The development of Bayesian inference has the data likelihood as a fundamental concept (Lesaffre and Lawson, 2012).

3.5.1 The Likelihood Function

Let $y_i,\ i = 1, ..., n$ be a random variable with probability density function $\pi(y_i|\phi)$, where $\phi = (\phi_1, ..., \phi_p)$ is a vector of relative risk parameters. The likelihood function of y_i is defined as

$$\pi(\mathbf{y}|\phi) = \prod_{i=1}^{n} \pi(y_i|\phi) \tag{3.5.1}$$

Equation 3.5.1 is based on the assumption that the sample values of $\mathbf{y} = (y_1, ..., y_p)'$ given the parameters ϕ are independent (Lesaffre and Lawson, 2012).

3.5.2 The Prior Distribution

As previously stated, Bayesian methods are based on prior belief about the parameters of interest. However, this belief about a parameter is captured in a density function referred to as a prior distribution. In the case of poor

or small sample sizes, the analysis will be dominated by the the prior distribution. According to Lesaffre and Lawson (2012), Prior distributions give extra "data" for a problem and may thus be utilized to improve parameter estimation.

The Propriety

The condition of improper prior distribution is when the integration of a random variable ϕ's prior distribution over ω is infinity. Mathematically we write:

$$\int_{\omega} \pi(\phi) d\phi = \infty \tag{3.5.2}$$

If the normalizing constant of a prior distribution is finite, it is said to be proper Bayarri and Berger (2004). It should be remembered that, while impropriety is a restriction to any prior distribution, an inappropriate prior does not always imply impropriety in the posterior.

Conjugate Prior

If a prior $\pi(\phi|y)$ and the posterior $\pi(\phi)$ are both coming from the same family of distributions, then such prior is refereed to as a conjugate of the likelihood.

In the case of MNL model the conjugate prior is the Dirichlet distribution. if $y = (y_1, y_2, \ldots, y_q)$, and y has a multinomial distribution with parameters n and $\phi = (\phi_1, \phi_2, \ldots, \phi_q)$ denoted as Multin$(n;\phi)$ then the likelihood:

$$\pi(\phi|y) \propto \phi_1^{y_1} \phi_2^{y_2} \ldots \phi_q^{y_q},$$

where $\phi_j \geq 0$ for all $j = 1, \ldots, q$ and $\sum_{j=1}^{q} \phi_j = 1$.

The conjugate prior is the Dirichlet distribution $\mathrm{D}(\alpha_1, \ldots, \alpha_q)$:

$$\pi(\phi) \propto \phi_1^{\alpha_1 - 1} \phi_2^{\alpha_2 - 1} \ldots \phi_q^{\alpha_q - 1},$$

where $\phi_j, \alpha_j \geq 0$ for all $j = 1, \ldots, q$ and $\sum_{j=1}^{q} \phi_j = 1$. As a result, the posterior distribution that results is another dirichlet distribution denoted as $\mathrm{D}(\alpha_1 + y_1, \ldots, \alpha_q + y_q)$:

$$\pi(\phi) \propto \phi_1^{\alpha_1 + y_1 - 1} \phi_2^{\alpha_2 + y_2 - 1} \ldots \phi_q^{\alpha_q + y_q - 1}.$$

additional "data" for a problem, hence can be used to enhance estimation of parameters

Non-Informative Priors

A non-informative prior, also known as flat, or reference prior, was defined by Arango et al. (2002) as one that gives very little or no detail at all about the experiment or has a minor influence on the results relative to the data.

3.5.3 Posterior Distribution

The posterior distribution is determined by multiplying the prior distribution and the likelihood, which can be expressed mathematically as:

$$\pi(\phi|y) = \begin{cases} \dfrac{f(y|\phi)g(\phi)}{\sum_{\phi} f(y|\phi)g(\phi)}, & \text{for a discrete parameter} \\ \dfrac{f(y|\phi)g(\phi)}{\int f(y|\phi)g(\phi)d(\phi)}, & \text{for a continuous parameter} \end{cases}$$

Since the parameters of interest in this analysis are continuous, the emphasis is on Bayesian inference for continuous results. Hence using Bayes theorem we can rewrite the above equation with only the continuous parameters as:

$$\pi(\phi|y) = \frac{\pi(y|\phi)g(\phi)}{\pi(\phi)}, \tag{3.5.3}$$

with,

$$\pi(\phi) = \int \pi(y|\phi)g(\phi) \tag{3.5.4}$$

which is known as the normalising constant. Ignoring the constant we can write the marginal probability in equation (3.5.3) as:

$$\pi(\phi|y) \propto \pi(y|\phi) \times g(\phi) \tag{3.5.5}$$

more colloquially,

$$\text{posterior} \propto \text{likelihood} \times \text{prior}$$

where $\pi(y|\phi)$ denotes the likelihood function, $g(\phi)$ is the prior and $\pi(\phi|y)$ is the posterior distribution.

3.5.4 Prior and Posterior

In this research we desire to use a non-informative prior for our multinomial logistic regression model, we choose the multivariate normal distribution as a prior distribution for the mean parameters β, that is we assume that $\beta_0 \propto N(\beta_0, \sum_0)$ where β_0 and $\sum_0$ are hyper-parameters.

$$\boldsymbol{\beta} \sim MVN(\beta_0, \sum_0)$$

we can write the general formula for a prior distribution as follows:

$$\pi(\boldsymbol{\beta}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta}-\beta_0)^T \boldsymbol{\Sigma}_0^{-1}(\boldsymbol{\beta}-\beta_0)\right)$$

Where,

$$\boldsymbol{\sum}_0 = \begin{pmatrix} s_0^2 & \dots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \dots & s_1^2 \end{pmatrix} \quad \beta_0 = \begin{pmatrix} \beta_0 \\ \vdots \\ \beta_1 \end{pmatrix}$$

From Ebenezer and Lougue (2019) multivariate normal prior does not have to be made up from independent components, hence the posterior distribution will be multivariate normal $(\beta_1, \sum_1)$ where

$$\sum_1 = \sum_0^{-1} + \sum_{MLE}^{-1}$$

and

$$\beta_1 = \sum_1 | \sum_{MLE}^{-1} | \hat{\beta}_{MLE} + \sum_1 | \sum_0^{-1} | \beta_0$$

where $\sum_{MLE}$ is the covariance matrix of the maximum likelihood estimate (MLE) vector with inverse $\sum_{MLE}^{-1}$ that is defined as:

$$\sum_{MLE}^{-1} = X^T \sum_0^{-1} X$$

while $\hat{\beta}_{MLE}$ represents the maximum likelihood vector. After observing the data we can arrive at the posterior distribution

$$\pi(\boldsymbol{\beta}|y_i) \propto \pi(y_i|\boldsymbol{\beta})\pi(\boldsymbol{\beta})$$

where $\pi(y_i|\beta)\pi(\beta)$ is the likelihood function and the prior distribution. The likelihood has a form:

$$\pi(y|\boldsymbol{\beta}) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - b_{LS})^T \boldsymbol{\Sigma}_{LS}^{-1}(\boldsymbol{\beta} - b_{LS})\right)$$

Considering the fixed effects alone, the posterior distribution for a multinomial logistic regression is expressed as:

$$p(\boldsymbol{\beta}|y) \propto \exp\left(-\frac{1}{2}(\boldsymbol{\beta} - \beta_1)^T \boldsymbol{\Sigma}_1^{-1}(\boldsymbol{\beta} - \beta_1)\right)$$

3.5.5 Markov Chain Monte Carlo (MCMC)

In this section, we will provide the elementary notion of MCMC algorithm used to compute the posterior distribution described above, since integration over the product of likelihood (observed data) and a prior is often analytically infeasible.

According to Ntzoufras (2011) these MCMC methods are the reason why quantitative researchers are now able to accurately estimate posterior distributions of highly complicated models with ease. MCMC has made a significant contribution to the growth and dissemination of Bayesian theory Ntzoufras (2011).

The Markov chain methods include building MCMC that ultimately "converges" to the desired (stationary) distribution. (Ntzoufras, 2011). The target distribution in this dissertation is the posterior distribution $\pi(\phi|y)$. In the next Section, we explain how MCMC algorithms work.

3.5.6 Markov Chain Monte Carlo Algorithm

Let $\phi^1, \phi^2, ..., \phi^G$ be a sample of size G from the posterior distribution $\pi(\phi|y)$. A Markov Chain is a stochastic process defined by $\phi^1, \phi^2, ..., \phi^G$ such that

$\pi\left(\phi^{g+1}|\phi^{g},...,\phi^{1}\right)=\pi\left(\phi^{g+1}|\phi^{g}\right)$.

That is, the distribution of ϕ at time $g+1$ given all the preceding ϕ values for $g, g-1, ..., 1$ depends only on the value ϕ^{g} of the previous sequence g. As $g \to \infty$, the distribution ϕ^{g} converges to its equilibrium, which is independent of the initial value of the chain ϕ^{0} (Ntzoufras, 2011).

Metropolis Algorithm and Gibbs sampling are commonly used MCMC methods but their draw back is the use of random walk which results in slow convergence. This study will avoid the issue by considering the use of Hamiltonian Monte-Carlo which only need few thousand iterations for the same problem where Gibbs sampling would need millions of iterations.

3.5.7 Hamiltonial Monte Carlo (HMC)

HMC is an MCMC approach that generates efficient transformations spanning the posterior using the derivatives of the density functions that are being sampled (Betancourt and Girolami, 2015). It employs a numerical integration-based approximate Hamiltonian dynamics simulation.

Momentum Variable

Hamiltonian Monte Carlo introduces ρ as an auxiliary variable that samples from the posterior distribution.

$$\pi(\rho,\phi)=\pi(\rho|\phi)\pi(\phi) \tag{3.5.6}$$

In **Rstan** ρ does not depend on ϕ and it is distributed as,

$$\rho \sim \mathsf{MultiNormal}(0,\Sigma). \tag{3.5.7}$$

The Hamiltonian

According to Carpenter et al. (2017), the Hamiltonian is defined by density $\pi(\rho, \phi)$ such that:

$$\begin{aligned} H(\rho, \phi) &= -\ln\left(\pi(\rho, \phi)\right) \\ &= -\ln\left(\pi(\rho|\phi)\right) - \ln\left(\pi(\phi)\right). \\ &= T(\rho|\phi) + U(\phi), \end{aligned} \tag{3.5.8}$$

where

$$T(\rho|\phi) = -\ln\left(\pi(\rho|\phi)\right), \tag{3.5.9}$$

is the potential energy and,

$$U(\phi) = -\ln\left(\pi(\phi)\right), \tag{3.5.10}$$

expresses the kinetic energy.

Generating Transitions

Starting with the current value of parameter ϕ, we create a transition to a new state in two steps before submitting it to the Metropolis accept step. (Carpenter et al., 2017).

First, a momentum value is computed independently of the underlying parameter values,

$$\rho \sim \mathsf{MultiNormal}(0, M). \tag{3.5.11}$$

Next, the joint system made up of the current parameter values and new momentumϕ is evolved via Hamilton's equations,

Following that, Hamilton's equations are used to develop the joint system (ϕ, ρ) made up of the existing parameter values ϕ and new momentum ϕ.

$$\frac{d\phi}{dt} = +\frac{\partial H}{\partial \rho} = +\frac{\partial T}{\partial \rho} \tag{3.5.12}$$

$$\frac{d\rho}{dt} = -\frac{\partial H}{\partial \phi} = -\frac{\partial T}{\partial \phi} - \frac{\partial U}{\partial \phi}. \tag{3.5.13}$$

$$\frac{d\phi}{dt} \quad = +\frac{\partial T}{\partial \rho}\frac{d\rho}{dt} \quad = -\frac{\partial U}{\partial \phi}. \tag{3.5.14}$$

To solve these differential equations, the **R** package **Rstan** uses a numerical integration called method called leapfrog. From there the algorithm move to Metropolis acceptance step, once that is done, we either update to a new state (ϕ^*, ρ^*) or maintain the current state (ϕ, ρ) (Hoffman et al., 2014).

3.5.8 Assessing and Improving Markov Chain Monte Carlo Convergence

It is crucial to figure out how many iterations to use to describe the posterior density to make sure the Markov chain has converged. However it is worth noting that a model's convergence does not always mean that it's a strong model. Model evaluation is just the beginning. These tests are used to assess whether the algorithm has achieved its target distribution (posterior distribution).

- **Autocorrelation Function (ACF) Plots** (Lesaffre and Lawson, 2012) stated that non vanishing Autocorrelation at long lags means

that each iterate provides less knowledge about the posterior, implying that a large sampling size is needed to cover the parameter space. Autocorrelation is a situation when there exists a correlation between model parameters in the MCMC. Usually autocorrelation can be eliminated by storing every i^{th} iteration, that process is called "thinning"(Congdon, 2010). Thinning reduces MCMC error and storage requirements especially when long runs are being carried out (Larsson and Stanfors, 2014).

- **Kernel Density Plots Kernel Density Plots**: A more satisfactory density plot for a converged chain would look more bell-shaped or parameters whose marginal posterior densities are approximately normal.

- **Gelman and Rubin Multiple Chain Convergence** :The use of two or more parallel chains with distinct beginning values is the basis for Gelman and Rubin convergence diagnostics (Lesaffre and Lawson, 2012). Lesaffre and Lawson (2012) stated that multiple chain convergence diagnostics provide evidence for the robustness of convergence across different subspaces. Standardizing variables and the unstructured random effect can help MCMC Chains to converge.

3.5.9 Software for Bayesian Data Analysis

There are various software programs which can fit models using the Bayesian approach. We show how to utilize **RStan** for this study, which uses Hamiltonian Monte Carlo (HMC) techniques instead of the Gibbs samplers and Metropolis-Hastings algorithms used in previous packages, such as BUGS and MLwiN (Browne and Rasbash, 2009). For complicated models, Hamil-

tonian Monte Carlo estimation offers significant advantages, see McElreath (2020) for an introduction of the approach that includes tips on interpretation and convergence diagnostics. The random effects logistic regression model will be fitted using HMC sampling in **Rstan**. The data preparation for the Bayesian random effects logistic regression models will be done in **R** (Core Team, 2013).

3.6 Criteria for Model Selection

In this section, we describe methods used to select the best fitting model from a set candidate models. Though technology to fit complex models through the Bayesian hierarchical models is widely available, there is no clear criteria to compare models and select best models. The most widely used criteria is how to measure and appropriately penalised the complexity of a hierarchical model.

3.6.1 Akaike Information Criterion (AIC)

One of the popular criterion for models comparison is the Akaike Information Criterion(AIC). The test was developed by Akaike (1973) for the aim of picking the best model from a pool of alternative models. The AIC chooses the model that minimizes the gap between fitted and anticipated true values, and it has the form,

$$AIC = -2\ln(L) + 2k \tag{3.6.1}$$

The AIC, on the other hand, has a habit of selecting models with an excessive number of parameters in cases of large samples.

3.6.2 Bayesian Information Criterion (BIC)

Schwarz et al. (1978) suggested the Bayesian Information Criterion, which was a modified version of the AIC when reasoning from a Bayesian perspective. This information criterion is defined by,

$$BIC = -2\ln(L) + k\ln(n), \tag{3.6.2}$$

where n is the sample size, L is the maximized likelihood and k is the number of regressors including the intercept.

3.6.3 Deviance Information Criterion (DIC)

The most commonly and widely used criteria for comparing hierarchical models is the Deviance Information Criterion proposed by Spiegelhalter et al. (2003). The DIC works in a similar manner like that Bayesian Information Criterion (BIC) (Schwarz et al., 1978). The DIC includes terms for both the fit and the complexity of a model. Spiegelhalter et al. (2003) proposed to estimate k. Given the likelihood function, $\pi(y|\phi)$, the deviance is usually defined as $D(\phi) = -2\ln\pi(y|\phi)$ and the posterior average deviance $\hat{D}$.

$$\begin{aligned} DIC &= kD + \hat{D} \\ &= 2\hat{D}(\phi) - D(\hat{\phi}), \end{aligned} \tag{3.6.3}$$

where $D(\hat{\phi})$ is the deviation calculated using the parameters' posterior mean. For non-hierarchical models, the DIC is seen as a generalization of the Akaike's Criterion (AIC), where DIC $\approx$ AIC.

3.6.4 Widely Applicable Information Criterion (WAIC)

According to Gelman et al. (2014) WAIC is given by:

$$lpd = \sum_{i=1}^{n} \log \left[\frac{1}{S} \sum_{s=1}^{S} p(y_i|\phi_s) \right] \tag{3.6.4}$$

$$P_{waic} = \sum_{i=1}^{n} var[\log[p(y_i|\phi)]] \tag{3.6.5}$$

$$WAIC = -2(lpd - P_{waic}), \tag{3.6.6}$$

where lpd is the natural log of DIC. The number of simulation draws S is usually assumed to be big enough to completely capture the posterior distribution. Gelman et al. (2014).

Chapter 4

Data Analysis

Introduction

This chapter presents sources of data and the data analysis. The chapter is divided into two sections. The first section describes the source of data used. In the second section, using the methodology explained in chapter 3, the captured data from the qualitative research is presented, analysed, described and interpreted, followed by a discussion of the research findings.

4.1 Data collection

4.1.1 Sources of data

The study uses a quantitative research method. The South African Demographic and Health Survey (SADHS) conducted from January to September 2016 is the major source of data used in this study. The data was downloaded free of charge from http://www.measuredhs.com after permission was granted to do so by USAID − Macro International. The SADHS provided a secondary data for this study, and only that will be used for analysis

without any other primary data to complement it. Questionnaires were administered to all sampled women in each of sampled households, collecting data on fertility, family planning, and child health, in addition to demographic and socio-economic data. The 2016 (SADHS) is the latest South African national data-set with information on contraceptive use. The data were intended for use by programme managers and policy-makers in order to evaluate and improve family planning and maternal and child health programmes.

4.1.2 Target population

The population was reduced to women under risk of conception, that is non pregnant sexually active women by the time of the survey. This thesis define sexually active women as "Women who had a sexual intercourse 4 weeks before the survey". All non users of contraceptives were excluded in the analysis, this reduced the sample to 4025 sexually active women coming from 689 clusters.

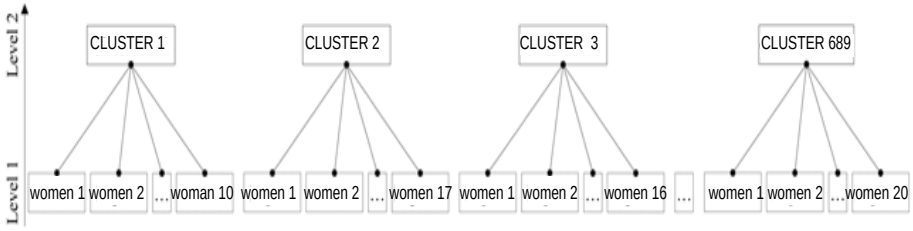


Figure 4.1: Structure of the data considered for analysis.

4.2 Exploratory data analysis (EDA)

Exploratory data analysis (EDA) is crucial and should be a first step in the the whole process of analysing data. EDA helps one to understand the relationships or correlations that may exist between the variables in the data, as well as any anomalies that might arise.

4.2.1 Response Variable

Table 4.1 below displays the percentage distribution of sexually active women in South Africa by the contraceptive method they currently use. Overall, the most popular methods are injections (50.04%) and Oral (23.88%). Less than 10% use a more permanent contraceptive methods (9.29% use long-acting methods (IUD/sterilization/implants)).

Table 4.1: Frequency and percentage distribution of sampled woman by contraceptive method currently used.

Contraceptive Method choice	Frequency	percentage (%)
Injection	2014	50.04
Oral	961	23.88
Barrier	652	16.20
Sterilization/IUD/Implants	374	9.29
Others	24	0.60
Total	4025	100

4.2.2 Explanatory Variables

Age group

Table 4.2 exhibits age divided into 6 categories. South Africa is considered a young population by age structure. In order to reflect the age structure, substantial proportion (proximately 49%) of the sampled women were between the ages of 15 and 30 years. As can be seen in the table below, the proportion of respondents in each age group decreases as they get older.

Table 4.2: Frequency and percentage distribution of respondents' age groups.

Age group	Frequency	percentage (%)
$15-24$	1147	28.5
$25-29$	833	20.70
$30-34$	734	18.23
$35-39$	567	14.09
$40-44$	443	11.01
$45-49$	301	7.48
Total	4025	100

Marital status

Individual women's marital status is observed as a binary attribute in which the respondent is either married/in-union or single/not in-union. In this study in-union category comprises both married women and any woman who is living together with her sexual partner. The marital status of the respondents is predominantly not in union, accounting for slightly over 60% of the sample (see Table 4.3).

Table 4.3: Frequency and percentage distributions of respondents' marital status.

Marital status	Frequency	percentage (%)
Not in union	2495	61.99
In union	1530	38.01
Total	4025	100

Number of children

Table 4.4 report the frequency and percentage of respondents in each of the number of children categories. The results in the table below indicate that a significant proportion of the sampled women has at most 3 children. Which is very close to South Africa's TFR (total fertility ratio) of 2.3 births per woman (United Nations, 2019).

Table 4.4: Frequency and percentage(%) distribution of women by number of children they have.

No. of living children	Frequency	percentage (%)
0	608	15.11
1	1186	29.47
2	1158	28.77
3	646	16.05
4+	427	10.61
Total	4025	100

Community aggregates

The SADHS did not gather data that could be used to explicitly portray the clusters' characteristics. Nonetheless, we aggregated individual woman's characteristics within their clusters to generate community variables.

Table 4.5 presents statistical summaries of some of the key determinants

Table 4.5: Summary statistics for aggregated cluster level covariates.

variables	No. of clusters	Mean	SD
Educated woman within PSU(%)	689	12.02	19.01
Cluster mean age at first sex	689	17.37	1.59
Black woman within PSU(%)	689	87.38	30.09

of contraceptive choices, aggregated at community level to get level two estimates of the determinants of contraceptive choices. In this thesis, percentage of educated woman at community level is defined as the percentage of woman who have attained any post secondary level qualification. Female educational attainment was disappointingly low with an average of 12% of educated woman per community. For a total of 4025 sexually active women that live in 689 communities of South Africa. On average 87.38% of sampled woman in each community were black. Relatively smaller promotions were sampled from whites and other races to mirror South Africa's population which is predominantly black.

4.3 Bivariate Analyses

Figure 4.2 below shows the distribution of each community level or cluster specific variables on the diagonal. Below the main diagonal are the bivariate scatter plots of the covariates with a fitted line. The correlation value and its p-value are displayed above the diagonal.

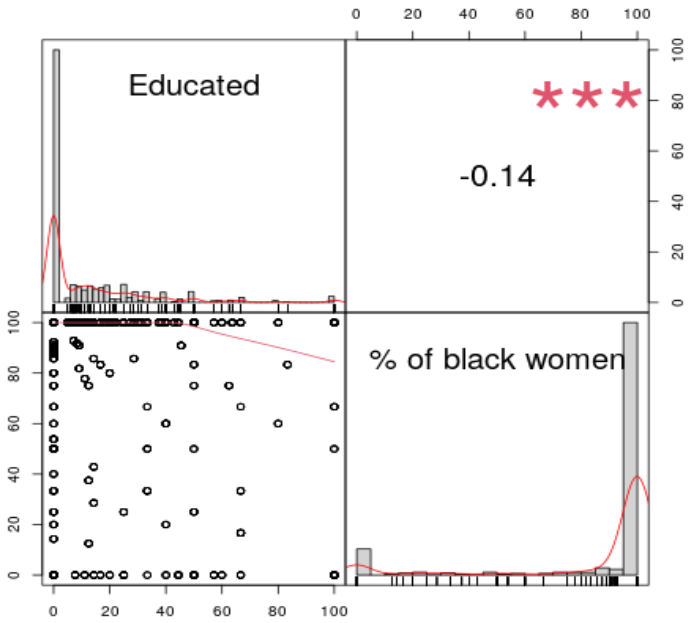


Figure 4.2: Distributions and Correlations of the aggregated level 2 covariates.

A p-value significance is denoted by: (0 '***'0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' '1). The figure was obtained using the **PerformanceAnalytics** package in R. From the figure we observe that there is significant correlation between the covariates. We also observe that there is negative correlation between percentage of black woman in a community and percentage of educated woman within the community. This denotes that a community with more black women is associated wit less percentage of educated woman.

The correlation $\rho = -0.14$ between the community level covariates, indicates that the independent variables do not depend on one another, thus we wont have to worry about multicollinearity, which is often a problem in regression models.

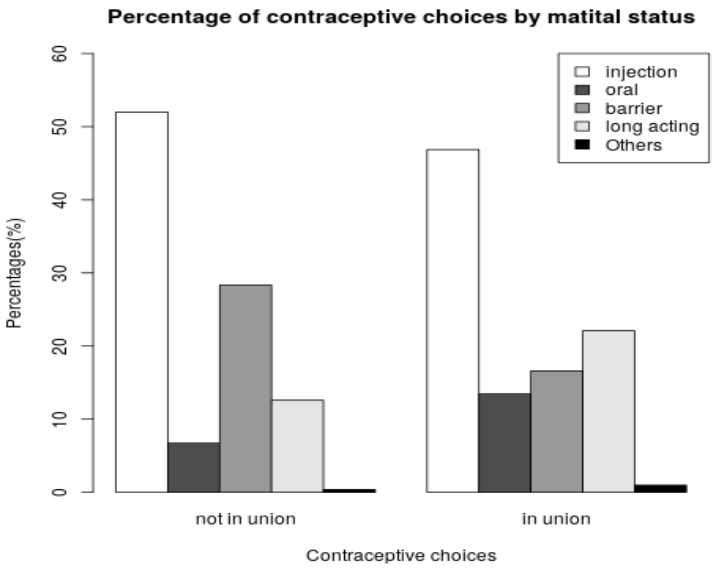


Figure 4.3: Bar plot of Contraceptive choices by marital status.

Now coming to contraceptive choices by marital status, for both in union and not in union women, injections shows a continued dominance, constituting (approximately 52% and 48% respectively).

Table 4.6: Cross tabulation for marital status by number of children.

	Marital status			
	Not inunion		In union	
No. of children	Number	%	Number	%
0	554	13.76	54	1.34
1	952	22.98	261	6.48
2	667	15.81	551	13.68
3	268	6.66	378	9.39
4+	141	3.50	286	7.11

Table 4.6 displays marital status by number of children contingency ta-

H

Table 4.7: Cross tabulation depicting relationships between contraceptive choices and some demographic and maternal characteristics.

Variables	Contraceptive choices					Df	χ^2	p
	s	Oral	Barrier	Long-acting	Others			
Age group						20	272.86	< 0.001
15-24	684	74	272	112	5			
25-29	475	84	184	84	6			
30-34	374	84	159	112	5			
35-39	237	60	153	113	4			
40-44	166	45	96	134	2			
Marital Status						4	164.93	< 0.001
Not in-union	1297	168	707	314	9			
In union	717	206	254	388	15			
No of children						16	339.46	< 0.001
0	238	49	272	45	4			
1	678	96	277	127	8			
2	621	140	213	178	6			
3	289	59	121	172	5			
4+	188	30	78	130	1			

ble. Considerable proportion of women with at least 2 children are in union, and the pattern is reversed for not in union woman, with majority(36%) accounting for less than 2 children.

4.3.1 Chi-squre(χ^2) Test of Association

The results of the χ^2 are presented in this section. The test was performed on all categorical potential independent variables. Table 4.7 summarises the results of the cross tabulation.

Age groups is an important factor in answering the questions about the contraceptive method choice. The results in Table 4.7 indicate that the test statistic is $\chi^2(20) = 272.86$, with the corresponding p of the test statistic $p < 0.001$. Since the $p < \alpha$, this is enough evidence to conclude that age groups and contraceptive method choices are significantly associated.

Positive association was found between current number of children a woman has and her choice of contraceptive method ($\chi^2(16) = 339.46$,$p < 0.001$). Another statistically significant association was found between contraceptive choice and marital status with the Chi-square test $\chi^2(4) = 164.93$ and $p < 0.00$.

4.4 Classical Multivariate Analysis

In this section, we present fitted models and their estimated effects. The MIXED command in SPSS is used to fit the multilevel multinomial models. These multilevel multinomial regression models were considered to predict the probability of a woman choosing any other contraceptive method over injection. Since the outcome variable is nominal, we consider a generalised logit link. All fitted models will be based on the model building strategies laid out in Table 4.8.

Table 4.8: Detailed description of Model constructing strategy.

Model A	Model B	Model C
No predictors, but only random effects for the cluster	Model A+ level 1 fixed effects related to Demographic and maternal characteristics of an individual women such as Age, Number of children ever born (1,2,3 and 4 or more) and Marital status.	Model B + level 2 fixed effects or community level characteristics, which include percentage of educated woman within PSUs, Mean age at first sexual intercourse and Percentage of Black women within PSUs
This model captures the percentage variation in contraceptive choice explained by level 2 units(Communities)	Results indicate the relationship between individual level predictors and the choice of contraception	The results show whether including community level variables enhances model fit

4.4.1 Classical Multilevel Multinomial Empty Model

Here we investigate a random intercept model in which just the intercept is used to fit the model (Model A), this model is often refereed to as the

variance component model. The results of fitting a random intercept empty model are summarised in Table 4.9.

Table 4.9: Classical Multilevel multinomial random intercept Model A.

Covariates	$\boldsymbol{\beta}$	SE
Fixed effects		
Oral (intercept)	-1.691**	0.062
Barrier(intercept)	-0.802**	0.049
Long-acting(intercept)	-1.130**	0.050
Others(intercept)	-4.506**	0.227
Random effects Varience of random errors at cluster level: var(u_0)=σ^2_{u0}	0.67**	0.070
Model fit		
-2*log likelihood	69879.335	
AIC	69887.35	
BIC	69912.48	

Significance codes, * $p < 0.1$; ** $p < 0.01$

From the results presented in Table 4.9, it is observed that the community level variance σ^2_{u0} is estimated as 0,67. In logit models, the residual variance σ^2_ϵ of an individual women (level 1) within a cluster is zero but the variance is considered fixed at $\pi^2/3(3.29)$. Hence the intra-cluster correlation coefficient

$$ICC = \sigma^2_u/(\sigma^2_u + \sigma^2_\epsilon) = 0.67/(0.67 + 3.29) = 0.169$$

According to Heck and Thomas (2015), $ICC = 0.05$ is the "cuttoff" for the evidence of clustered observations. Thus, our computed ICC estimate of 0.169 or approximately 17% is indicative of substantial clustering between communities. The clustering variability greater than 10% which is due to between group differences justifies the need for a multilevel modelling, since ignoring it may result in overstatement of the significance of our model parameters.

4.4.2 Results of Random Intercept Multinomial Models

Using our modelling process described in Table 4.8, we add various combinations of fixed effects to Model A in order to get Model B and Model C. In Model B, we add individual level characteristics such as age group, marital status and no children ever born. Finally, to get model C we augment Model B by adding community level variables. For all three different models we consider a simple random structure (i.e., only a single random intercept for each cluster, $u_{0jk} = u_{0j}$ for all $k = 1, \ldots, 5$ contraceptive choice categories).

For simple comparison only parameter estimates of the fixed effects of Three fitted models are presented in Table 4.10 below.

Model A

$$\begin{aligned} \eta_{ijk} = \log\left(\frac{\pi_{ijk}}{\pi_{ij4}}\right) \quad &= \beta_{0j} \\ &= \gamma_{00} + u_{0j} \end{aligned}$$

Model B

$$\begin{aligned} \eta_{ijk} = \log\left(\frac{\pi_{ijk}}{\pi_{ij4}}\right) \quad &= \beta_{0j} + \beta_{1j} Age(25-29)_{ijk} \\ &+\beta_{2j} Age(30-34)_{ijk} + \beta_{4j} Age(35-39)_{ijk} \\ &+\beta_{4j} Age(40-44)_{ijk} + \beta_{5j} Age(45-49)_{ijk} \\ &+\beta_{6j} Marital(inunion)_{ijk} + \beta_{7j} Children(1)_{ijk} \\ &+\beta_{8j} Children(2)_{ijk} + \beta_{9j} Children(3)_{ijk} \\ &+\beta_{10j} Children(4+) \end{aligned}$$

$$\beta_{0j} \quad = \gamma_{00} + u_{0j}$$

Model C

$$\begin{aligned}
\eta_{ijk} = \log\left(\frac{\pi_{ijk}}{\pi_{ij4}}\right) &= \beta_{0j} + \beta_{1j}Age(25-29)_{ijk} \\
&+\beta_{2j}Age(30-34)_{ijk} + \beta_{4j}Age(35-39)_{ijk} \\
&+\beta_{4j}Age(40-44)_{ijk} + \beta_{5j}Age(45-49)_{ijk} \\
&+\beta_{6j}Marital(inunion)_{ijk} + \beta_{7j}Children(1)_{ijk} \\
&+\beta_{8j}Children(2)_{ijk} + \beta_{9j}Children(3)_{ijk} \\
&+\beta_{10j}Children(4+) \\
\beta_{0j} &= \gamma_{00} + \gamma_{01}(cluster_{educated})_j + \gamma_{02}(cluster_{ageAtSex})_j \\
&+\gamma_{03}(cluster_{blacks})_j + u_{0j}
\end{aligned}$$

The reference category is "injection"; and so, we estimate a model for "Oral" relative to "injection", "barrier" relative to "injection", "long-acting" relative to "injection" and again a model for "others" relative to "injection". The parameter estimates in model A, model B and model C above are presented in Table 4.10 below.

Table 4.10: Fixed effects of Three fitted classical multilevel multinomial logit random intercept Models.

Characteristics		Model A	Model B	Model C
Oral	Intercept	-1.691**	-1.832**	-2.759**
Age	45-49		1.454**	1.241**
	40-44		1.209**	1.012**
	35-39		1.126**	0.967**
	30-34		0.879**	0.747**
	25-29		0.662**	0.564**
Marital status	In union		0.772**	0.620
Number of children	4+		-1.671**	-1.299**
	3		-1.239**	-0.952**
	2		-0.898**	-0.700**
	1		-0.813**	-0.666**
% of educated woman				0.018**
Mean age at first sex				0.049
% of black woman				-0.002
Barrier	Intercept	-0.802**	-0.092	-3.800
	45-49		2.678**	2.661**
	40-44		1.820**	1.751**
	35-39		1.909**	1.885**
	30-34		1.271**	1.224**
	25-29		0.815**	0.775**
	In union		-0.407**	-0.349**
Number of children	4+		-2.531	-2.539**
	3		-2.276**	-2.279**
	2		-2.198**	-2.223**
	1		-1.556**	-1.590**
% of educated women				0.010**
mean age at first sex				0.133**
% of Black women				0.015**
Long-acting	Intercept	-1.130**	-1.745	-3.270**
Age	45-49		1.667	1.395**
	40-44		1.279**	1.044**
	35-39		0.766**	0.575**
	30-34		0.397*	0.254
	25-29		-0.017	-0.115
Marital status			0.225*	0.055
Number of children	4+		0.244	0.674
	3		0.326	0.632**
	2		-0.140	0.049
	1		-0.167	-0.027
%of educated woman				0.017**
Mean age at first sex				0.106**
% of black women				-0.006**
Others	Intercept	-4.506**	-4.601	-7.908**
Age	45-49		1.704	1.227
	40-44		0.924	0.526
	35-39		1.274	0.941
	30-34		0.852	0.530
	25-29		0.781	0.524
Marital status	In union		1.235*	0.991*
No of children	4+		-2.827	-1.987
	3		-1.411	-0.750
	2		-1.731*	-1.276
	1		-0.872	-0.492
% of educated woman				0.026**
Mean age at first sex				0.172
% of Black women				-0.002

Significance codes, * $p < 0.1$; ** $p < 0.01$

Results of the multilevel multinomial logistic regression models in Table 4.10, presents parameter estimates of three fitted models. The models were fitted in such a way that model A is nested under model B and model B is nested under Model C . The reason for that was to make it possible to use The likelihood ratio hypotheses test for nested models to compare the three models based on their deviances (-2*log likelihood).

4.4.3 Model Comparison and Selection

In this section We verify that the statistical models fitted to the data is appropriate by assessing goodness of fit tests described in section 3.6 of Chapter 3

Table 4.11: Results of model comparison statistics

Covariates	Model A	Model B	Model C
Random effects			
Error variance of the random intercept	0.670 **	0.601 **	0.532**
ICC	0.169	0.151	0.139
Model fit			
-2*log likelihood	69879.34	68686.04	65961.76
AIC	69887.35	68694.05	65969.75
BIC	69912.48	68719.20	65994.93

Significance codes, * $p < 0.1$; ** $p < 0.01$

Having accounted for the three explanatory variables in Model B we notice reduction in the variability of the group levels. On the basis of deviance (-2*log likelihood) which is much smaller in Model C as compared to other fitted models, this suggests that Model C fit data better than Model A and Model B as the drop in deviance is statistically significant. Other comparison measures like AIC and BIC values leads to similar conclusions, Hence we choose Model C to model the determinants of contraceptive method choices. The full results of Model C are presented in the Table 4.12 below.

Table 4.12: Fixed effects of the frequentist multinomial logistics regression.

Variables		$\hat{\beta}$	SE	Pr($>\|z\|$)	95% CI of $\hat{\beta}$		$p-value$
					Lower	Upper	
Oral	Intercept	-2.759	0.7263	-3.799	-4.183	-1.335	0.000**
Age	45-49	1.241	0.2977	4.169	0.658	1.825	0.000**
	40-44	1.012	0.2517	4.023	0.519	1.506	0.000**
	35-39	0.967	0.2321	4.164	0.512	1.422	0.000**
	30-34	0.747	0.2083	3.584	0.338	1.155	0.000**
	25-29	0.564	0.1934	2.918	0.185	0.943	0.004**
Marital status	In union	0.620	0.1327	4.667	0.359	0.880	0.000**
Number of children	4+	-1.299	0.3128	-4.152	-1.912	-0.686	0.000**
	3	-0.952	0.2682	-3.548	-1.477	-0.426	0.000**
	2	-0.700	0.2302	-3.039	-1.151	-0.248	0.002**
	1	-0.666	0.2120	-3.141	-1.082	-0.250	0.002**
% of educated woman		0.018	0.0030	5.921	0.012	0.024	0.000**
Mean age at first sex		0.049	0.0391	1.250	-0.028	0.125	0.211
% of black women		-0.002	0.0019	-1.083	-0.006	0.002	0.279
Barricidal	Intercept	-3.800	0.6733	-5.643	-5.120	-2.480	0.000**
Age	45-49	2.661	0.2160	12.319	2.238	3.085	0.000**
	40-44	1.751	0.1920	9.120	1.375	2.128	0.000**
	35-39	1.885	0.1724	10.938	1.547	2.223	0.000**
	30-334	1.224	0.1567	7.811	0.917	1.531	0.000**
	25-29	0.775	0.1411	5.489	0.498	1.051	0.000**
Marital status	In union	-0.349	0.1051	-3.322	-0.555	-0.143	0.001**
Number of children	40+	-2.539	0.2213	-11.475	-2.973	-2.105	0.000**
	3	-2.279	0.1925	-11.838	-2.657	-1.902	0.000**
	2	-2.223	0.1642	-13.537	-2.545	-1.901	0.000**
	1	-1.590	0.1385	-11.481	-1.861	-1.318	0.000**
% of educated woman		0.010	0.0029	3.560	0.005	0.016	0.000**
Mean age at first sex		0.133	0.0353	3.765	0.064	0.202	0.000**
% of black women		0.015	0.0022	6.605	0.010	0.019	0.000**
Long-acting	Intercept	-3.270	0.6375	-5.129	-4.520	-2.020	0.000**
Age	45-49	1.395	0.2231	6.253	0.958	1.833	0.000**
	40-44	1.044	0.1960	5.325	0.659	1.428	0.000**
	35-39	0.575	0.1918	3.000	0.199	0.951	0.003**
	30-34	0.254	0.1788	1.421	-0.096	0.605	0.155
	25-29	-0.115	0.1738	-0.662	-0.456	0.226	0.508
Marital status	In union	0.055	0.1080	0.505	-0.157	0.266	0.613
Number of children	4+	0.674	0.2550	2.645	0.175	1.174	0.008**
	3	0.632	0.2377	2.660	0.166	1.098	0.008**
	2	0.049	0.2200	0.223	-0.382	0.480	0.824
	1	-0.027	0.2023	-0.135	-0.424	0.369	0.893
% of educated woman		0.017	0.0027	6.354	0.012	0.022	0.000**
Mean age at first sex		0.106	0.0341	3.120	0.039	0.173	0.002**
% of Black women		-0.006	0.0015	-4.281	-0.009	-0.004	0.000**
Others	Intercept	-7.908	2.5213	-3.136	-12.851	-2.965	0.002**
Age	45-49	1.227	1.0025	1.224	-0.738	3.193	0.221
	40-44	0.526	0.9689	0.543	-1.373	2.426	0.587
	35-39	0.941	0.8202	1.147	-0.667	2.549	0.251
	30-34	0.530	0.7510	0.706	-0.942	2.003	0.480
	25-29	0.524	0.6675	0.785	-0.785	1.833	0.433
Marital status	In union	0.991	0.4940	2.006	0.023	1.960	0.045
Number of children	4+	-1.987	1.2875	-1.544	-4.512	0.537	0.123
	3	-0.750	0.8688	-0.863	-2.453	0.954	0.388
	2	-1.276	0.7915	-1.612	-2.828	0.276	0.107
	1	-0.492	0.6842	-0.719	-1.833	0.850	0.472
% of educated woman		0.026	0.0086	2.982	0.009	0.043	0.003**
Mean age at first sex		0.172	0.1312	1.312	-0.085	0.429	0.190
% of black woman		-0.002	0.0063	-0.339	-0.014	0.010	0.735

The only level 2 variable that was continuously associated with the contraceptive choice methods was education within communities. Among level 1 (or individual variables), all factors were significantly associated with the use of oral or barrier choices respectively. In order to be able to facilitate the interpretation of the model parameters, we further evaluated the odds ratios for the estimates that are presented in Table 4.12. The results are displayed in Table 4.13.

Table 4.13: Odds ratios and 95% confidence interval of the odds of parameters from a classical multilevel-multinomial model, estimating contraceptive method choice among women in South Africa.

Parameters ($\boldsymbol{\beta}$)		Oral			Barrier			Long-acting			Others		
		Exp($\hat{\boldsymbol{\beta}}$)	95% CI for Exp($\boldsymbol{\beta}$)		Exp($\hat{\boldsymbol{\beta}}$)	95%CI for Exp($\boldsymbol{\beta}$)		Exp($\hat{\boldsymbol{\beta}}$)	95%CI for Exp($\boldsymbol{\beta}$)		Exp($\hat{\boldsymbol{\beta}}$)	95%CI for Exp($\boldsymbol{\beta}$)	
			Lower	Upper		Lower	Upper		Lower	Upper		Lower	Upper
Intercepts		0.063	0.015	0.263	0.022	0.006	0.084	0.038	0.011	0.133	0.000	0.000	0.005
Individual variables													
Age groups	45 – 49	3.460	1.930	6.204	14.317	9.373	21.867	4.036	2.606	6.251	3.411	0.478	24.350
	40 – 44	2.752	1.680	4.508	5.726	3.954	8.395	2.839	1.934	4.169	1.692	0.253	11.311
	35 – 39	2.629	1.668	4.144	6.588	4.699	9.236	1.778	1.221	2.589	2.563	0.513	12.796
	30 – 34	2.110	1.402	3.174	3.401	2.501	4.624	1.289	0.908	1.831	1.699	0.390	7.410
	25 – 29	1.758	1.203	2.569	2.170	1.645	2.862	0.891	0.634	1.253	1.689	0.456	6.250
	15 – 24												
Marital Status	in union	1.858	1.432	2.410	0.705	0.570	0.867	1.056	0.855	1.305	2.694	1.023	7.098
	not inunion												
Number of childern	4+	0.273	0.148	0.504	0.079	0.051	0.122	1.936	1.191	3.236	0.137	0.011	1.711
	3	0.386	0.228	0.653	0.102	0.070	0.149	1.882	1.181	2.999	0.473	0.086	2.595
	2	0.497	0.316	0.780	0.108	0.078	0.149	1.050	0.682	1.617	0.279	0.059	1.317
	1	0.514	0.339	0.779	0.204	0.156	0.268	0.973	0.655	1.447	0.612	0.160	2.339
	0												
Community level variables													
% of educated woman within PSU		1.018	1.012	1.024	1.010	1.005	1.016	1.017	1.012	1.022	1.026	1.009	1.043
Mean Age at First sexual intercourse		1.050	0.973	1.134	1.142	1.066	1.224	1.112	1.040	1.183	1.188	0.918	1.536
% of black woman within PSU		0.998	0.994	1.002	1.016	1.010	1.019	0.994	0.991	0.996	0.998	0.986	1.010

Table 4.13 presents the 95% confidence interval of the odds ratio (OR) from the classical multilevel multinomial random intercept regression model. A significant association exists between contraceptive method choices and marital status, age group, number of children a woman have, percentage of educated women in a community(cluster), community specific mean age at first sex and percentage of black women in a community. For interpretation this study will use a combination of both point estimates of the odds and their 95% confidence interval to interpret the results, because confidence interval does not only point out the region/point where the parameter is most likely estimated, it also gives information about the accuracy of the estimate. In that sense it contains more information which makes it more meaningful.

The results show that, controlling for all other variables in the model, we can say with 95% confidence that, relative to women aged $15-19$, the odds of using any method other than injections increase with age, peaking at $45-49$, for the first three choices; OR=3.460, 95%CI [1.93, 6.20]; OR = 14.317, 95%CI [9.37, 21.87]; OR=4.036, 95%CI [2.61, 6.25]; for oral, barrier and IUD/sterilization/implants method choices respectively. Our results are in line with the findings from (Stephenson et al., 2007)
In union women had the reduced odds of using barrier methods over injection, which is understandable since married people would be looking into building a family.

Compared to not in union women, the chance of in union women using oral over injection was slightly higher than that of not in union women (OR = 1.86 , 95% CI [1.43, 2.41]). Moreover, relative to woman with less than 1 children, the odds of a woman with 4 or more children using long increases

with number of children; (OR = 1.936 , 95% CI [1.91, 3.23]).

Interestingly, at the community level, the likelihood of using any other method over injection was associated with a higher percentage of well educated women within the community; OR = 1.018, 95%CI [1.01, 1.02]; OR=1.010, 95%CI [1.01, 1.02]; OR = 1.017, 95%CI [1.01, 1.02]; OR = 1.026, 95%CI [1.01, 1.04]; for oral, barrier, UID/sterilization/implants and other methods respectively. This demonstrates that educated woman are more likely to be selective rather than going with the majority.

Furthermore, among women who came from communities in which the mean age at first sex was higher it was more common to use Barrier (OR=1.142, 95%CI [1.07, 1.22]), IUD/sterilization/implants (OR=1.11, 95%CI [1.04, 1.18]), compared to injection. For the random effects, the results give the cluster level variability, with the non-zero intra-cluster correlation coefficient, this justifies the multilevel approach for analysis. Based on the result of estimated intercept for clusters, there exists statistically significant variation between the different contraceptive choices that woman can access in South Africa. The co-variance parameter estimate was used for the computation of intraclass correlation coefficient.

$$ICC = \frac{\tau_{00}}{\tau_{00} + \sigma^2} = 0.139$$

The result of intraclass correlation coefficient indicates how much of the total variation in the likelihood of woman choosing a particular form of contraception over non using. The intraclass correlation coefficient is calculated as 0.139 representing about 14% of the total variation in the outcome variable is accounted for by Clusters.

The cluster-level random intercept term remained significant for all method-choice categories after the inclusion of the individual level variables in the model.

4.5 Results of Bayesian Modelling Framework

In this section we present the results of Bayesian multilevel multinomial logistic regression models in explaining the relationship between demographic, maternal characteristics of women and their contraceptive choices. The Bayesian multilevel multinomial logistic regression models were fitted using **rstan** package within **R** software. We ran two chains for 10000 iterations with the first 5000 discarded as a burn in period. The NUST sampler extends static Hamiltonian Monte Carlo chains via automated tuning: the user need not to specify neither number of steps nor step size. NUTS determines the number of steps via a sophisticated tree building algorithm, which we briefly described in chapter 3, We set flat multivariate normal priors on fixed effects.

4.5.1 Bayesian Empty Model

The results of a Null or empty model are summarized in Table 4.14

Table 4.14: Bayesian Multilevel multinomial epty model (Model I).

Covariates	$\boldsymbol{\beta}$	SE
Fixed effects		
Oral (intercept)	-1.85	0.00
Barrier(intercept)	-0.90	0.00
Long-acting(intercept)	-1.20	0.00
Others(intercept)	-5.97	0.01
Random effects		
Varience of random errors at cluster level: var(u_0)=$\sigma^2_{u_0}$	1.05	0.01
ICC	0.242	
Model fit		
Widely Applicable Information Criterion (WAIC)	9574.4	

This model contains no predictors, it is focussed on the assessment of the heterogeneity of contraceptive choices among communities (clusters). The ICC shown in Table 4.14 gives an estimate of 0.242 or just over 24% of the variability in contraceptive choices is allocated between community differences. We now add the same explanatory variables as those used in the frequentist approach to get Model II and Model III respectively and the results are presented below.

Table 4.15: Summary of the Posterior Distributions for Multilevel-multinomial Models estimating the determinants of contraceptive method choice among women in South Africa.

Variables		Model I	Model II	Model III
Oral	intercept	-1.85	-1.93	-2.86
Age	25-29		0.67	0.59
	30-34		0.93	0.80
	35-39		1.20	1.03
	40-44		1.26	1.06
	45-49		1.53	1.30
Marital status	In union		0.74	0.63
No of children	1		-0.87	-0.72
	2		-0.97	-0.76
	3		-1.32	-1.04
	4+		-1.79	-1.42
% of educated woman				0.02
Mean age at first sex				0.05
% of black woman				0.00
Barrier	intercept	-0.90	-0.11	-4.03
Age	25-29		0.87	0.82
	30-34		1.37	1.31
	35-39		2.07	2.02
	40-44		1.95	1.86
	45-49		2.87	2.83
Marital status	In union		-0.45	-0.38
Number of children	1		-0.70	-1.71
	2		-2.39	-2.39
	3		-2.49	-2.47
	4+		-2.76	-2.75
% of educated woman				0.01
Mean age at first sex				0.14
% of black woman				0.02
Long-acting	intercept	-1.20	-1.82	-3.33
Age	25-29		-0.03	-0.12
	30-34		0.40	0.26
	35-39		0.79	0.59
	40-44		1.32	1.07
	45-49		1.72	1.43
Marital status	In union		0.23	0.05
Number of children	1		-0.17	-0.03
	2		-0.14	0.04
	3		0.35	0.64
	4+		0.27	0.68
% of educated woman				0.02
Mean age at first sex				0.11
% of black women				-0.01
Others	intercept	-5.97	-6.63	-11.4
Ages	25-29		0.74	0.53
	30-34		0.80	0.49
	35-39		1.30	0.98
	40-44		0.63	0.24
	45-49		1.48	0.98
Marital status	In union		1.33	1.14
Number of children	1		-1.78	-0.44
	2		-1.74	-1.33
	3		-1.31	-0.69
	4+		-3.31	-2.57
% of educated woman				0.03
Mean age at first sex				0.23
% of Black women				-0.00

4.5.2 Model Assessment and Comparison

The most commonly used criterion DIC, is not implemented in **rstan** because it is not nearly as good as WAIC estimators produced by **loo** package.

Table 4.16: WAIC of Bayesian Multilevel models .

Model	WAIC (SE)	Effective parameters
Model I	9574.4 (112.97)	231.5
Model II	9267.0 (113.26)	324.5
Model III	8721.6 (123.68)	284.3

In the Bayesian paradigm, Table 4.16 above shows model diagnostics for all equipped models. A model with a low WAIC value fits the data well. Hence based on the comparison of model complexity and goodness of fit, we choose model III

4.5.3 Result of Final Bayesian Multilevel Multinomial Regression Model

Table 4.17 show parameters, the approximation of the average of the posterior distribution of the model parameter coefficient; an approximation of the standard deviation of the posterior distribution and computational accuracy of the mean. Furthermore it shows percentiles which include the 97.5^{th} percentile or an approximation of the upper endpoint of the 95% credible interval and the 2.5^{th} percentile or an approximation of the lower end point of the 95% credible interval. It also suggests that Markov chain has converged with $\hat{R} = 1$. That can also be backed by Mean errors which are low.

Table 4.17: Summary of the Posterior Distribution for a Multilevel-multinomial Model Parameters estimating the determinants of contraceptive method choice among women in South Africa

Covariates		Mean	SE_{mean}	2.5th%	97.5th%	$\hat{R}$
Oral	intercept	-2.86	0.10	-4.40	-1.36	1.00
Age	25-29	0.59	0.00	0.20	0.97	1.00
	30-34	0.80	0.00	0.038	1.21	1.00
	35-39	1.03	0.00	0.57	1.50	1.00
	40-44	1.06	0.00	0.56	1.57	1.00
	45-49	1.30	0.00	0.68	1.89	1.00
Marital status	In union	0.63	0.00	0.37	0.89	1.00
No of children	1	-0.72	0.00	-1.14	-0.29	1.00
	2	-0.76	0.00	-1.23	-0.30	1.00
	3	-1.04	0.00	-1.58	-0.49	1.00
	4+	-1.42	0.00	-2.07	-0.80	1.00
% of educated woman		0.02	0.00	0.01	0.03	1.00
Mean age at first sex		0.05	0.00	-0.03	0.13	1.00
% of black woman		0.01	0.00	-0.01	0.01	1.00
Barrier	intercept	-4.03	0.01	5.48	-2.6	1.00
Age	25-29	0.82	0.00	0.54	1.11	1.00
	30-34	1.31	0.00	1.00	1.63	1.00
	35-39	2.02	0.00	1.67	2.37	1.00
	40-44	1.86	0.00	1.48	2.26	1.00
	45-49	2.83	0.00	2.39	3.28	1.00
Marital status	In union	-0.38	0.00	-0.59	-0.10	1.00
Number of children	1	-1.71	0.00	-2.00	-1.42	1.00
	2	-2.39	0.00	-2.74	-2.05	1.00
	3	-2.47	0.00	-2.87	-2.07	1.00
	4+	-2.75	0.00	-3.21	-2.30	1.00
% of educated woman		0.01	0.00	0.00	0.02	1.00
Mean age at first sex		0.14	0.00	0.06	0.22	1.00
% of black woman		0.02	0.00	0.01	0.02	1.00
Long-acting	intercept	-3.33	0.11	-4.62	-2.0	1.00
Age	25-29	-0.12	0.00	-0.47	0.22	1.00
	30-34	0.26	0.00	-0.09	0.61	1.00
	35-39	0.59	0.00	0.21	0.96	1.00
	40-44	1.07	0.00	0.68	1.45	1.00
	45-49	1.43	0.00	0.98	1.88	1.00
	In union	0.05	0.00	-0.16	0.27	1.00
	1	-0.03	0.00	-0.43	0.37	1.00
	2	0.04	0.00	-0.39	0.48	1.00
	3	0.64	0.00	0.17	1.12	1.00
	4+	0.68	0.00	0.18	1.19	1.00
% of educated woman		0.02	0.00	0.01	0.02	1.00
Mean age at first sex		0.11	0.00	0.04	0.18	1.00
% of black women		-0.01	0.00	-0.01	-0.00	1.00
Others	intercept	-11.4	0.70	-19.72	-5.20	1.00
Ages	25-29	0.53	0.00	-0.91	2.00	1.00
	30-34	0.49	0.06	-1.21	2.16	1.00
	35-39	0.98	0.04	-0.88	2.78	1.00
	40-44	0.24	0.20	-2.11	2.31	1.00
	45-49	0.98	0.11	-1.51	3.14	1.00
Marital status	In union	1.14	0.00	0.06	2.29	1.00
Number of children	1	-0.44	0.00	-1.92	1.18	1.00
	2	-1.33	0.00	-3.06	0.50	1.00
	3	-0.69	0.12	-2.66	1.36	1.00
	4+	-2.57	0.20	-6.19	0.25	1.00
% of educated woman		0.03	0.00	0.01	0.05	1.00
Mean age at first sex		0.23	0.01	-0.08	0.59	1.00
% of Black women		0.00	0.01	-0.02	0.02	1.00

Before we can make inference we need to check if parameters have converged

4.5.4 Convergence Diagnostics

To asses convergence this study used autocorrelation plots, probability density plots, trace plots and geweke plots. The posterior probability density functions (pdfs) for the multinomial logistic regression model parameters are given in this section. We discuss the results of the First 9 parameters and the posterior distributions of the rest of parameters are summarized in Table 4.17

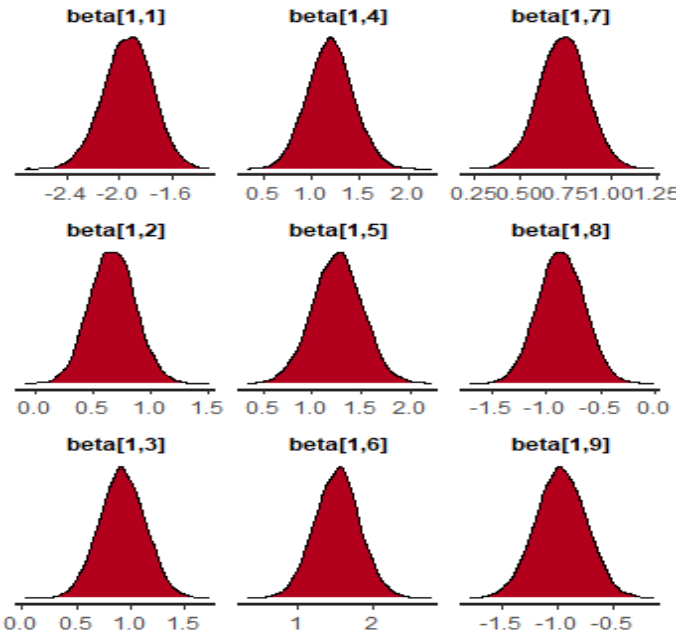


Figure 4.4: Kernel/ density plots of Markov Chain Monte Carlo for the first 9 parameters.

Figure 4.4 Displays the posterior distribution of the first 9 model parameters.

probability density estimate of the parameter. The peak of the distribution (the posterior mode) is the most likely. More satisfactory kernel density plots for parameters of interest would be more bell-shaped or symmetric. Hence, the density plots above show that convergence of the chain has been reached.

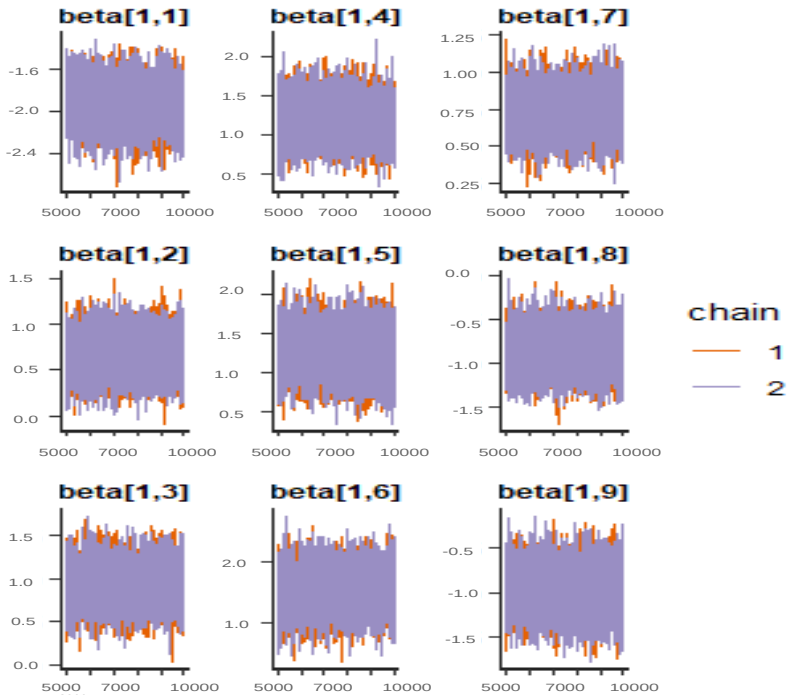


Figure 4.5: Trace plots for the first 9 parameters.

Figure 4.5 displays time series of a parameters in the model as MCMC iterates. The caterpillars are fuzzy indicating that the MCMC chains have mixed well. The Gelman and Rubin trace plots show the convergence of the two parallel chains (Chains with different initial values).

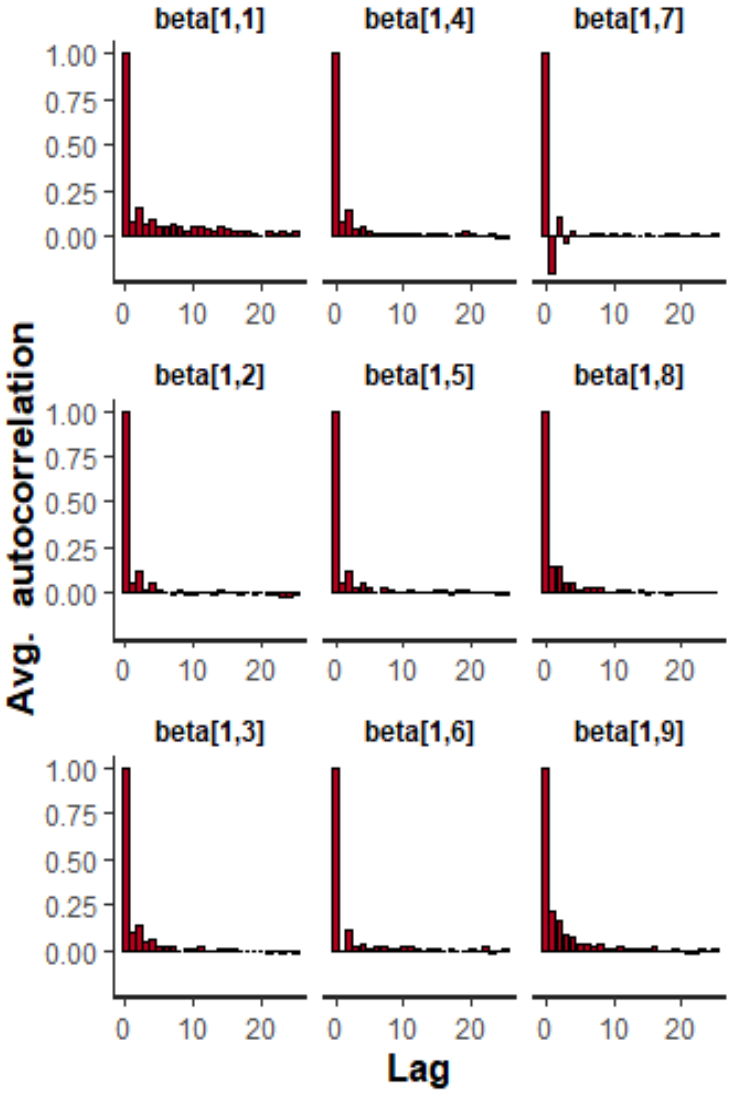


Figure 4.6: Trace plots of Markov Chain Monte Carlo for the first 9 covariates .

Figure 4.6 shows the autocorrelation plots. These plots appear to dampen quickly; therefore, this provides evidence of the convergence of the Markov chain and suggests that it may be appropriate to average Markov chains

output.

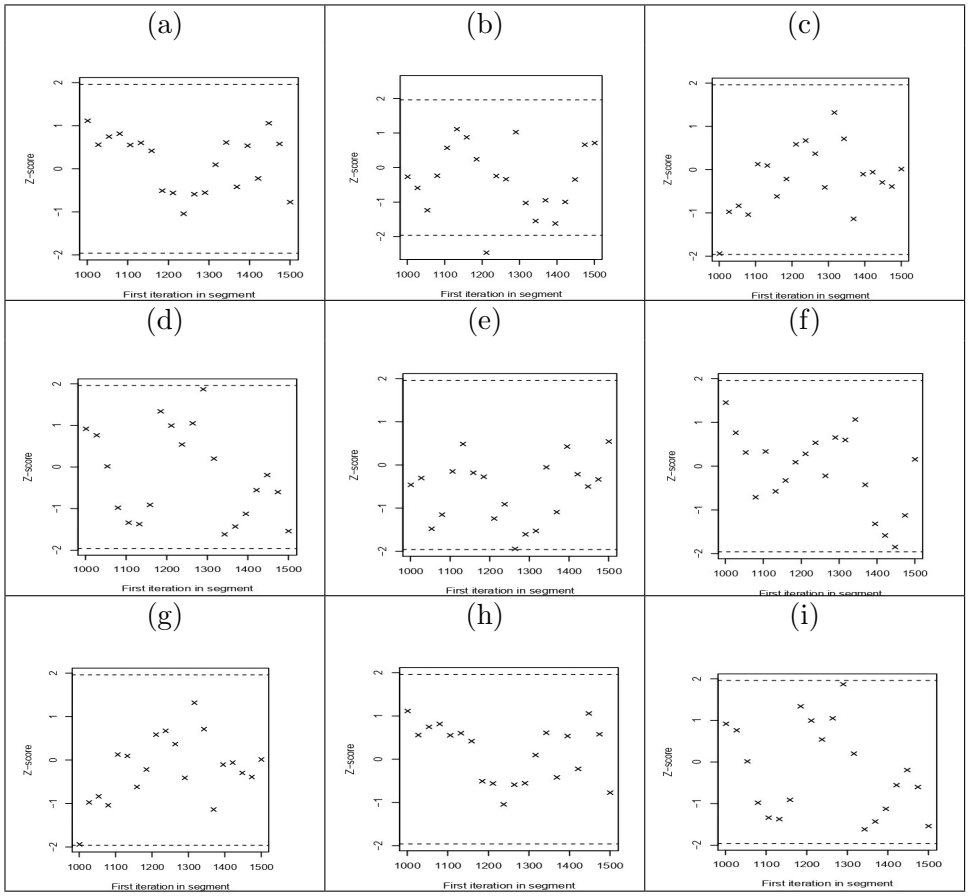


Figure 4.7: Geweke plots for the first nine coefficients from the posterior distribution

Geweke plots for the first few parameters are displayed in Figure 4.7. There is no significant proportion of Z−scores outside the two-standard deviation bands (ignorable amount can be seen in (b)). Based on the rule of thumb, the plots are indicative of a chain that have converged after 10000 iterations. Since MCMC chain have converged, we can make inferences from our model.

To interpret parameters in Table 4.17, we compute odds ratios and the results are stored in Table 4.18.

Table 4.18: Odds ratios of the parameters of a final multilevel Bayesian model.

Variables		OR	95% Credible interval of OR	
			Lower	Upper
Oral	intercept	0.06	0.01	0.26
Age	25-29	1.80	1.22	2.64
	30-34	2.23	1.04	3.35
	35-39	2.80	1.77	4.48
	40-44	2.89	1.75	4.81
	45-49	3.67	1.97	6.62
Marital status	In union	1.88	1.45	2.44
No of children	1	0.49	0.24	0.75
	2	0.47	0.29	0.74
	3	0.35	0.21	0.61
	4+	0.24	0.13	0.45
% of educated woman		1.02	1.01	1.03
Mean age at first sex		1.05	0.97	1.14
% of black woman		1.00	0.99	1.00
Barricidal	intercept)	0.02	0.004	0.07
Age	25-29	2.27	1.72	3.03
	30-34	3.82	2.72	5.10
	35-39	7.54	5.31	10.70
	40-44	6.42	4.39	9.58
	45-49	16.94	10.91	26.58
Marital status	In union	0.68	0.55	0.90
Number of children	1	0.18	0.14	0.24
	2	0.09	0.06	0.13
	3	0.08	0.06	0.13
	4+	0.06	0.04	0.10
% of educated woman		1.01	1.00	1.02
Mean age at first sex		1.15	1.06	1.25
% of black woman		1.02	1.01	1.02
Long-acting	intercept	0.04	0.01	0.14
Age	25-29	0.89	0.63	1.25
	30-34	1.30	0.91	1.84
	35-39	1.80	1.23	2.61
	40-44	2.92	1.97	4.26
	45-49	4.18	2.66	6.55
Marital status	In union	1.05	0.85	1.31
Number of children	1	0.97	0.65	1.45
	2	1.04	0.67	1.62
	3	1.90	1.19	3.06
	4+	1.97	1.20	1.21
% of educated woman		1.02	1.01	1.03
Mean age at first sex		1.12	1.04	1.20
% of black women		0.99	0.99	1.00
Others	intercept	0.00	0.00	0.001
Ages	25-29	1.70	0.40	7.39
	30-34	1.63	0.30	8.67
	35-39	2.66	0.41	16.12
	40-44	1.27	0.12	10.07
	45-49	2.66	0.22	23.10
Marital status	In union	3.13	1.06	9.87
Number of children	1	0.64	0.15	3.25
	2	0.26	0.05	1.65
	3	0.50	0.07	3.90
	4+	0.08	0.002	1.28
% of educated woman		1.03	1.01	1.05
Mean age at first sex		0.23	0.92	1.80
% of Black women		1.00	0.98	1.02

For the Bayesian method, stan (Bayesian Inference using HMC) is used to fit the model.

Relative to women aged $15-19$, the odds of using contraceptives increase with age, peaking at $35-39$, for long acting and barrier methods; OR=1.80, 95%CI [1.23, 2.61]; OR =7.54, 95%CI [5.37, 10.07]; OR=4.036, 95%CI [2.61, 6.25]; for oral, and long-acting methods respectively. The results further show that, under barrier methods, there is a general increase in the odds of using barrier over injection; with age peaking at $40-44$ and then falls thereafter. The higher odds associated with those in age group $35-39$ may be because of the elevated sexual habits that are characteristic of people in that age range.

Compared to women without children, those who have 3 children were more likely to use long acting methods, which include sterilisation (a permanent method choice); OR=1.90, 95%CI [0.90,0.36], it would be reasonably to think that woman with 3 or more children would have reached their desired number of children, since 3 is above South African's total fertility ratio. Thus explains the higher odds of using Sterilization which is a more permanent method.

With reference to unmarried (not in-union) women, the odds of a married woman using oral contraceptives over injection are almost double those of an unmarried woman, (OR = 1.88, 95% CI = [1.45 -2.44]), holding the effects of other variables in the model constant.

For age at first intercourse (OR = 1.15, 95% CI: [1.06, 1.25]), one unit increase in age at first sex results in 1.15 times the likelihood of using a barrier

method over injection. perhaps this may be due to the fact that people who practice late sex, they tend to be responsible since they start having sex while they are already matured.

For community level education, one unit increase in the percentage of educated woman inflates the odds of choosing any other contraceptive method over injection. OR=1.03, 95%CI [1.01, 1.05]; OR =1.02, 95%CI [1.0,1.02]; OR=1.02, 95%CI [1.01, 1.03]; for oral, barrier, long-acting and other methods respectively.

Discussion and Partial Conclusion

In this section we reported empirical results of applying multilevel Bayesian statistics in modelling individual and community level factors that influence individual woman's choice of contraceptive.

The use of non informative prior afforded the data an opportunity to speak for it self. This study shows that Bayesian have a better way of handling uncertainty, we see that from Bayesian results which uses credible intervals of a probability density to estimate a parameter of interest, shifting away from confidence interval of point estimates in the frequentist approaches.

4.5.5 Statistical Computing

We used the package **rstan** in R to compute the posterior distributions using an MCMC algorithm. In **rstan** we specified non informative priors for the fixed effects with very large standard deviation. as suggested by Ebenezer and Lougue (2019). The rstan package uses NUST to run well behaved MCMC chains. For model comparison we computed the WAIC from **loo**

package in **R**,

Chapter 5

Discussion and Conclusions

5.1 Introduction

This chapter presents discussion based on the study objectives followed by conclusion and future research.

5.2 Discussion

In the last chapter we fitted multilevel models to analyse the SADHS data. The frequentist the novel Bayesian multilevel models are the models considered. The SADHS dataset had a three level hierarchical structure where women were nested within households and households within communities/ primary sampling units(PSUs). In our analysis we restricted the data to two levels, where women are nested within communities, we ignored the household level because in most cases there were few women per household.

Both the frequentist and the Bayesian multilevel models yielded similar inference about the parameters. We noted differences in standard errors of

estimate. Standard errors of estimates from the frequentist model were lower than the ones obtained in Bayesian analysis. Another notable difference was that the odds ratios of frequentist models are generally lower than those from the Bayesian models. Now, in comparison of the parameter estimates, the results from Bayesian model with non informative prior were similar to those obtained by the frequentist model. Thus decision of choosing between frequentist and Bayesian with non informative prior should depend on other factors like computation speed, though Bayesian results have more information, because one can choose to interpret posterior means or median, a choice that does not exist under the frequentist modelling framework.

In an attempt to find the determinants of contraceptive choices, fitted multilevel models revealed that there exists community variations in determinants of contraceptive choices, since all level 1 and level 2 covariates were found to be significantly associated with contraceptive choices.

Increase in percentage of well educated woman in a community changes the contraceptive behaviour of women, especially reducing the percentage of injection use. Injection is by far the most common contraceptive method in South Africa. However, for well-educated women, injection is not the main choice; they often choose barrier methods or more permanent methods.
It was found that, women who are not in union have reduced odds of using barrier methods over injection. This is alarming because unmarried women are less likely to be in stable relationships. The fact that the odds of using any type increases with age, it only means a lot has to be done to encourage youth since it might be too late when they start using contraceptives. The fact that young ones are less likely to use any method including condoms

which prevent both conception and STIs leaves them at a very high of contracting and spreading STIs including HIV.

The results from this study shed a light on how the community dynamics can influence women's choice of contraceptive methods. Policy makers and program managers may use this new information to help better shape implementation and provision of family planning programs. Lastly empowering women by providing more educational opportunities would have a significant effect on their choices.

5.3 Conclusion

To this date contraceptive studies remains heavily invested in modelling the use, the current study addressed that issue by modelling choices. Our study has extended the application of qualitative response analysis to contraceptive studies, and the use of Bayesian analysis to model choices as called for by earlier studies. Specifically, we used multinomial logit analysis to identify key variables that affect individual woman's choice of contraceptive method. The results suggest that the key determinants are, education, Age and number of children a woman has.

5.4 Future research

There is still a need for exploration of multilevel multinomial models most especially in the case of non responses. For contraceptive studies it would help to explore the male population as well.

References

H. Akaike. Maximum likelihood identification of gaussian autoregressive moving average models. *Biometrika*, 60(2):255–265, 1973.

K.A. Aragaw. Application of logistic regression in determining the factors influencing the use of modern contraceptive among married women in Ethiopia. *Am J Theor Appl Stat*, 4(3):156–62, 2015.

R. Arango, J. Lemaitre, E. Burbano, C.and Lamprea, and P. Rueda. Jurisprudencia constitucional sobre el derecho al mínimo vital, 2002.

M.J. Bayarri and J.O. Berger. The interplay of Bayesian and frequentist analysis. *Statistical Science*, pages 58–80, 2004.

T. Bayes. An essay towards solving a problem in the doctrine of chances, by the late Rev. Mr. Bayes, F. communicated by Mr Price, in a letter to John,c. *Philosophical transactions of the Royal Society of London*, (53): 370–418, 1763.

M. Betancourt and M. Girolami. Hamiltonian monte carlo for hierarchical models. *Current trends in Bayesian methodology with applications*, 79 (30):2–4, 2015.

W. J. Browne and J. Rasbash. *MCMC estimation in MLwiN*. Citeseer, 2009.

B. Carpenter, A. Gelman, M.D. Hoffman, D. Lee, B. Goodrich, M. Betancourt, M.A. Brubaker, J. Guo, P. Li, and A. Riddell. Stan: a probabilistic programming language. *Grantee Submission*, 76(1):1–32, 2017.

B. Chigbu, S. Onwere, C. Aluka, C. Kamanu, O. Okoro, and P.J.C Feyi-Waboso. Contraceptive choices of women in rural South Eastern Nigeria. *Health & place*, 14(4):841–852, 2008.

J. K. Cochran, K. K. Turekian, and G. Keller. Radiocarbon and 210pb distribution in submersible-taken deep-sea cores from project famous. *Earth and Planetary Science Letters*, 34(2):167–173, 1977.

P.D. Congdon. *Applied Bayesian hierarchical methods*. CRC Press, 2010.

R. Core Team. R: A language and environment for statistical computing, 2013.

F. Cowan and A. Pettifor. HIV in adolescents in Sub-Saharan Africa. *Current Opinion in HIV and AIDS*, 4(4):288–293, 2009.

E. Demidenko. *Mixed models: theory and applications with R*. John Wiley & Sons, 2013.

M. Do and N. Kurimoto. Women's empowerment and choice of contraceptive methods in selected African countries. *International perspectives on sexual and reproductive health*, pages 23–33, 2012.

O. R. Ebenezer and S. Lougue. Bayesian generalized linear mixed modeling of breast cancer. *Iranian journal of public health*, 48(6):1043, 2019.

T. Ferede. Multilevel modelling of modern contraceptive use among Rural and Urban population of Ethiopia. *American Journal of Mathematics and Statistics*, 3(1):1–16, 2013.

W.H. Finch, J.E. Bolin, and K. Kelley. *Multilevel modeling using R*. Crc Press, 2016.

A. Gelman, J. Hwang, and Aki V. Understanding predictive information criteria for Bayesian models. *Statistics and computing*, 24(6):997–1016, 2014.

R.H. Heck and S.L. Thomas. *An introduction to multilevel modeling techniques: MLM and SEM approaches using Mplus*. Routledge, 2015.

A.I. Hermalin. *Fertility regulation and its costs: A critical essay*. University of Michigan, Population Studies Center, 1983.

R. Hodes. The culture of illegal abortion in south africa. *Journal of Southern African Studies*, 42(1):79–93, 2016.

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

I.N.K. Indongo. *Contraceptive use among young women in Namibia: Determinants and policy implications*. PhD thesis, University of Pretoria, 2007.

A. Kaida, F. Laher, S. A. Strathdee, D. Money, P. A. Janssen, R. S. Hogg, and G. Gray. Contraceptive use and method preference among women in Soweto, South Africa: the influence of expanding access to HIV care and treatment services. *PloS one*, 5(11):e13868, 2010.

A. Kulczycki. The determinants of withdrawal use in Turkey: A husband's imposition or a woman's choice? *Social Science & Medicine*, 59(5):1019–1033, 2004.

W. Kungu, A. Agwanda, and A. Khasakhala. Trends and determinants of contraceptive method choice among women aged 15-24 years in Kenya. *F1000Research*, 9(197):197, 2020.

C. Larsson and M. Stanfors. Women's education, empowerment, and contraceptive use in sub-saharan Africa: findings from recent demographic and health surveys. *African Population Studies*, 28:1022–1034, 2014.

E. Lesaffre and A.B. Lawson. *Bayesian biostatistics*. John Wiley & Sons, 2012.

V.L. Luvai. *Hierarchical logistic regression model for Multilevel analysis: an application on use of Contraceptives among women in Reproductive age in Kenya*. PhD thesis, Maseno University, 2017.

M.A. Magadi and S. Curtis. Trends and determinants of contraceptive method choice in Kenya. *Studies in family planning*, 34(3):149–159, 2003.

A.D Martin, K.M. Quinn, J.H. Park, and M.J.H. Park. Package 'mcmcpack', 2020.

R. McElreath. *Statistical rethinking: A Bayesian course with examples in R and Stan*. CRC press, 2020.

D. McFadden and K. Train. Mixed mnl models for discrete response. *Journal of applied Econometrics*, 15(5):447–470, 2000.

C.C. Monnahan, J.T. Thorson, and T.A. Branch. Faster estimation of bayesian models in ecology using hamiltonian monte carlo. *Methods in Ecology and Evolution*, 8(3):339–348, 2017.

W.D. Mosher and J. Jones. Use of contraception in the United states: 1982-

2008. *Vital and health statistics. Series 23, Data from the National Survey of Family Growth*, (29):1–44, 2010.

I. Ntzoufras. *Bayesian modeling using WinBUGS*, volume 698. John Wiley & Sons, 2011.

K.H. Park and P.M. Kerr. Determinants of academic performance: A multinomial logit approach. *The Journal of Economic Education*, 21(2):101–111, 1990.

E. Radovich, M.L. Dennis, K.L.M. Wong, M. Ali, C.A. Lynch, J. Cleland, O. Owolabi, M. Lyons-Amos, and L. Benova. Who meets the contraceptive needs of young women in sub-saharan Africa? *Journal of Adolescent Health*, 62(3):273–280, 2018.

J. Ross and J. Stover. Use of modern contraception increases when more methods become available: analysis of evidence from 1982–2009. *Global Health: Science and Practice*, 1(2):203–212, 2013.

F. J. Samaniego. *A comparison of the Bayesian and frequentist approaches to estimation*. Springer Science & Business Media, 2010.

Carl-Erik Särndal, Ib Thomsen, Jan M Hoem, DV Lindley, O Barndorff-Nielsen, and Tore Dalenius. Design-based and model-based inference in survey sampling [with discussion and reply]. *Scandinavian Journal of Statistics*, pages 27–52, 1978.

G. Schwarz et al. Estimating the dimension of a model. *The annals of statistics*, 6(2):461–464, 1978.

A. Scott and A. Glasier. Evidence based contraceptive choices. *Best practice & research clinical obstetrics & Gynaecology*, 20(5):665–680, 2006.

L. Seutlwadi, K. Peltzer, G. Mchunu, and B.O. Tutshana. Contraceptive use and associated factors among South African youth (18-24 years): A population-based survey. *South African Journal of Obstetrics and Gynaecology*, 18(2), 2012.

D. Spiegelhalter, A. Thomas, N. Best, and D. Lunn. Winbugs user manual, 2003.

J. Starkweather and A.K. Moske. Multinomial logistic regression, 2011.

R. Stephenson, A. Baschieri, S. Clements, M. Hennink, and N. Madise. Contextual influences on modern contraceptive use in sub-Saharan Africa. *American journal of public health*, 97(7):1233–1240, 2007.

R. Stephenson, A. Beke, and D. Tshibangu. Community and health facility influences on contraceptive method choice in the Eastern Cape, South Africa. *International family planning perspectives*, pages 62–70, 2008a.

R. Stephenson, A. Beke, and D. Tshibangu. Contextual influences on contraceptive use in the Eastern Cape, South Africa. *Health & place*, 14(4): 841–852, 2008b.

K. Tanfer, L.A. Cubbins, and K.L. Brewster. Determinants of contraceptive choice among single women in the United States. *Family Planning Perspectives*, pages 155–173, 1992.

United Nations. 2019 revision of world population prospects. 2019.

E. Wagenmakers, M. Lee, T. Lodewyckx, and G.J. Iverson. Bayesian versus frequentist inference. In *Bayesian evaluation of informative hypotheses*, pages 181–207. Springer, 2008.

World Health Organization and Partners in Health. *Management of MDR-TB: A Field Guide, a Companion Document to Guidelines for the Programmatic Management of Drug-resistant Tuberculosis*. World Health Organization, 2009.

Apendices

Selected Rstan codes

```
library(rstan)
data {
int K; // number of contraceptive choices (outcome categories)
int K1; // K-1 alternative categories
int N; // number womens
int P; //Number of women level covariates
int y[N]; // contraceptive choice for each women coded 1 to K
vector[P] x[N]; //women level covariates
int G; // number regions or provinces
int map[N]; // map woman  to regions or provinces
}
transformed data {
vector[K1] zero;
real baseline;
zero = rep_vector(0, K1);
baseline = 0;
}
parameters {
matrix[K1,P] beta; // fixed effects
corr_matrix[K1] omega; // ranef correlations
vector<lower=0,upper=10>[K1] sigma; // ranef scales
vector[K1] u[G]; // random intercepts
}
```

```
transformed parameters{
cov_matrix[K1] V;
V = quad_form_diag(omega, sigma);
}


model {
// prior for beta (vectorized)
for(i in 1:K1) {
    beta[i]~ normal(0,100);
}

// prior/hyper prior for random effects
omega~ lkj_corr(2);
for(g in 1:G) {
u[g] ~ multi_normal(zero, V);
}
{ // local block for linear predictorf
vector[K] xb;
for(n in 1:N) {
xb = append_row(baseline, beta*x[n] + u[map[n]]);
y[n] ~ categorical_logit(xb);
}
}
}
```